



DISCUSSION PAPER SERIES

IZA DP No. 3330

On the Dynamics of Interstate Migration: Migration Costs and Self-Selection

Christian Bayer
Falko Juessen

February 2008

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

On the Dynamics of Interstate Migration: Migration Costs and Self-Selection

Christian Bayer

IGIER, Bocconi University

Falko Juessen

*Technische Universität Dortmund
and IZA*

Discussion Paper No. 3330
February 2008

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

On the Dynamics of Interstate Migration: Migration Costs and Self-Selection^{*}

This paper develops a tractable dynamic microeconomic model of migration decisions that is aggregated to describe the behavior of interregional migration. Our structural approach allows us to deal with dynamic self-selection problems that arise from the endogeneity of location choice and the persistency of migration incentives. Keeping track of the distribution of migration incentives over time has important consequences for the econometrical treatment, because the dynamics of this distribution influences the estimation of structural parameters, such as migration costs. For US interstate migration, we obtain a cost estimate of less than one-half of an average annual household income. This is substantially smaller than the migration costs estimated by previous studies. We attribute this difference to the treatment of the dynamic self-selection problem.

JEL Classification: C61, C20, J61, R23

Keywords: dynamic self-selection, migration, indirect inference

Corresponding author:

Falko Juessen
Department of Economics
Technische Universität Dortmund
44221 Dortmund
Germany
E-mail: falko.juessen@uni-dortmund.de

^{*} We would like to thank Francesc Ortega, Andreas Schabert and conference participants at the North American Summer Meeting of the Econometric Society 2006, the Society of Economic Dynamics Meeting 2006, the Annual Meeting of the European Economic Association 2006, the Annual Meeting of the Verein für Socialpolitik 2006, the Spring Meeting of Young Economists 2007, the Meeting of the Society of Computational Economics 2007, the Meeting of the European Regional Science Association 2007, and at seminars held at IZA, Universität Bonn and Università Bocconi for their helpful comments and suggestions. Part of this paper was written while C. Bayer was visiting fellow at Yale University and Jean Monnet fellow at the European University Institute. He is grateful for the support of these institutions. Financial support by the Rudolf Chaudoire Foundation is gratefully acknowledged. The research has been supported by DFG under Sonderforschungsbereich 475. We would like to thank Christian Wogatzke for excellent research assistance. A previous version of the paper has been circulated under the title "A generalized options approach to aggregate migration with an application to US federal states".

1 Introduction

Migration decisions are important economic decisions. Migration allows individual agents to evade adverse shocks to their income and it is an important way of macroeconomic adjustment (Blanchard and Katz, 1992, and Decressin and Fatas, 1995). Many factors influence the decision to migrate and there is a vast empirical literature that links migration decisions to economic incentives (see Greenwood, 1975, 1985, and 1997 and Cushing and Poot, 2004 for survey articles). At the same time, most of this literature has remained relatively silent about the actual costs of migration to individual agents. Albeit, migration costs are a structural parameter of high interest, both at an individual level as well as from an aggregate perspective (Sjaastad, 1962), for example since more generous unemployment insurance schemes will receive more political support in countries where migration costs are high, see Hassler et al. (2005).

While the literature has in general remained silent about migration costs, a small number of recent studies does report estimates, despite their estimation being technically demanding. For example, Davies, Greenwood, and Li (2001) report a cost estimate of about US\$ 180,000 for each migration between US states, and Kennan and Walker (2003, 2006) conclude that, for a typical move, migration costs are between US\$ 176,000 and US\$ 270,000.¹ This magnitude of migration costs corresponds to roughly 4-6 average annual household incomes. Such an estimate appears very high.

Kennan and Walker (2003) suggest that some kind of omitted variable problem may drive the high cost estimate. In particular, they suggest that an unobservable wage component is correlated to the decision to stay. We argue that the endogeneity of the location choice will always lead to such correlation between unobserved idiosyncratic wage components and location. This endogeneity problem, put in simple words, refers to the fact that agents are in a certain region most likely because they moved there in the past for the reason that they are better off living there. If all observable things are equal, it must be some unobserved component of their preferences that is in favor of the place in which they actually live.

These considerations motivate us to develop a microeconomic structural model of migration which can be aggregated and used to describe the simultaneous evolution of unobservable migration incentives and migration rates at an aggregate level. Our modelling strategy follows Caballero and Engel's (1999) paper on investment, which highlights the interaction of lumpy investment and the evolution of investment incentives.

¹These estimates do not yet include mark-ups for distance and other factors that influence the psychic costs of migration. Return migration is usually associated with lower, but still substantial costs.

For migration, the evolution of incentives is driven by income shocks and past migration. Say a household living in one region is earning a low current income, but faces a substantially higher potential income in an other region. This household is very likely to migrate. As a result, the number of households facing large income differentials strongly decreases after migration decisions have been taken, while the number of households facing a smaller income differential changes less. Unless income differentials are fully observable, the resulting distribution of unobserved migration incentives will be endogenous. In particular, it will take a non-standard, unknown form, it will not be symmetric, and it will change over time. Only making explicit this simultaneous evolution in a simulation approach allows us to avoid the problem of unobservability of incentives.

While we can use our structural model to do so, i.e. to express and to track the dynamic evolution of migration incentives at the macroeconomic level, we are constrained to a bi-regional setup for numerical feasibility. This constraint arises because simulating the dynamic evolution of migration incentives is numerically intense. For the same reason, we have to focus exclusively on the migration-income relation, excluding other factors that may be important to migration decisions and shape the heterogeneity of the population in terms of their migration behavior.² Nonetheless, our focus allows us to analyze the importance of the aforementioned dynamic self-selection problem in modelling and analyzing migration behavior. Migration costs are found to be about US\$ 18,500, less than one-half of an average annual household income. This cost estimate is substantially lower than the cost estimates reported by previous studies. Moreover, we show that applying the techniques used in other papers, we would obtain higher cost estimates also from data generated by a simulation of our structural model.

Not only that we estimate substantially lower migration costs when taking into account dynamic self-selection, but we also document a migration dynamics at the micro level that differs from a model which does not keep track of the incentive distribution. For example, agents who have recently migrated are more likely to migrate again over the next years in our model, a fact which has frequently been documented by studies using micro data (e.g. Kennan and Walker, 2006). In our model, this effect is only due to the incentive dynamics, but not because there is any fundamental "mover-stayer" heterogeneity. Taking into account dynamic self-selection, our approach hence complements the structural approaches of Davies, Greenwood, and Li (2001) and Kennan and Walker (2006), who neglect self-selection but account for a richer set of factors that influence

²This research strategy links our paper to Coen-Pirani's (2006) island-economy model of regional migration.

migration decisions.³

Overall, we conclude that keeping track of the distribution of migration incentives over time has an important influence on the estimation of migration costs. This finding extends the role of self-selection problems to a dynamic setup, which so far have been highlighted in static frameworks (see for example Borjas, 1987, Borjas, Bronars, and Trejo, 1992, Tunali, 2000, and Hunt and Mueller, 2004).

The remainder of this paper is organized as follows: Section 2 gives a brief discussion of the difficulties of estimating structural migration models when the population dynamically self-selects into its preferred region. The section develops the main motive of our paper and illustrates why migration costs are hard to estimate by standard (discrete choice) estimation techniques. Motivated by these considerations, Section 3 presents a dynamic microeconomic model of the migration decision which assumes that an agent maximizes future expected well-being by location choice. In Section 4, we show how to aggregate this model. We derive the contemporaneous law of motion of the distribution of migration incentives and aggregate migration rates. Section 5 confronts the model with aggregate data on migration between US states and presents the estimates of the structural parameters of the model, in particular the estimates of migration costs. Section 6 investigates in more detail the role of self-selection for our estimation, both for the biases in static estimation and for the micro structure of migration. Section 7 concludes and an appendix provides detailed proofs as well as details on the data employed.

2 What makes migration costs so hard to measure?

Most micro studies and now also more macro studies on migration link the individual migration decision to a probabilistic model in which agent i migrates at time t if the gain in utility terms obtained by migration,

$$\left(u_{it}^{\text{move}} - u_{it}^{\text{stay}}\right) = \gamma x_{it} + \nu_{it}, \quad (1)$$

is large enough and exceeds some threshold value \bar{c} , see for example Davies, Greenwood, and Li (2001), Hunt and Mueller (2004), or Kennan and Walker (2006). This threshold value \bar{c} can be interpreted as migration costs in utility terms. The vector of covariates x_{it} is composed of information that describes economic incentives to migrate, i.e. the gains from migration. The vector of parameters γ measures the sensitivity of the migration

³Kennan and Walker's (2006) as well as our model are based on the real-options approach to migration suggested by Burda (1993) and Burda et al. (1998). However, the latter two papers only look at migration as a once and for all decision, so that they preclude return migration. Moreover, they do not study the evolution of migration incentives, to which past migration decisions feed back.

decision to these economic incentives. The stochastic component ν_{it} reflects differences across agents, omitted migration incentives, and/or some variability of migration costs.

Typically, we are interested in the structural parameters γ and \bar{c} and hence would estimate some version of (1) to infer these parameters. Unfortunately, such direct approach is very difficult as *potential* migration gains are unobservable to the econometrician.

To illustrate this difficulty, suppose that an agent only cares about the difference in income between home and destination region, regions A and B in the following. In such setting, x_{it} would be simply a measure of relative income potentials for an agent which she can realize by location choice. A rational agent then moves to the region where she earns the most, provided that her migration costs are covered by the discounted present value of the differences in future incomes.

However, the econometrician can only observe the income that an agent realizes in the region in which she is currently living. Therefore, the other, the unobserved, potential income has to be proxied. Typically, it is proxied by an income a similar agent realizes in the other region. One example for this approach is the paper by Hunt and Mueller (2004), which applies Mincer-type wage regressions to obtain the unobservable potential income. A similar example can be found in Burda et al. (1998) or Kennan and Walker (2006). At a macro level, this approach often means replacing agent-specific income differences by average income differences across regions, see for example Davies, Greenwood, and Li (2001).

If we proxy the unobservable income difference x_{it} for individual i in equation (1) by the average income difference $\bar{x}_{.t}$ between source and destination region, then we obtain

$$\left(u_{it}^{\text{move}} - u_{it}^{\text{stay}}\right) = \gamma \bar{x}_{.t} + \underbrace{\gamma (x_{it} - \bar{x}_{.t}) + \nu_{it}}_{\text{composed error term}}. \quad (2)$$

The composed error term $\gamma (x_{it} - \bar{x}_{.t}) + \nu_{it}$ now also includes the idiosyncratic component of income differences $\eta_{it} := (x_{it} - \bar{x}_{.t})$. In a microeconomic interpretation of (2), $\bar{x}_{.t}$ is the income component that is explained by all observable characteristics of the agent in a Mincer-type wage regression and η_{it} is the unexplained residual. Our argument applies to this microeconomic interpretation too.

Since we do not want to base our argument on a classical measurement error or omitted variable problem, assume that the idiosyncratic component to the income difference η_{it} is orthogonal to the average income difference $\bar{x}_{.t}$. For the ease of exposition, suppose in addition that the agent really just cares about income, so that the true stochastic component is actually identical to zero, $\nu_{it} \equiv 0$.

Under these assumptions, we can rewrite (2) as

$$\left(u_{it}^{\text{move}} - u_{it}^{\text{stay}}\right) = \gamma \bar{x}_{.t} + \gamma \eta_{it}. \quad (3)$$

In this equation, the regression residual only captures the distribution of idiosyncratic potential income differences around the mean.

While the migration decision is deterministic to the individual in this setting, it is stochastic to the econometrician due to his lack of knowledge of η_{it} . If the econometrician were to know the distribution of the unobserved component η_{it} , he would nonetheless be able to estimate γ with a suitable probabilistic discrete choice model. However, it is problematic to assume one of the standard distributions for η_{it} , e.g. a probit distribution.

Suppose agents are heterogeneous with respect to their potential incomes, so that the idiosyncratic component η_{it} has a non-degenerated distribution. In particular, assume that η_{it} is initially normally distributed as displayed in Figure 1 (a), so that in the initial situation a probit model would be appropriate. The figure displays the distribution of migration incentives, i.e. potential incomes, $x_{it} = \bar{x}_{.t} + \eta_{it}$. Low values of this sum imply that income in region *A* is favorable, high values of this sum imply better income prospects in region *B*. Correspondingly, all agents with $\bar{x}_{.t} + \eta_{it} < 0$ decide to live in region *A* and they decide to live in region *B* otherwise if we assume zero migration costs for the moment. In other words, agents self-select into the region that is favorable for them.⁴

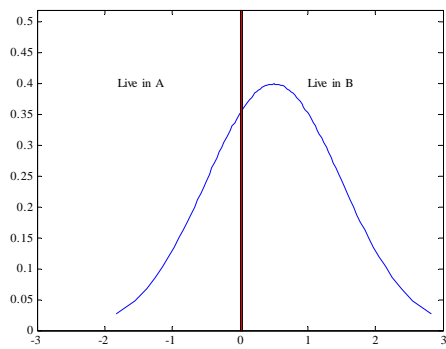
As a result of this self-selection, the distribution of income differences changes for the next period. No agent who lives in region *A* prefers to live in region *B*. This means that for those agents who live in region *A* the distribution of income differences is as displayed in Figure 1 (b). Effectively, the right-hand part of the distribution in Figure 1 (a) has been cut because all agents with higher income in region *B* have actually chosen *B* as the region to live in.

It can be seen that migration incentives $\bar{x}_{.t} + \eta_{it}$ are no longer normally distributed conditional on a household living in region *A*. Since the estimation residual $\gamma \eta_{it}$ in our setup results from a linear transformation of the migration incentive $\bar{x}_{.t} + \eta_{it}$, also the estimation residual $\gamma \eta_{it}$ is no longer normally distributed. Accordingly, the distributional assumptions to estimate (1) by standard maximum likelihood techniques are no longer fulfilled.

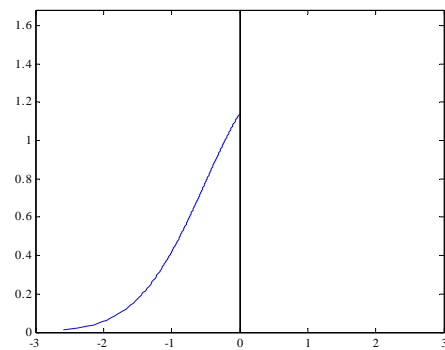
Even adding a normally distributed idiosyncratic income shock does not reestablish a

⁴This self-selection is driven directly by the heterogeneity of agents with respect to potential incomes, but it does not reflect immanent and fixed differences of the regions as in Borjas (1987) and Borjas, Bronars, and Trejo (1992).

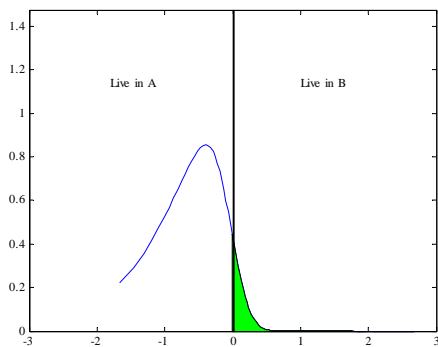
Figure 1: Distribution of potential incomes in region B relative to A



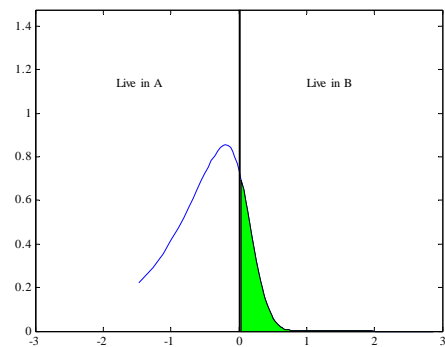
(a) overall population



(b) conditional on living in region A after migration



(c) conditional on living in region A after migration and idiosyncratic shocks



(d) conditional on living in region A after migration, idiosyncratic, and aggregate shocks

normal distribution of income differences if income differences are sufficiently persistent. Figure 1 (c) displays how mild idiosyncratic shocks alter the distribution displayed in Figure 1 (b). Again, the distribution is different from the standard distributions assumed in the estimation of discrete-choice models. The colored-in region indicates the set of agents that will migrate from A to B after the idiosyncratic shocks.

Besides idiosyncratic shocks, also aggregate shocks to the average income difference \bar{x}_t influence the migration decisions of agents. Figure 1 (d) shows the distribution of migration incentives as in Figure 1 (c), but after an adverse shock to region A . By comparing Figures 1 (c) and 1 (d), one can see that the shape of the distribution after migration (the not colored-in region) differs between both figures. In consequence, the distribution of migration incentives will not be strictly stationary, it will evolve over time, and it will depend on the history of aggregate shocks.

Now, how does this correspond to an unreasonable estimate of migration costs? If the parameter \bar{c} in (1) is normalized to 1, the parameter γ has a straightforward interpretation. It measures the sensitivity of migration decisions to income incentives and its inverse $\frac{1}{\gamma}$ is exactly the income differential at which an average agent is just indifferent between moving and not moving. Or to put it differently, $\frac{\bar{c}}{\gamma}$ is the money measure of average migration costs.

In turn, this implies that any bias in the estimate of \bar{c} or γ directly translates into a bias in estimated migration costs $\frac{\bar{c}}{\gamma}$. With the distribution of migration incentives mis-specified, \bar{c} and/or γ will be estimated with a bias most probably. The mis-specification of the distribution of migration incentives has two aspects. One is that the distribution will always be non-standard, i.e. neither normal nor logistic. The second aspect is that the distribution also changes over time as a result of aggregate shocks to income and the triggered migration decisions. Accordingly, one needs to keep track of the evolution of the incentive distribution and standard techniques to deal with self-selection cannot be applied in a straightforward way. Therefore, we develop a model based on dynamic optimal migration decisions in the presence of persistent shocks to income. This model can then be aggregated and used to simulate the evolution of migration and its incentives over time.

3 A simple stochastic model of migration decisions

We consider an economy with two regions, A and B . This economy is assumed to be inhabited by a continuum of infinitely lived agents of measure 1. We model the economy in discrete time and at each point in time an agent has to decide in which region to live and work. First, we consider the decision problem of an individual agent i living in

region $j = A, B$. Thereafter we discuss aggregation and the dynamics of the distribution of migration incentives.

Living in region j at time t gives the agent utility \tilde{w}_{ijt} . Although \tilde{w}_{ijt} could be understood as a catch-all variable for migration incentives, which can be interpreted as wage income, employment prospects, amenities, utility from social networks and so on, we restrict ourselves to an interpretation of \tilde{w}_{ijt} as income.

Migration incentives (incomes) are stochastic in our model. We assume them to be composed of a permanent (autocorrelated) component \tilde{w}_{ijt} and a transitory (i.i.d.) component φ_{ijt} .⁵ These components vary over time and across individuals. For simplicity, we assume that only the permanent component is observed before the migration decision. Consequently, in describing migration behavior, we can focus exclusively on the effect of permanent variations in potential incomes. Changes in the transitory component realize after migration and hence do not affect migration choice. However, when confronting our model to data, including data on aggregate income, we need to take transitory income fluctuations into account. This means that the microeconomic model does not need to include the transitory income component, while we have to take it into account in the aggregation of realized incomes.

Moving from one region to the other is not costless to an agent. When an agent moves, she is subject to a disutility c that enters additively in her utility function. Therefore, the instantaneous utility function $u_{it}(j, k)$ is given by

$$u_{it}(j, k) = \tilde{w}_{ijt} - \mathbb{I}_{j \neq k} c \quad (4)$$

for an agent that has lived in region k in the preceding period and now lives in region j . Here, \mathbb{I} denotes an indicator function, which equals 1 if the agent has moved from region k to j and 0 if the agent already lived in region j before.

The agent discounts future utility by factor $\beta < 1$ and maximizes the discounted sum of expected future utility by location choice. The agent knows the distribution of the permanent component of income \tilde{w}_{ijt} and forms rational expectations. With \tilde{w}_{ijt} being stochastic, the potential migrant waits for good income opportunities. In her migration decision the agent thus takes into account the option value to wait and learn more about future incomes.

The distribution of migration incentives, \tilde{w}_{ijt} , is assumed to be log-normal. In particular, we assume that log income, w_{ijt} , follows an AR(1) process with normally distributed

⁵See the evidence on transitory income fluctuations provided by Storesletten, Telmer, and Yaron (2004) for instance.

innovations ξ_{ijt} and autoregressive coefficient ρ :

$$\ln(\tilde{w}_{ijt}) =: w_{ijt} = \mu_j (1 - \rho) + \rho w_{ijt-1} + \xi_{ijt}, \quad j = A, B. \quad (5)$$

This process holds for the whole continuum of agents and each agent draws her own series of innovations ξ_{ijt} for both regions. The expected value of log income in region j is μ_j . The innovations ξ_{ijt} are composed of aggregate as well as idiosyncratic components. They have mean zero, are serially uncorrelated, but may be correlated across regions A, B (see Section 4.2.2).⁶

Income and cost distributions, together with the utility function and the discount factor define the decision problem for the potential migrant. This is an optimization problem, which is described by the following Bellman equation:

$$V(k, w_{iAt}, w_{iBt}) = \max_{j=A, B} \left\{ \exp(w_{ijt}) - \mathbb{I}_{\{k \neq j\}} c + \beta \mathbb{E}_t V(j, w_{iAt+1}, w_{iBt+1}) \right\}. \quad (6)$$

In this equation, \mathbb{E}_t denotes the expectations operator with respect to information available at time t .⁷

The optimal policy is relatively simple. The agent migrates from region k to region j if and only if the costs of migration are lower than the sum of direct benefits of migration $\exp w_{ijt} - \exp w_{ikt}$ and the expected value gain

$$\Delta V(w_{iAt}, w_{iBt}) := \beta \mathbb{E}_t [V(B, w_{iAt+1}, w_{iBt+1}) - V(A, w_{iAt+1}, w_{iBt+1})]. \quad (7)$$

This means that the agent migrates if and only if

$$c \leq \exp(w_{ijt}) - \exp(w_{ikt}) + \Delta V(w_{iAt}, w_{iBt}) =: \bar{c}(w_A, w_B). \quad (8)$$

This gives a critical level of costs $\bar{c}(w_A, w_B)$ at which an agent living in region A is indifferent between moving and not moving to region B. A person moves from A to B if and only if

$$c \leq \bar{c}_A := \bar{c}(w_A, w_B).$$

Conversely, a person living in region B moves to region A if and only if

$$c \leq \bar{c}_B := -\bar{c}(w_A, w_B).$$

⁶For technical reasons, we assume boundedness of ξ_{ijt} , so that ξ_{ijt} is in fact only *approximately* normal.

⁷Existence and uniqueness of the value function is proved in the appendix.

Note that \bar{c} can be positive as well as negative. If \bar{c} is positive, region B is more attractive. If it is negative, region A is more attractive and a person living in region A would only have an incentive to move to region B if migration costs were negative.

4 Aggregate migration and the dynamics of income distributions

4.1 Aggregate migration

Given this trigger rationale for migration, the hazard rate

$$\Lambda_j(w_A, w_B) := \begin{cases} 1 & \text{if } c \leq \bar{c}_j(w_A, w_B) \\ 0 & \text{otherwise} \end{cases}, \quad j = A, B$$

defines whether a person in region j moves to the other region if she faces the potential incomes (w_A, w_B) .

Now, consider the distribution F_t of (potential) incomes (w_A, w_B) and household locations. Suppose this joint income and location distribution is the distribution after the income shocks ξ_{ijt} have been realized, but before migration decisions have been taken. Let f_{jt} denote the conditional density of this distribution, conditional on the household living in region j at time t . Then, the actual fraction $\bar{\Lambda}_{jt}$ of households living in j that migrate to the other region evaluates as

$$\bar{\Lambda}_{jt} := \int \Lambda_j(w_A, w_B) \cdot f_{jt}(w_A, w_B) dw_A dw_B. \quad (9)$$

The aggregate migration hazard can be thought of as a weighted mean of all microeconomic migration hazards $\Lambda_j(w_A, w_B)$, weighted by the density of income pairs (w_A, w_B) from distribution F_t .

4.2 Dynamics of income distributions

The distribution F_t itself (and hence f_{jt}) evolves over time and is a result of direct shocks to income just as it is a result of past migration. We need to characterize the law of motion for F_t to close our model and to obtain the sequence of aggregate migration rates.

4.2.1 The effect of migration on income distributions

In order to follow the evolution of F_t we need to characterize both the evolution of the fraction P_{jt} of households living in each region and the conditional distribution of incomes f_{jt} (conditional on a household actually living in a specific region j).

The proportion of households living in region j at time $t + 1$ is a result of migration

decisions at time t . The law of motion for P_{jt} is thus given by

$$P_{jt+1} = (1 - \bar{\Lambda}_{jt}) P_{jt} + \bar{\Lambda}_{-jt} P_{-jt}. \quad (10)$$

The first part of the sum reflects the fraction of households that remain in region j , where $(1 - \bar{\Lambda}_{jt})$ is the probability to stay in region j . The second part is the fraction of households that migrate from region $-j$ to region j .

Since the microeconomic migration hazard depends on (w_A, w_B) , different potential incomes result in different propensities to migrate. In consequence, migration changes not only the fraction P_{jt} of households living in region j at time t , but also the conditional distribution of income, f_{jt} . For example, households living in region A , earning a low current income, w_A , but facing a substantially higher potential income in B , w_B , will probably migrate. As a result, the number of those households will drop to zero in region A after migration decisions have been taken, while the number of households facing a smaller income differential might not change, recall Figure 1.

The distribution of migration incentives is thus a function of past migration decisions, and we can express the new density of households with income (w_A, w_B) in region j after migration, \hat{f}_{jt} , by

$$\hat{f}_{jt}(w_A, w_B) = [1 - \Lambda_{jt}(w_A, w_B)] \frac{f_{jt}(w_A, w_B) P_{jt}}{P_{jt+1}} + \Lambda_{-jt}(w_A, w_B) \frac{f_{-jt}(w_A, w_B) P_{-jt}}{P_{jt+1}}. \quad (11)$$

The probability $[1 - \Lambda_{jt}(w_A, w_B)]$ is again the probability to stay in region j . The term $f_{jt}(w_A, w_B) P_{jt}$ weights this probability and is the *unconditional* income density for region j before migration has taken place. To obtain the *conditional* density after migration, the unconditional income density, $f_{jt}(w_A, w_B) P_{jt}$, is divided by P_{jt+1} , which is the fraction (or probability) of households living in region j after migration (i.e. in time $t + 1$). Analogously, the second part of the sum is constructed.

4.2.2 The effect of income shocks on the income distribution

Besides migration, also shocks to income change the distribution of income pairs, F_t . The shocks to income can be differentiated along two dimensions: One dimension is aggregate vs. idiosyncratic, the other one is region-specific vs. economy-wide. For a single agent we can decompose the total shock ξ_{ijt} to her potential income in region j (see equation 5) into an aggregate regional-component θ_{jt} and an individual-specific

regional-component ω_{ijt} :⁸

$$w_{ijt} = \mu_j (1 - \rho) + \rho w_{ijt-1} + \xi_{ijt}$$

$$\xi_{ijt} = \theta_{jt} + \omega_{ijt}, \quad j = A, B.$$

The aggregate shock θ_{jt} for region j hits all agents equally and changes their potential income for region j . Note that this shock does not depend on the actual region the agent is living in. For example, a positive shock $\theta_{At} > 0$ increases the *potential* income in region A for agents that are currently living in this region as well as for agents that are currently living in region B . They realize this potential income by deciding to actually live in region A . The importance of economy-wide business cycles relative to the size of region-specific aggregate fluctuations is reflected by the correlation ψ_θ between aggregate shocks θ_{At} and θ_{Bt} .

Such aggregate shocks are typically only a minor source of income variation for an agent, though. Agents differ in various personal characteristics that result in different income profiles over time. Individuals differ in their skills and while the demand may grow for the skill of one person, demand may deteriorate for another person's skills. This heterogeneity is captured by the idiosyncratic shocks $(\omega_{iAt}, \omega_{iBt})$. If ω_{iAt} is positive, income prospects of the individual agent i increase in region A . The correlation ψ_ω between ω_{iAt} and ω_{iBt} reflects economy-wide demand shifts for a person's individual skills. Since we assume aggregate and idiosyncratic shocks to be independent, the variance of the total shock to income, ξ_{ijt} , is the sum of the variances of idiosyncratic and aggregate shocks: $\sigma_\xi^2 = \sigma_\omega^2 + \sigma_\theta^2$.

Persistency in incomes is captured by the autoregressive parameter ρ in equation (5). We abstain from the inclusion of permanently fixed *individual* differences (fixed effects) primarily because this makes the model numerically more tractable.⁹

Together, aggregate and idiosyncratic shocks to income as well as income persistency determine the transition from \hat{f}_{jt} to f_{jt+1} , details are provided in the appendix. The latter density now determines migration decisions in time $t + 1$, starting the cycle over

⁸Note that this is equivalent to assuming that both aggregate and idiosyncratic (deviations from aggregate) potential incomes follow AR-1 processes with the same autocorrelation ρ to which θ and ω are the respective innovations.

⁹While the solution of the restricted dynamic programming problem of the agent can be obtained quickly, the simulation of the distribution of migration incentives is numerically much more involved. The computation time for the estimation amounts to ca. 24h on a Xeon 3GHz machine.

If we were to include fixed effects that reflect different types of agents, the model had to be solved for each different type of agent in the way it is now solved for the single type of agent. Thus, any modelling of k heterogeneous types of migrants would mean a time need of k days for the estimation.

again. As a result it is both past income shocks and past migration decisions that drive the incentives to migrate. Making this explicit and keeping track of the distributional dynamics of migration incentives is the key element of our model, as it distinguishes our approach from other empirical models of migration.

4.3 Aggregate Income

To link our model to aggregate data, we finally need to describe the evolution of aggregate regional realized incomes. For region j , the fraction of income which is persistent and realized after migration reads as the conditional expected value of $\exp(w_{ijt})$

$$\int \exp(w_j) \hat{f}_{jt}(w_A, w_B) dw_A dw_B.$$

Besides persistent income also transitory components (orthogonal to the persistent part) add to the fluctuations of income in observed data, so that we obtain for log aggregate income \bar{w}_{jt}

$$\bar{w}_{jt} = \ln \left(\int \exp(w_j) \hat{f}_{jt}(w_A, w_B) dw_A dw_B \right) + \varphi_{jt}.$$

The transitory income component φ_{jt} measures fluctuations in income at a high frequency that are irrelevant to the migration decision. More generally, it captures in the empirical approach the idea that in reality income measures migration incentives imperfectly. One reason is that the empirical income concept is noisy as such. Personal income—the income concept we use—comprises labor income as well as some components of capital incomes. Whether each of these elements is too widely or too narrowly defined as a migration incentive is not clear a priori. The inclusion of φ_{jt} reflects this agnostic view.

5 Estimation

5.1 Estimation technique

We rely on an indirect inference procedure in order to find the parameters of our model that allow us to match closest the observed patterns of migration that are in the data. In particular, we apply a method of simulated moments (MSM) as has been proposed by Gourieroux, Monfort, and Renault (1993) to obtain estimates of structural parameters when the likelihood function of the structural model becomes intractable, as in our setting. This estimator relies on numerical simulation of the model, details on the numerical simulation are provided in the appendix.

The idea behind indirect inference is to choose a set of moments that captures the

characteristics of the data, and then to calibrate and simulate the structural economic model such that the moments are best replicated by the simulation. Accordingly, we have to 1) discuss how to use our model of two regions to address migration between a richer set of regions, 2) decide which parameters of our model shall be estimated and which are treated as fix, 3) decide on an informative set of moments along which simulated and real-world data are compared, and 4) decide how to measure the closeness of the simulated and actual moments.

A simulation of our model yields migration and income data for two regions. Of course, the actual migrant faces a more complex decision problem than the one simulated in our model of two regions. Including D.C. as a destination region, an agent has to decide between 50 possible alternative states where she can move to. To make this comparable to our model, the 50 alternatives in the data have to be aggregated to a single complementary region.¹⁰ The average income of the alternative region is proxied by the population-weighted average income over all alternative 50 states.¹¹ With these assumptions, we can use our model to simulate a dataset for migration that has the same size as the Internal Revenue Service (IRS) migration data, which is our empirical benchmark. This database contains annual area-to-area migration flow data for US states for the period 1989-2004.¹² Accordingly, we simulate our model for 51 pairs of regions and 81 years, but we drop the first 55 years for each region to minimize the influence of our initial choice of F_0 . In order to minimize simulation uncertainty, we replicate each simulation 5 times and use the averages over the simulations.

Naturally, we cannot estimate all parameters of the model, since this would be numerically infeasible since every dimension added to the parameter search increases the estimation time exponentially. Our primary interest is to estimate migration costs. Besides migration costs, we restrict ourselves to the estimation of those parameters of our model which cannot be inferred from raw data alone. One such set of estimated

¹⁰Generating artificial bi-regional data means that we assume technically that the best income opportunity over all alternative regions follows the log-normal distribution assumed in our model. Note that an approximation of this sort cannot be avoided by assuming an extreme value distribution for incomes. This would only work if migration incentives were serially uncorrelated.

¹¹Income data is taken from the REIS database, CPI deflated, and in logs.

¹²A detailed data description of both IRS and REIS data can be found in the data appendix. Alternative migration data for the US, such as the Census, are less appropriate for our analysis as we focus on the effect of income dynamics. While the Census reports changes in the place of residence over a period of 5 years and is only available once every decade, the IRS data are available on a yearly basis. The Census data suggests an approximate annual migration rate of 2%, which is significantly lower than the migration rate of 3.9% documented in the IRS data. This difference may stem from the fact that the Census data cannot take into account return migration over a 5 year period, which can be expected to be of sizable importance (see the discussion in Coen-Pirani, 2006 or the results of Kennan and Walker, 2006).

parameters is the correlation of shocks to permanent income across regions, ψ_ω and ψ_θ . Another is the correlation of income shocks across individuals, i.e. the fraction of aggregate shocks, ϕ . These correlations refer to *potential* incomes and are therefore inherently unobservable. We assume that aggregate and individual correlation coefficients are equal, i.e. $\psi_\omega = \psi_\theta$, so that we only need to estimate one common parameter, ψ . Since migration smooths income, the counterparts to ϕ and ψ in terms of a covariance structure in *realized* incomes are substantially influenced by the magnitude of migration costs. At the same time, we also expect ϕ and ψ to have a significant influence on the behavior of aggregate migration itself. Consequently, we can neither estimate them from raw realized income data, nor fix them to arbitrary values.

The transitory shock is irrelevant to the migration decision, and hence its effect on realized incomes cannot be smoothed by migration. For this reason, we fix the correlation of the transitory income shock, ψ_φ , to the value of the observed correlation of incomes in the REIS data, taking fixed effects and a linear time trend into account. However, the variance of the transitory shock, σ_φ^2 , is estimated. The transitory component does not only capture purely transitory fluctuations in income, but also various forms of measurement error. Since we find it difficult to guess their quantitative importance, we abstain from fixing its variance or estimating it outside the model. Our complete set of estimated parameters is $\Theta = (\text{migration costs}, \psi, \phi, \sigma_\varphi^2)$.

In turn, all other parameters $(\beta, \rho, \mu, \sigma_\xi^2, \psi_\varphi)$ in our model have to be fixed to take reasonable values. As we work with annual data, we choose the discount factor $\beta = 0.95$. In order to characterize the process for income, we need to specify the autocorrelation parameter ρ , the mean μ , and the long-run variance of income. We take these parameters from the recent paper by Storesletten, Telmer, and Yaron (2004). They estimate the dynamics of idiosyncratic labor market risk for the US based on the Panel Study of Income Dynamics. Thus the paper conveys information on both income variances and autocorrelation of log household income. Besides, the paper reports a mean household income of US\$ 45,000. To approximately match this figure, we choose the mean of the *log* income to be $\mu_A = \mu_B = 10.5$. Storesletten, Telmer, and Yaron (2004) find an annual autocorrelation of incomes of roughly 0.95 and a standard deviation of idiosyncratic income shocks ranging from 0.09 to 0.14 for business cycle expansions and from 0.16 to 0.25 for business cycle contractions (see Storesletten, Telmer, and Yaron 2004, Table 2). They report a frequency weighted average of 0.17 for those standard deviations in their preferred specification (Storesletten, Telmer, and Yaron, 2004, pp. 711). Since we do not model different variances of idiosyncratic shocks to income along the business cycle, we use their preferred average value of 0.17 for the estimation. Combining both elements,

the autocorrelation and the variance of idiosyncratic shocks to income, we calculate the long-run variance of income to be $\frac{\sigma_w^2}{1-\rho^2} = 0.30$.¹³

A next step is to decide on an informative set of moments, ϱ . We select the standard deviation of migration rates, the standard deviation of average incomes, and the correlation of average incomes across regions as the first three moments to be matched. To this set of moments we add the estimated parameters from a reduced form regression of migration rates on the incomes of the destination and the source region. To make the regression scale-invariant with respect to incomes, we use log-deviations from average incomes as the income variables, i.e. we estimate

$$m_{it} = \alpha_0 + \alpha_1 (w_{it} - \bar{w}_i) + \alpha_2 (w_{-it} - \bar{w}_{-i}) + u_{it}.$$

The parameters α_1 and α_2 reflect income sensitivities of migration. The intercept α_0 captures the average of migration rates.

We simulate our model for a given vector of model parameters Θ and calculate the distance between the moments obtained from this simulation $\hat{\varrho}(\Theta)$ and the sample moments ϱ_S . We use the covariance matrix of ϱ_S obtained by 10000 bootstrap replications as a weighting matrix so that our distance and goodness-of-fit measure is

$$L = (\varrho_S - \hat{\varrho}(\Theta))' \text{cov}(\varrho_S)^{-1} (\varrho_S - \hat{\varrho}(\Theta)).$$

The actual estimation is carried out by minimizing the distance measure L numerically by using a direct search algorithm.

5.2 Estimation results

Table 1 displays the point estimates of the matched moments calculated from the IRS and REIS data and the corresponding moments obtained from the simulation of our model with the estimated parameters. In particular the column "All Moments Matched" refers to the estimation results from our baseline specification. The last two columns report robustness checks discussed in the next subsection. Overall our model is able to replicate the observed moments closely. In fact, the $\chi^2(2)$ -distributed overidentification test reported at the bottom of the table does not reject our model.

¹³The choice to fix σ_ξ^2 is somewhat problematic, but it arises as a necessary restriction from the use of macro data. The realized fluctuation of income is a function of migration costs. Still we use statistics calculated from realized income fluctuations to calibrate a parameter for potential income fluctuations. Yet, the standard deviation of idiosyncratic potential income can be understood as a pure scaling parameter and even at zero migration costs, realized income inherits 89% of the standard deviation of potential income. Consequently, at most we underestimate costs by 15% due to this.

Table 1: Simulated moments estimation: moments estimates

Moment	Actual Data Moments	Simulated Moments		
		All Moments Matched	w/o Average Migration Rate	CRRA (log utility)
Migration Rates				
standard deviation	0.0036	0.0038	0.0037	0.0033
Aggregate Regional Income				
standard deviation	0.0299	0.0294	0.0298	0.0283
correlation across regions	0.5807	0.5714	0.5762	0.5686
Reduced Form Regression				
intercept (\emptyset migration rate)	0.0393	0.0394	(0.0386) ¹	0.0383
sensitivity to income of destination region	0.0603	0.0663	0.0633	0.0494
sensitivity to income of source region	-0.0624	-0.0627	-0.0598	-0.0504
<hr/>				
Overidentification test $\chi^2(2)$		3.3583	1.652 ²	56.90
p-value		0.1865	0.19869	0.000

¹ Not matched; ² Only one degree of freedom.

The column ‘Actual Data Moments’ refers to the moments estimated from the combined REIS/IRS data set, with data on 50 US states and D.C. over the period 1989-2004. The columns ‘Simulated Moments’ refer to the moments estimated from the simulation of the model using the parameters given in Table 2. Both actual and simulated data are within-transformed and linearly de-trended. The simulations generate a panel of 51 region-pairs and a 81-year history of migration and income data. The first 55 years of simulated data are dropped in order to minimize the influence of initial values. Each simulation is repeated 5 times and data moments are compared to the average over the 5 replications of the simulation.

Table 2: Simulated moments estimation: structural parameter estimates

Parameter	All Moments Matched	w/o Average Migration Rate	CRRRA (log utility)
Migration Costs	18,285 ¹ (2, 211)	18,571 ¹ (6756)	0.4762 ² (0.6863)
Correlation of Income Shocks Across Regions Ψ	0.2482 (0.1947)	0.2178 (0.2446)	0.1286 (0.1852)
Fraction of Income Shock due to Aggregate Fluctuations ϕ	0.0041 (0.0011)	0.0040 (0.0034)	0.0035 (0.0011)
Standard Deviation of Transitory Income Shock σ_φ	0.0266 (0.0012)	0.0265 (0.0020)	0.0262 (0.0012)

¹Migration cost estimate \hat{c} in US\$ terms. ² $\exp(\hat{c}) - 1$ measures the relative income gain necessary to offset migration costs.

Standard errors in parenthesis. Estimation is carried out using the simulated moments estimator by Gourieroux, Monfort, and Renault (1993), which chooses structural model parameters by matching the moments from a simulated panel of regions with data moments as displayed in Table 1. The simulations generate a panel of 51 region-pairs and a 81-year history of migration and income data. The first 55 years of simulated data are dropped in order to minimize the influence of initial values. Each simulation is repeated 5 times and data moments are compared to the average over the 5 replications of the simulation.

Table 2 presents the estimates of the model parameters. The estimated migration costs are US\$ 18,285. This is a substantially smaller number than the estimates reported in previous contributions such as Davies, Greenwood, and Li (2001) or Kennan and Walker (2006).

The estimated correlation of income shocks across regions is 24.82%. This is significantly smaller than the observed correlation of realized incomes (58.07%, see Table 1). The realized incomes co-move more strongly than the shocks to income because migration ties together more closely the average incomes in both regions than they were tied together without migration. This drives a wedge between the correlation of income shocks and the correlation of average realized incomes, the latter being larger than the

former.

The estimated fraction of income shocks that is aggregate amounts to 0.41%. To put this number into perspective, we calculate the long-run variance of per capita state income, accounting for a linear time trend and fixed effects, and set this number relative to the variance of household incomes as reported by Storesletten et al. (2004). While this benchmark ratio is with 0.6% somewhat larger, it also points towards aggregate income shocks being very small compared to idiosyncratic ones.

There is a significant transitory income component (measurement error) in the aggregate income fluctuations, which has an estimated standard deviation of 0.0266. This means that transitory fluctuations in aggregate income add a variance term that has about 36.5% of the long-run variance of the sum of *potential* incomes and measurement error. However, migration smooths realized income so that transitory shocks make up 82% of the aggregate variance in *realized* incomes (again the sum of persistent and transitory fluctuations). As outlined before, the transitory income component of our model relates to two sources: to truly transitory fluctuations in incomes and to the fact that income does not measure migration incentives perfectly.

5.3 Robustness checks

One might argue that our relatively low migration-cost estimate is driven by the fact that we attribute all migration to be driven by the income incentive. Correspondingly, the average migration rate our model should match had to be lower. The downside of such an argument is that it would be hard to tell which was the right migration rate to match. Therefore, taking an agnostic view on this point, we provide a robustness check of our estimation results, excluding the average migration rate from the set of moments our model is calibrated to. The estimation results are reported in the corresponding columns of Tables 1 and 2. As one can see, the point estimate for migration costs changes only slightly. Since we drop one moment condition, the standard errors of the estimates increase.

As an additional robustness check, we replace the assumption of a risk-neutral agent and assume constant relative risk aversion instead. Instead of linear utility from income, we now assume logarithmic utility. While one can interpret the risk-neutrality assumption as a short cut for modelling an agent who has access to perfect capital markets, the assumption of logarithmic utility from income can be thought of as a model of restricted capital market access. The respective last column of Tables 1 and 2 report the corresponding estimation results. The estimated cost parameter c is no longer readily interpretable as a US\$ value, instead it reflects the *relative* income gain necessary to

offset the costs of migration. This means that the money measure of migration costs at average income is obtained by multiplying $\exp(c) - 1$ with average income. Our estimate of $c = 0.4762$ implies an estimated migration cost of US\$ 32,738.

6 Behind the Scenes & Further Aspects

A simulation based estimation technique like a method of simulated moments is, by construction, somewhat of a black box. Therefore, we try to shed some light on our estimation results by running counterfactual simulations, by comparing the estimation to static estimation techniques, and by investigating the properties of the individual (micro) migration data our model generates.

6.1 Zero migration costs

We start with a simulation of the model with migration costs set to zero. In a situation in which unobservable migration incentives are serially uncorrelated, i.e. drawn completely anew every period, migration rates would be 50% on average in the absence of migration costs. In such situation of zero costs and i.i.d. incentives, every period half of the population in one region is better off living in the other one. By contrast, this is no longer true if there is serial correlation of incentives. Agents self-select into the region where they are better off, and only those agents that have been on the margin, on the verge of moving, in the previous period are likely to migrate in the current period. To illustrate this point we simulate our model for the counterfactual case of zero migration costs, setting all other parameters to their estimated values. In Table 3 we report some summary statistics for this experiment.

The most important result of this simulation experiment concerns the average migration rate. Only 10.2% of the population migrates each year even if migration costs are zero. This shows why a dynamic view on migration incentives leads to much lower estimates of migration costs.

Besides we see that in a world with costless migration, realized incomes are more strongly correlated, they vary less and are significantly higher (+23.7%) than without any migration. Compared to the simulation results under the estimated migration costs (log average income: 10.876, i.e. +22.6%), the increase is only mild, however. Only moves that lead to small income gains are added by setting migration costs to zero.

6.2 Cost estimates from static models

The evidence from the zero cost specification suggests that it is indeed the dynamic nature of the problem, with agents sorting themselves into their preferred region, that

Table 3: Simulation results: Zero Migration Costs

Moment	Data Moment	Simulated Moment	Moment	Data Moment	Simulated Moment
Migration rates			Reduced form regression ³		
standard deviation	0.0036	0.0074	intercept (\emptyset migration rate)	0.0393	0.1017
Income			sensitivity of migration to income of:		
log average income ¹	10.65 ⁴	10.887	destination region	0.0603	0.0577
standard deviation	0.0299	0.0294	source region	-0.0624	-0.0542
correlation across regions ²	0.5808	0.6390			

¹ Cross-sectional average income of a region; ² Partial correlation controlling for a linear time trend and fixed effect; ³ Coefficients of a reduced form regression of migration rates on incomes in both regions; ⁴ Value used for simulation $(\mu + \frac{\sigma_\xi}{2})$. Original data has been rescaled to have this mean value after de-trending.

The zero cost specification assumes one US\$ of migration costs for numerical feasibility. We simulate data on 51 region-pairs and a 81 year history of migration and income data. The first 55 years of simulated data are dropped in order to minimize the influence of initial values. The simulation is repeated 5 times. The table reports averages over the 5 repetitions.

changes the estimation of migration costs substantially. To provide further evidence, we run three alternative estimation experiments.

- First, we re-estimate our model, setting the autocorrelation of income to zero, keeping fixed the long term variation of income. This experiment tells us which role it plays in our model and estimation to keep track of the distribution of migration incentives.
- Second, we apply a static random utility model to data generated from a simulation of our model using the previously estimated parameters. In this setup, we know the exact size of migration costs by construction. Hence, we can obtain an idea about the bias introduced by static modelling.
- Third, we estimate an approximate version of our model, where the dynamic self-selection that shapes incentive distributions is ignored. In this experiment, we replace the conditional density in (9) by its unconditional counterpart. If self-selection played no role, this replacement was innocent. The former place of residence was of no informational content for unobservable migration incentives and conditional and unconditional distributions coincided. Hence we should obtain similar estimation results as in our baseline estimation specification, if self selection was of no concern.

As we will see, all three versions that ignore the dynamic setup will lead to higher estimated migration costs.

6.2.1 Estimation with zero autocorrelation

As first experiment is to estimate the model from real data, but setting the autocorrelation of income to zero. Effectively this makes our model a static model, as the expected value in the Bellman equation (6) becomes independent of the current state of migration incentives. Table 4 reports the estimation results from this experiment. Migration cost estimates are with US\$ 57,713 substantially higher in this setting than in the setting with autocorrelated income. Moreover, the model is no longer able to match the observed data moments, so that the model specification test now rejects our model. Compared to the estimates reported in the literature, estimated migration costs are still on the low side. However, two aspects need to be taken into account: First, with assumed zero autocorrelation of gains from migration, the net present value of these gains is very limited. In fact, the (naive) net-present value gain from a once and for all location decision is

Table 4: Simulated moments estimation: Estimation results for zero autocorrelation

Parameter	Risk-Neutrality, No Autocorrelation	Risk-Neutrality, Autocorr. = 0.95
Migration Costs	57625 ¹ (3601)	18,285 ¹ (2, 211)
Correlation of Income Shocks Across Regions Ψ	0.1494 (0.1349)	0.2482 (0.1947)
Fraction of Income Shock due to Aggregate Fluctuations ϕ	0.0001 (0.0007)	0.0041 (0.0011)
Standard Deviation of Transitory Income Shock σ_φ	0.0283 (0.0017)	0.0266 (0.0012)
p-value, $\chi^2(2)$ overident. test	0.0000	0.1865

¹Migration cost estimate \hat{c} in US\$ terms.

Standard errors in parenthesis. Estimation is carried out using the simulated moments estimator by Gourieroux, Monfort, and Renault (1993), which chooses structural model parameters by matching the moments from a simulated panel of regions. See Table 1 for further details. The simulations generate a panel of 51 region-pairs and a 81-year history of migration and income data. The first 55 years of simulated data are dropped in order to minimize the influence of initial values. Each simulation is repeated 5 times and data moments are compared to the average over the 5 replications of the simulation.

about 10 times smaller with no autocorrelation than in our baseline specification. Second, we estimate substantial measurement error in the no-autocorrelation specification, which dampens the increase in estimated migration costs. Under zero autocorrelation, the estimated measurement error is with a standard deviation of 0.0283 substantially larger, making up 96% of the sum of the long-run variances of *potential* income and measurement error..

6.2.2 Random static utility model

To corroborate our conclusion in an alternative way, we simulate our model to generate a data set that comprises 51 pairs of regions and 16 years of data. The parameters

of the model are fixed to their estimated values reported in the column "All Moments Matched" in Table 2. Then we estimate a static random utility model of migration from this data.

In particular, we apply a conditional-probit approach similar to Davies, Greenwood, and Li's (2001) conditional logit model to describe the migration decision. Simplifying Davies, Greenwood, and Li's (2001) model and adapting it to our bi-regional framework with normally distributed errors, the likelihood of the conditional probit model becomes

$$\ln \mathcal{L} = \sum_t \sum_{i=1,2} P_{it} \left[\begin{array}{l} \bar{\Lambda}_{it} \ln [\Phi(\gamma(\exp \bar{w}_{it} - \exp \bar{w}_{-it}) - c)] \\ + (1 - \bar{\Lambda}_{it}) \ln [1 - \Phi(\gamma(\exp \bar{w}_{it} - \exp \bar{w}_{-it}) - c)] \end{array} \right], \quad (12)$$

where Φ is the cumulative distribution function of a normal distribution.

While Davies, Greenwood, and Li (2001) include a set of other variables to describe the utility gains from location choice, our simulated model just allows for log income as an explanatory variable. This means that the form of the likelihood function in (12) assumes that utility is composed of an income component (with sensitivity $\gamma > 0$) and a disutility from migration $c > 0$. The estimated money measure of this disutility is $\frac{\hat{c}}{\hat{\gamma}}$.¹⁴

Since our argument is not based on the classical measurement error that we additionally took into account in our estimation, i.e. migration incentives being only imperfectly measured by income, we run two versions of the static random utility experiment. In the first version, we set all parameters to their estimated values except for the variance of the measurement error, which is set to zero instead. In the second version, we also include the measurement error into the estimation. This allows us to decompose the total bias into two sources, classical measurement error and dynamic self-selection.

Throughout, the migration costs taken from our estimation are US\$ 18,285. For the version without measurement error the conditional-logit estimation suggests a cost of US\$ 40,576, a number that is substantially higher than the costs used in the simulation. In comparison, 'true' and estimated costs correspond to 0.3 and 0.77 average annual incomes, respectively. This comparative exercise shows that the estimation of structural parameters is likely to be subject to a bias if the unobserved dynamics of the distribution of incentives is not taken into account. Adding measurement errors to the simulated income data drives the cost estimate up to US\$ 99,457. These numbers suggest that while the bias from self-selection is not exactly as important as the bias from classical measurement error, still it is substantial.

¹⁴Since our model is composed of two regions only, we cannot estimate γ and c from a cross-section as Davies, Greenwood, and Li (2001) do, but have to pool the simulated data instead.

Table 5: Simulation results: comparison to cost estimate based on a static random utility model

	Simulation	
	without classical measurement error	with error
Migration Cost (from estimation, used for simulation)	18,285	18,285
Average Annual Income	53,060	53,041
Migration Cost Estimate Based on a Static Random Utility Model		
- in terms of annual incomes	76.5%	187.4%
- in US \$	40,576	99,457

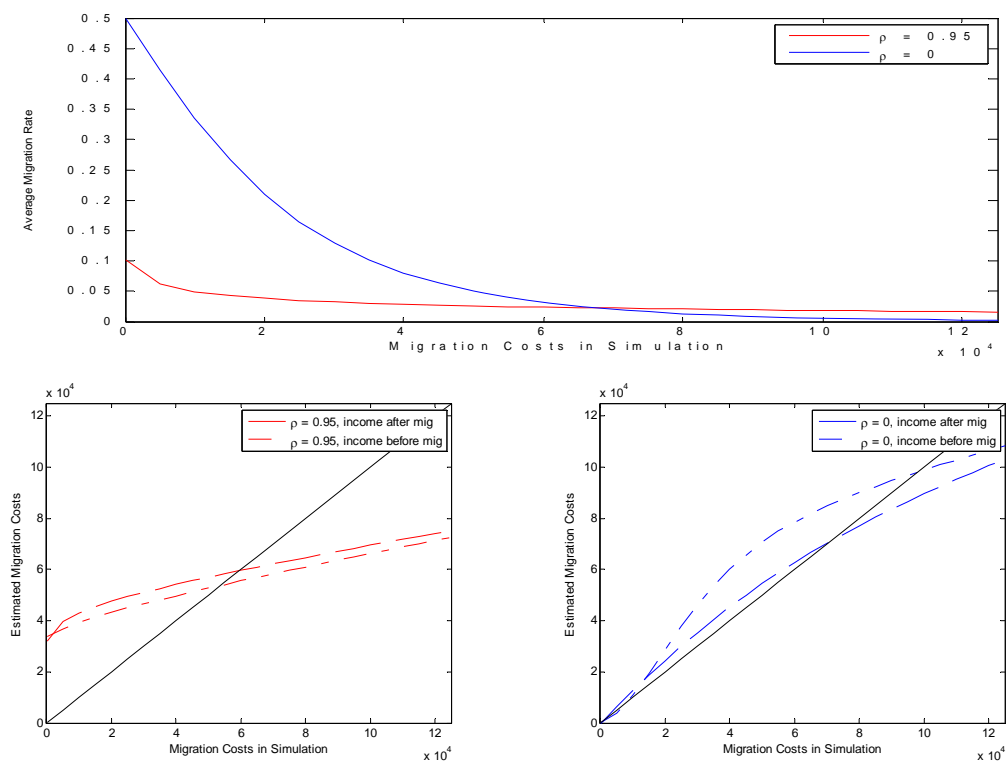
Figure 2 displays the cost estimates from the static random utility model as a function of costs used in the simulation. There are two sources of bias linked to the static random utility model in this setup. One is the dynamic self-selection bias we highlighted, the other is a bias from misspecification as the static model assumes no autocorrelation in income differences. The latter bias stems from the fact that the static model considers current incomes instead of net present values of income streams; it disregards future income gains as a result of location choice.

We see that the bias from self-selection becomes more important, the lower the true costs of migration are. If migration costs are very high, there is simply little migration each year and hence not much self-selection. Instead, the mis-specification (downwards) bias of a static model dominates the bias from self-selection.

For comparison we plot static cost estimates from a simulation of our model with autocorrelation ρ set to zero. In this setup the static random utility model is correctly specified and leads to almost unbiased estimates if we assume income to be observed before migration.¹⁵

¹⁵At very high costs the static random utility model ceases to estimate the costs used in the simulation correctly, but this is most likely the case because the approximation quality of the income process becomes weak in the then relevant domain of income fluctuations. In fact, migration rates are already very low when the static model ceases to estimate the cost parameter well, see Figure 2.

Figure 2: Comparison of cost estimate based on a static random utility model



6.2.3 Estimation with fixed distribution of migration incentives

As a final experiment, we estimate the model with autocorrelated income, but use the unconditional distribution of migration incentives to evaluate migration rates. This estimation therefore ignores dynamic self-selection, but leaves the parameters of the problem unchanged otherwise. For example, from the micro perspective, the decision problem is still dynamic and, naive net-present values of location choices remain unaltered.

Though this model is inconsistent, one could be persuaded by the logic that migration rates are small each year, making it innocent to approximate the conditional incentive distribution (conditional on the place of residence) in (9) by the unconditional distribution. Table 6 reports the estimation results. The parameter estimates from the ‘consistent’ model with dynamic self-selection are repeated for comparison.

Neglecting self-selection seems all but harmless. The point estimates of all model parameters change substantially. Most importantly—and in line with our argument—the estimated migration costs are with US\$ 363,300 substantially larger. The bias in this last experiment is largest since there is no offsetting downwards bias from a misrepresentation of the income process, i.e. of implied expected income streams.

Hence, treating migration as a dynamic decision problem at the micro level without taking care of dynamic self-selection in the aggregation may lead to a more severe bias than ignoring the dynamic structure of the migration decision altogether.

6.3 Implied micro characteristics & mover patterns

So far, we focused exclusively on the estimation of migration costs and the fact that dynamic self-selection plays an important role for this estimation. However, current contributions to the empirical study of migration behavior go beyond the representative agent assumption of our homogeneous migration cost model (with heterogeneous incentives). They argue on the basis that individual migration behavior displays a significant autocorrelation that there should be a mover-stayer dichotomy in the population. Some agents would then have permanently lower migration costs than others due to some deeper psychological factors.

However, inferring serially correlated migration costs from the observation of serially correlated migration behavior may be hasty. High serial correlation of migration incentives, implies on the one hand that those agents which have a strong incentive to remain in state j in period t most likely keep this strong incentive also in period $t + 1$. On the other hand, it also implies that those agents which have recently migrated can only be mildly in favor of their new place of residence, otherwise they had moved earlier—at least if migration costs are not too large. These considerations as a whole suggest that

Table 6: Simulated moments estimation: Model ignoring dynamic self selection

Parameter	Ignoring Dyn. Self-Selection	With Dynamic Self-Selection
Migration Costs	363,300 ¹ (9750)	18,285 ¹ (2, 211)
Correlation of Income Shocks Across Regions Ψ	0.40524 (0.0367)	0.2482 (0.1947)
Fraction of Income Shock due to Aggregate Fluctuations ϕ	0.014 (0.0015)	0.0041 (0.0011)
Standard Deviation of Transitory Income Shock σ_φ	0.0000 (0.0891)	0.0266 (0.0012)
p-value, $\chi^2(2)$ overident. test	0.0000	0.1865

¹Migration cost estimate \hat{c} in US\$ terms.

Standard errors in parenthesis. Estimation is carried out using the simulated moments estimator by Gourieroux, Monfort, and Renault (1993), which chooses structural model parameters by matching the moments from a simulated panel of regions. See Table 1 for further details. The simulations generate a panel of 51 region-pairs and a 81-year history of migration and income data. The first 55 years of simulated data are dropped in order to minimize the influence of initial values. Each simulation is repeated 5 times and data moments are compared to the average over the 5 replications of the simulation.

Table 7: Comparison of Dynamic incentives model and Binomial model of migration: Mover-Stayer Heterogeneity

Migration Characteristics		Stayers	Single Movers	2 or More Moves	3 or More Moves
Dynamic Model (D)	as % of population	47.7%	32.8%	19.5%	5.1%
	as % of movers		62.7%	37.3%	9.8%
Binomial Model (B)	as % of population	44.5%	36.7%	18.7%	4.3%
	as % of movers		66.2%	33.8%	7.8%
D relative to B	as % of population	1.078	0.894	1.043	1.186
	as % of movers		0.946	1.105	1.256

Frequencies of migration behavior over a 20 period interval. The frequencies are obtained by simulating the behavior of a cross-section of 50000 households for 70 years, using the migration behavior of the households over the time interval $[t + 1, t + 20]$, $20 < t \leq 50$. Reported frequencies are obtained by averaging over all overlapping subintervals and 5 repeated simulations.

our model should be able to reproduce at least some "mover-stayer heterogeneity" in migration data, even though agents are homogeneous with respect to their (psychic) costs of migration.

Table 7 provides evidence on this. We simulate a sample of 50000 households over 70 years and track their migration behavior. We calculate migration probabilities over the $[t + 1, t + 20]$ year interval for those who migrated in year t and those who did not move in year t , for $20 < t \leq 50$. We compare the migration probabilities from our model to the migration probabilities generated from a binomial model of migration that assumes a constant migration probability equal to the average migration rate in the sample.¹⁶ Table 7 displays the probability of a household to never migrate ("Stayer"), migrate exactly once ("Single mover"), migrate at least twice, and migrate at least three times over the 20 year period. As one can see, substantially more households than predicted by

¹⁶As a side result of this experiment, we see that average migration rates (when starting with the ergodic distribution of potential income) reach their normal levels quickly, after about 3 periods. Consequently, we can expect our results to carry over to a finite time framework as well, which is technically much more difficult to analyze though, since income distributions for each age group of agents would need to be simulated.

the binomial model do not move. At the other extreme, more households than predicted by the binomial model move more than two or three times. Observing only the data in Table 7, one might conjecture that our model exhibits a "mover-stayer heterogeneity", which is not the case though. Yet, compared to the micro evidence provided for example by Kennan and Walker (2006), our model—and hence dynamic self-selection—is unlikely to explain all of the mover-stayer heterogeneity documented.

7 Conclusion

We have provided a tractable model of aggregate migration with a sound microeconomic foundation. The paper is a contribution to the recently evolving literature on structural models of migration. We explicitly deal with the problem that potential gains from migration are unobservable and display a dynamic character. This dynamic character of migration incentives has two aspects: First, the individual gains from migration evolve stochastically over time, but will typically be highly persistent. Second, at an aggregate level, the distribution of migration incentives is a result of past migration decisions themselves.

Starting from the microeconomic decision problem allows us to keep track of the dynamics of the incentive distribution. This dynamics is driven by (dynamic) self-selection. Neglecting this self-selection results in biased estimates of structural parameters, such as migration costs. In our application to US interstate migration, we estimate migration costs to be substantially lower than those reported in previous studies. The estimated migration costs amount to about US\$ 18,500, which corresponds to less than one-half of an average annual income.

Our analysis calls for a careful treatment of the self-selection problem when economic incentives are not fully observable. What makes this issue particularly relevant for the analysis of migration is that unobservable incentives are highly autocorrelated though not perfectly persistent. Rather than being drawn every period anew, migration incentives have a long memory. One example of this long memory of migration incentives is the persistency that income displays. We integrated the persistency of unobserved migration incentives in a structural dynamic microeconomic model of the migration decision. This consequently allowed us to simulate the joint behavior of the observed migration rates, of the unobserved migration incentives, and of their observable proxies, i.e. incomes.

Addressing the partial unobservability of migration incentives may not only be important to macro-studies of migration. Even at a micro level, potential incomes are typically unobservable and have to be proxied. However, such approximation regularly neglects self-selection. If households live in their preferred place of residence as a result

of their location choice, and if all observable things are equal, then it must be the unobserved component of their preferences that is in favor of the place where they actually live in. Besides unobservable parts of income, this unobservable component of preferences can also comprise different valuations of different amenities and social networks. Even these factors can be expected to exhibit persistency.

Future research calls for a more complex microeconomic model that integrates more information into the macroeconomic analysis, for example labor market conditions and amenities. Additionally, it would be desirable to extend our bi-regional approach to the case of multiple regions, as in Davies, Greenwood and Li (2001), and Kennan and Walker (2006). Further aspects, such as the interaction of migration and local labor markets, could be analyzed in a general equilibrium framework as in Coen-Pirani (2006), but our results call for an explicit treatment of the dynamic structure and persistency of migration incentives. However, all this goes beyond what is currently numerically feasible, in particular if the model is meant to be estimated. Beyond the application to migration decisions, our treatment of the dynamic self-selection problem may also be applicable to other important discrete choices in an economy, for example labor-market participation or product switching.

8 Appendix

8.1 Existence and uniqueness of the value function

We begin with proving existence and uniqueness of the value function. Notation is as in the main text throughout this appendix, unless stated otherwise.

For the ease of exposition, we assume that the income process is only approximately log-normal. In particular, we assume that income has a finite support.

Definition 1 Let $\mathbb{W} = [\underline{W}, \overline{W}]$ be the support of w .

Definition 2 Define a mapping T according to the migration problem of a household, that is

$$T(u)(k, w_{iAt}, w_{iBt}) = \max_{j=A,B} \left\{ \exp(w_{ijt}) - \mathbb{I}_{\{k \neq j\}}c + \beta \mathbb{E}_t u(j, w_{iAt+1}, w_{iBt+1}) \right\}. \quad (13)$$

The mapping T is defined on the set of all real-valued, bounded functions \mathbb{B} that are continuous with respect to $w_{A,B}$ and have domain $D = \{A, B\} \times \mathbb{W}^2$.

Lemma 3 The mapping T preserves boundedness.

Proof. To show that T preserves boundedness one has to show that for any bounded function u also Tu is bounded. Consider u to be bounded from above by \bar{u} and bounded from below by \underline{u} . Then, Tu is bounded, because

$$Tu = \max_{j=A,B} \left\{ \exp(w_{ijt}) - \mathbb{I}_{\{k \neq j\}}c + \beta \mathbb{E}_t u(j, w_{iAt+1}, w_{iBt+1}) \right\} \leq \exp(\overline{W}) + \beta \bar{u} < \infty, \quad (14)$$

and

$$Tu = \max_{j=A,B} \left\{ \exp(w_{ijt}) - \mathbb{I}_{\{k \neq j\}}c + \beta \mathbb{E}_t u(j, w_{iAt+1}, w_{iBt+1}) \right\} \quad (15)$$

$$\geq \max_{j=A,B} \left\{ \exp(w_{ijt}) - \mathbb{I}_{\{k \neq j\}}c + \beta \underline{u} \right\} \geq \exp(\underline{W}) + \beta \underline{u} > -\infty. \quad (16)$$

■

Lemma 4 The mapping T preserves continuity.

Proof. Since Tu is the maximum of two continuous functions, it is itself continuous.

■

Lemma 5 The mapping T satisfies Blackwell's conditions.

Proof. First, we need to show that for any $u_1(\cdot) < u_2(\cdot)$ the mapping T preserves the inequality. Since both the expectations operator and the max operator preserve the

inequality, also T does so. Second, we need to show that $T(u + a) \leq Tu + \gamma a$ for any constant a and some $\gamma < 1$. Straightforward algebra shows that

$$T(u + a) = Tu + \beta a. \quad (17)$$

Since $\beta < 1$ by assumption, T satisfies Blackwell's conditions. ■

Proposition 6 *The mapping T has a unique fixed point on \mathbb{B} , and hence the Bellman-equation has a unique solution.*

Proof. *Follows straightforwardly from the last three Lemmas.* ■

8.2 Effect of Income Shocks on the Distribution of Income

Idiosyncratic shocks, aggregate shocks, and the persistency of the income process determine the transition of the distribution of income incentives after migration to the distribution of migration incentives before migration in the next period. The income distribution at the beginning of period $t + 1$, F_{t+1} , results from adding idiosyncratic and aggregate shocks to the distribution of income after migration in period t , \hat{F}_t , of which $\hat{f}_{jt}(w_A, w_B)$ is the conditional density, see (11). This means that a household having an income w_{ijt+1} in period $t + 1$ can result from any possible combination of w_{ijt} and $\xi_{ijt+1} = \theta_{jt+1} + \omega_{ijt+1}$ for which

$$w_{ijt+1} = \mu_j(1 - \rho) + \rho w_{ijt} + \theta_{jt+1} + \omega_{ijt+1} \quad (18)$$

holds. Solving this equation for w_{ijt} , we obtain

$$w_j^*(w_{ijt+1}, \theta_{jt+1}, \omega_{ijt+1}) := w_{ijt} = \frac{w_{ijt+1} - (\theta_{jt+1} + \omega_{ijt+1})}{\rho} - \mu_j \frac{(1 - \rho)}{\rho}. \quad (19)$$

This $w_j^*(w_{ijt+1}, \theta_{jt+1}, \omega_{ijt+1})$ is the time- t potential income in region j that is consistent with a future potential income of w_{ijt+1} and realizations of shocks $\theta_{jt+1} + \omega_{ijt+1}$ at the beginning of period $t + 1$. Now suppose that both kinds of shocks, θ and ω , have been realized. Then, $w_{A,B}^*$ is a one-to-one mapping of future incomes (w_{iAt+1}, w_{iBt+1}) to current income (w_{iAt}, w_{iBt}) .

The conditional density of observing the future income pair (w_{iAt+1}, w_{iBt+1}) can thus be obtained from a retrospective. The income pair (w_A^*, w_B^*) of past incomes corresponds uniquely to a future income pair (w_{iAt+1}, w_{iBt+1}) . Consequently, we can express the density of the income distribution at time $t + 1$ using the income distribution after

migration \hat{F}_t , and its conditional density \hat{f}_{jt} . The density of the income distribution F_{t+1} conditional on the region and the vector of shocks is given by

$$\begin{aligned} f_{jt+1}(w_A, w_B | \theta_{At+1}, \theta_{Bt+1}, \omega_{iAt+1}, \omega_{iBt+1}) \\ = \hat{f}_{jt}(w_A^*(w_A, \theta_{At+1}, \omega_{iAt+1}), w_B^*(w_B, \theta_{Bt+1}, \omega_{iBt+1})). \end{aligned} \quad (20)$$

Weighting this density with the density of the idiosyncratic shocks $h(\omega_{iAt+1}, \omega_{iBt+1})$ yields the density of observing the future income pair (w_A^*, w_B^*) together with the idiosyncratic shocks $(\omega_{iAt+1}, \omega_{iBt+1})$:

$$\hat{f}_{jt}(w_A^*(w_A, \theta_{At+1}, \omega_{iAt+1}), w_B^*(w_B, \theta_{Bt+1}, \omega_{iBt+1})) \cdot h(\omega_{iAt+1}, \omega_{iBt+1}).$$

Integrating over all possible idiosyncratic shocks $(\omega_{iAt+1}, \omega_{iBt+1})$ yields the density f_{jt+1} of the income distribution before migration in period $t+1$ for a certain combination of aggregate shocks $(\theta_{At+1}, \theta_{Bt+1})$:

$$\begin{aligned} f_{jt+1}(w_A, w_B | \theta_{At+1}, \theta_{Bt+1}) = \\ \int \hat{f}_{jt}(w_A^*(w_A, \theta_{At+1}, \omega_A), w_B^*(w_B, \theta_{Bt+1}, \omega_B)) \cdot h(\omega_A, \omega_B) d\omega_A d\omega_B, \quad j = A, B. \end{aligned} \quad (21)$$

For given aggregate shocks, this new distribution determines migration from region j to region $-j$ according to equation (9) for period $t+1$.

The evolution of income distributions can thus be summarized as follows. Between two consecutive periods, the conditional distribution of potential incomes first evolves as a result of migration decisions, moving the density from f_{jt} to \hat{f}_{jt} . Thereafter, the distribution is again altered by aggregate and idiosyncratic shocks to income, moving the density from \hat{f}_{jt} to f_{jt+1} . The latter density now determines migration decisions in period $t+1$, starting the cycle over again. In other words, migration incentives are not only a result of past income shocks, but also a result of past migration decisions.

8.3 Invariant distribution

We prove that migration decisions and idiosyncratic shocks to income imply that potential income follows an ergodic Markov-process if there are no aggregate shocks. Therefore, there is an invariant distribution the sequence of income distributions converges to. For simplicity, we present the proof for an arbitrary discrete approximation of the model with a continuous state-space for income.

Lemma 7 *Assume an arbitrary discretization of the state space with n points for the*

potential income in each of the regions. Then, we can capture the transition from f_t to f_{t+1} , which are the unconditional densities of the distribution of households over both regions and potential incomes, in a matrix $\Gamma = \begin{pmatrix} \Pi(I - D_A) & \Pi D_B \\ \Pi D_A & \Pi(I - D_B) \end{pmatrix} \in R^{2n^2 \times 2n^2}$.¹⁷ In this matrix, Π denotes the transition matrix that approximates the AR(1)-process for income by a Markov-chain, see Adda and Cooper (2003, pp. 56) for details. Matrix D_j , $j = A, B$ is the $n^2 \times n^2$ diagonal matrix with the migration hazard rates for each of the n^2 income pairs of the income grid.

Proof. First, we take a discrete state-space of n possible incomes for each region, $w_{A1} \dots w_{An}$ and $w_{B1} \dots w_{Bn}$. Second, we denote the vector of probabilities that describes the distribution of potential incomes and household locations in the following form

$$f = \left(f(A, w_{A1}, w_{B1}) \dots f(A, w_{An}, w_{B1}) \dots f(A, w_{An}, w_{Bn}) f(B, w_{A1}, w_{B1}) \dots f(B, w_{An}, w_{Bn}) \right)'. \quad (22)$$

Analogously, we define the distribution after migration but before idiosyncratic shocks, \hat{f} . Taking our law of motion from (21), we obtain as a discretized analog

$$f_{t+1} = \begin{pmatrix} \Pi & 0 \\ 0 & \Pi \end{pmatrix} \hat{f}_t. \quad (23)$$

Now, define $d_h \in \{0, 1\}$ as the fraction of households that migrate and are in the h -th income and location triple given our vectorization of the income grid. This means that $d_h = \Lambda_j(w_{Ak}, w_{Bl})$, $h = 1 \dots 2n^2$, where (j, w_{Ak}, w_{Bl}) is the h -th element in the vectorized grid. Moreover, define $D = \text{diag}(d)$ as the diagonal matrix with migration rates on the diagonal and D_A and D_B as the diagonal matrices with only the first n^2 and the last n^2 elements of d , respectively. Then, we can describe the transition from f_t to \hat{f}_t by

$$\hat{f}_t = \begin{pmatrix} I - D_A & D_B \\ D_A & I - D_B \end{pmatrix} f_t \quad (24)$$

Combining the last two equations, we obtain

$$f_{t+1} = \begin{pmatrix} \Pi(I - D_A) & \Pi D_B \\ \Pi D_A & \Pi(I - D_B) \end{pmatrix} f_t. \quad (25)$$

■

¹⁷Since we work with a discretization, strictly speaking f is not a density, but a vector of probabilities for drawing a location-income possibility vector from a given element of the grid.

Lemma 8 For any distribution of idiosyncratic shocks with support equal to \mathbb{W}^2 , matrix Π has only strictly positive entries.

Proof. If the idiosyncratic shocks have support equal to \mathbb{W}^2 , then every pair of potential incomes can be reached from every other pair of incomes as a result of the shock, because we assume the shocks to income to be approximately log-normal. Thus, all entries of Π are strictly positive. ■

Lemma 9 Γ^2 has only positive entries.

Proof. Due to $\bar{c}_A = -\bar{c}_B$, we can assume an ordering of states such that we can write $D_A = \begin{pmatrix} I_{n_a} & 0 \\ 0 & 0 \end{pmatrix}$ and $D_B = \begin{pmatrix} 0 & 0 \\ 0 & I_{n_b} \end{pmatrix}$, without loss of generality, where I_z is a $z \times z$ unit matrix. Accordingly, we define partitions of Π such that

$$\begin{aligned} \Pi &= \begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix} = \begin{pmatrix} B_1 & B_2 \\ B_3 & B_4 \end{pmatrix} \\ &= \begin{pmatrix} C_1 & C_2 \\ C_3 & C_4 \end{pmatrix} = \begin{pmatrix} D_1 & D_2 \\ D_3 & D_4 \end{pmatrix}, \end{aligned}$$

where $A_1 \in R^{(n^2-n_a) \times (n^2-n_a)}$, $B_1 \in R^{n_b \times n_b}$, $C_1 \in R^{n_a \times n_a}$, $D_1 \in R^{(n^2-n_b) \times (n^2-n_b)}$.

This yields for Γ^2 after some tedious algebra

$$\Gamma^2 = \begin{pmatrix} B_2 C_3 & A_2 A_4 & B_2 C_4 & A_2 B_4 \\ B_4 C_3 & A_4 A_4 & B_4 C_4 & A_4 B_4 \\ D_1 C_1 & C_1 A_2 & D_1 D_1 & C_1 B_2 \\ D_3 C_1 & C_3 A_2 & D_3 D_1 & C_3 B_2 \end{pmatrix}.$$

Each entry of this matrix is positive, since Π and hence its partitions are positive.

Proposition 10 Under the assumptions of the above Lemmas, migration and idiosyncratic shocks define an ergodic process with a stationary distribution $F_0 = \lim_{n \rightarrow \infty} B^n e_i$.

Proof. The above Lemma directly implies the ergodicity of the Markov chain. ■

■

8.4 Data

Data on migration between US federal states is provided by the US Internal Revenue Service (IRS). The IRS uses individual income tax returns to calculate internal migration flows between US states. In particular, the IRS compiles migration data by matching

Table 8: Descriptive statistics

	Mean	Std. Dev.	Min	Max
Migration Rate	.0393	.0178	.0144	.1146
Migration Rate Filtered	.0393	.0036	.0234	.0629
Income per Capita	10.71	.1644	10.39	11.27
Income per Capita Filtered	10.71	.0312	10.61	10.82
Complementary Income per Capita	10.76	.0624	10.67	10.86
Complementary Income per Capita Filtered	10.76	.0239	10.73	10.80

the Social Security number of the primary taxpayer from one year to the next. The IRS identifies households with an address change since the previous year, and then totals migration to and from each state in the US to every other state. Given these bilateral migration flows, we compute aggregate gross immigration for the 50 US states and the District of Columbia as the sum of all immigrations from other US states to a particular state. Migration rates are calculated by expressing gross immigration as proportions of the number of non-migrants reported in the IRS dataset. The IRS state-to-state migration-flow data is available for the years 1989 - 2004.

Income per capita data is taken from the Regional Economic Information System (REIS) compiled by the Bureau of Economic Analysis (BEA). We use personal income as the relevant income concept. The REIS data is available online at www.bea.gov/bea/regional/reis/. The income-per-capita figure for the alternative region is computed as the population-weighted mean of all per capita incomes outside a specific state.

We remove a linear time trend from all data and express all variables as deviations from their unit-specific means (re-scaled by their overall mean), i.e. we apply a within-transformation. Table 8 reports descriptive statistics for the original as well as for the transformed data.

In order to examine the time-series properties of the data employed, we perform a

unit-root analysis for migration rates and income data. In a sample of size $T = 16$ and $N = 51$ either a Breitung and Meyer (1994) or a Levin, Lin, and Chu (2002) unit-root test is most appropriate. For the Breitung and Meyer (1994) test, we determined the optimal augmentation lag length by sequential t -testing. Taking into account three augmentation lags and time-specific effects, we can reject the null hypothesis of a unit root at the 5% level of significance. Similarly, the Levin, Lin, and Chu (2003) test rejects the null hypothesis of a unit root taking a linear time trend into account.

8.5 Numerical aspects

The first step in solving the model numerically is to obtain a solution to (6). We do so by value-function iteration.¹⁸ For this value-function iteration, we first approximate the bivariate process of potential incomes for an individual agent in regions A and B

$$w_{ijt} = \mu_j(1 - \rho) + \rho w_{ijt-1} + \xi_{ijt} \quad (26)$$

by a Markov chain. Because w_A and w_B are correlated through the correlation structure in ξ , it is easier to work with the orthogonal components (w_A^+, w_B^+) of (w_A, w_B) in the value function iteration.

We evaluate the value function on an equi-spaced grid for the orthogonal components with a width of $\pm 3.5\sigma_{A,B}^+$ around their means, where $\sigma_{A,B}^+$ denote the long-run standard deviations of the orthogonal components. The grid is chosen to capture almost all movements of the income distribution F later on.¹⁹ Given this grid, we use Tauchen's (1986) algorithm to obtain the transition probabilities for the Markov-chain approximation of the income process in (26).

We apply a multigrid algorithm (see Chow and Tsitsiklis, 1991) to speed up the calculation of the value function. This algorithm works iteratively. It first solves the dynamic programming problem for a coarse grid and then doubles the number of grid points in each iteration until the grid is fine enough. In between iterations the solution for the coarser grid is used to generate the initial guess for the value-function iteration on the new grid by spline interpolation. The initial grid has 16×16 points (income A \times income B) and the final grid has 128×128 points.

¹⁸See for example Adda and Cooper (2003) for an overview of dynamic programming techniques.

¹⁹The choice of $\pm 3.5\sigma_{A,B}^+$ is motivated as follows. We obtain in the estimation that about 99% of the income shocks is due to the idiosyncratic component. Therefore, we can expect 99.9% of the mass of the income distribution to fall within $\pm 3.29 \cdot \sqrt{0.99}\sigma_{A,B}^+ \cong \pm 3.27\sigma_{A,B}^+$ around the mean of the distribution for any given year. Additionally, the mean income for each year moves within the band $\pm 3.29 \cdot \sqrt{0.01}\sigma_{A,B}^+ \cong \pm 0.33\sigma_{A,B}^+$ in again 99.9% of all years. Since the sum of both components is $\pm 3.6\sigma_{A,B}^+$, a grid variation of $\pm 3.5\sigma_{A,B}^+$ should not truncate the income distribution.

The solution of (6) yields the optimal migration policy and thus the microeconomic migration hazard rates Λ_j . With these hazard rates, we can obtain a series of aggregate migration rates for a simulated economy as described in detail in Section 4.2 for any realization of aggregate shocks $(\theta_{jt})_{t=1\dots T}^{j=A,B}$ and an initial distribution F_0 .

This means that we need an initial distribution of income F_0 to solve the sequential problem. Following Caballero and Engel's (1999) suggestion, we use the ergodic distribution of income \bar{F} that would be obtained in the absence of aggregate income shocks.²⁰ This ergodic distribution \bar{F} is a natural starting guess for F_0 as Caballero and Engel (1999) argue.

To simulate a series of migration rates which correspond to the aggregate migration hazards $(\bar{\Lambda}_{At,Bt})_{t=1\dots T}$, we draw a series of aggregate shocks (to the orthogonal basis) $(\theta_{At}^+, \theta_{Bt}^+)_{t=1\dots T}$ from a normal distribution with variance $\phi \cdot (\sigma_{A,B}^+)^2$, $\phi \in [0, 1]$. The weight ϕ measures the relative importance of aggregate shocks, relative to idiosyncratic shocks, i.e. $\sigma_\omega^2 = (1 - \phi) \sigma_\xi^2$ and $\sigma_\theta^2 = \phi \sigma_\xi^2$. Correspondingly, the orthogonal components of the idiosyncratic shocks have variance $(1 - \phi) \cdot (\sigma_{A,B}^+)^2$.

As we did to approximate the microeconomic income process for value function iteration, we also discretize the distribution of migration incentives over the chosen grid of income to simulate its evolution. Accordingly, we replace the conditional density in (9) by discrete probabilities. This means that for grid points \bar{x}_{jk} , $j = A, B$; $k = 1\dots 64$, (k being the index of grid points) with a distance of $2h$ in between points, we calculate the probabilities initially (for $t = 0$) as

$$\bar{p}_{k,0} = \int_{\bar{x}_{A,k-h}}^{\bar{x}_{A,k+h}} \int_{\bar{x}_{B,k-h}}^{\bar{x}_{B,k+h}} f_0(x_1, x_2) dx_1 dx_2.$$

An aggregate shock θ_j now implies that the off-grid pair $(\bar{x}_{A,k} + \theta_A, \bar{x}_{B,k} + \theta_B)$ occurs with probability $\bar{p}_{k,0}$. To re-obtain on-grid probabilities, we use spline interpolation methods to find $p_{k,1}$, restricting p to take values between 0 and 1. That is, for each t we define a function z with $z(\bar{x}_{A,k} + \theta_{A,t+1}, \bar{x}_{B,k} + \theta_{B,t+1}) := \bar{p}_{k,t}$ and obtain $\bar{p}_{k,t+1}$ as

$$\bar{p}_{k,t+1} = \hat{z}(\bar{x}_{A,k}, \bar{x}_{B,k})$$

where \hat{z} is the interpolation of z . Idiosyncratic shocks are accounted for by multiplying after-aggregate-shock, on-grid income probabilities with the transition probability matrix obtained from Tauchen's algorithm. The effect of migration on the distribution of

²⁰This distribution is calculated by assuming that idiosyncratic shocks ω have the full variance of ξ .

migration incentives is captured by using a discretized version of (11).

We calculate aggregate migration rates this way, i.e. by directly simulating the evolution of the incentive distribution instead of using a Monte Carlo method based on drawing a sample of agents, for the reason that the latter is not adequate in our case. We focus on aggregate behavior, but aggregate shocks turn out to be relatively small (being responsible for less than 1% of the total variation in income, see the discussion in Section 5.2). Hence sampling variation would exceed the true aggregate variation of income most likely if we applied a Monte Carlo approximation.

References

- [1] Adda, J. and R. Cooper (2003): "Dynamic Economics: Quantitative Methods and Applications", MIT Press, Cambridge.
- [2] Blanchard, O. and L. Katz (1992): "Regional evolutions", *Brookings Papers on Economic Activity*, 1, 1-75.
- [3] Breitung, J. and W. Meyer (1994): "Testing for unit roots in panel data: are wages on different bargaining levels cointegrated?", *Applied Economics*, 26, 353-361.
- [4] Burda, M. (1993): "The Determinants of East-West German Migration: Some First Results", *European Economic Review*, 37, 452-462.
- [5] Burda, M., M. Müller, W. Härdle, and A. Werwatz (1998): "Semiparametric Analysis of German East-West Migration Intentions: Facts and Theory", *Journal of Applied Econometrics*, 13, 525-541.
- [6] Borjas, G. J. (1987): "Self-selection and the earnings of immigrants", *American Economic Review*, 77, 531-553.
- [7] Borjas, G. J., S. Bronars, and S. Trejo (1992): "Self-selection and internal migration in the United States", *Journal of Urban Economics*, 32, 159-185.
- [8] Caballero, R. and E. M. R. A. Engel (1999): "Explaining Investment Dynamics in U.S. Manufacturing: A Generalized (S,s) Approach", *Econometrica*, 67, 783-826.
- [9] Chow, C. S. and J. N. Tsitsiklis (1991): "An optimal multigrid algorithm for continuous state discrete time stochastic control", *IEEE Transactions on Automatic Control*, 36, 898-914.
- [10] Cushing, B. and J. Poot (2004): "Crossing boundaries and borders: Regional science advances in migration modelling", *Papers in Regional Science*, 83, 317-338.
- [11] Coen-Pirani, D. (2006): "Understanding Gross Workers Flows Across U.S. States", mimeo, Carnegie Mellon University.
- [12] Davies, P. S., M. J. Greenwood, and H. Li (2001): "A Conditional Logit Approach to U.S. State-to-State Migration", *Journal of Regional Science*, 41, 337-360.

- [13] Decressin, J. and A. Fatas (1995): "Regional Labor Market Dynamics in Europe", *European Economic Review*, 39, 1627-1655.
- [14] Gourieroux, C., A. Monfort, and E. Renault (1993): "Indirect inference", *Journal of Applied Econometrics*, 8, S85-118.
- [15] Greenwood, M. J. (1975): "Research on internal migration in the United States: A survey", *Journal of Economic Literature*, 13, 397-433.
- [16] Greenwood, M. J. (1985): "Human migration: Theory, models and empirical studies", *Journal of Regional Science*, 25, 521-544.
- [17] Greenwood, M. J. (1997): "Internal migration in developed countries", in: Rosenzweig, M. R. and O. Stark (eds.), *Handbook of population and family economics*, Volume 1B, Elsevier, Amsterdam.
- [18] Hassler, J., J. V. R. Mora, K. Storesletten, and F. Zilibotti (2005). "A Positive Theory Of Geographic Mobility And Social Insurance", *International Economic Review*, 46, 263-303.
- [19] Hunt, G. L. and R. E. Mueller (2004): "North American Migration: Returns to Skill, Border Effects and Mobility Costs", *Review of Economics and Statistics*, 86, 988-1007.
- [20] Kennan, J. and J. R. Walker (2003): "The Effect of Expected Income on Individual Migration Decisions", NBER Working Papers, No. 9585.
- [21] Kennan, J. and J. R. Walker (2006): "The Effect of Expected Income on Individual Migration Decisions", mimeo, University of Wisconsin.
- [22] Levin, A., C. F. Lin, C. S. J. Chu (2002): "Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties", *Journal of Econometrics*, 108, 1-24.
- [23] Sjaastad, L. (1962): "The costs and returns of human migration", *Journal of Political Economy*, 70, 80-93.
- [24] Storesletten, K., C. I. Telmer, and A. Yaron (2004): "Cyclical Dynamics in Idiosyncratic Labor-Market Risk", *Journal of Political Economy*, 112, 695-717.
- [25] Tauchen, G. (1986): "Finite state Markov-chain approximation to univariate and vector autoregressions", *Economics Letters*, 20, 177-181.
- [26] Tunali, I. (2000): "Rationality of Migration", *International Economic Review*, 41, 893-920.