A generalized options approach to aggregate migration with an application to US federal states

Christian Bayer,*Falko Jüssen‡
Universität Dortmund

Abstract

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 to deal with dynamic self-selection problems that arises from the endogeneity of location choice and the persistency of migration incentives. Keeping track of the distribution dynamics of migration incentives has important consequences, because these dynamics influences the estimation of structural parameters, such as migration costs. For US interstate migration, we obtain a cost estimate of approximately two average annual household incomes. This is at most half of the migration cost estimates reported in previous studies. We attribute this difference to the treatment of the self selection problem.

KEYWORDS: Self selection, migration, indirect inference, dynamic optimization
JEL-codes: C61, C20, J61, R23

1 Introduction

Migration decisions are important economic decisions. Migration allows indivdual agents to smooth income and is an important way of adjustment to macroeconomic shocks (Blanchard and Katz, 1992). Many factors influence the decision to migrate and there is a vast empirical literature that links migration decisions to economic incentives (see

*University of Dortmund, Faculty for Economics and Social Sciences, Department of Economics, 44221 Dortmund; phone: +49-231-755-3246; fax: +49-231-755-3027; email: christian.bayer@wiso.uni-dortmund.de

†University of Dortmund, Faculty for Economics and Social Sciences, Department of Economics, 44221 Dortmund; phone: +49-231-755-3291; fax: +49-231-755-3069; email: f.juessen@wiso.uni-dortmund.de

‡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 their support. We would like to thank Christian Wogatzke for excellent research assistance.

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. Nevertheless, migration costs are surely a structural parameter of high interest (Sjaastad, 1962).

Recently, there has been a small number of studies that actually do report estimates on migration costs. Davis, Greenwood, and Li (2001), henceforth DGL, report a cost estimate of about US$ 180 000 for each migration between US states, and Kennan and Walker (2005), henceforth KW, conclude that all other things equal migration costs are at about US$ 270 000.[1] In terms of average annual income, these migration costs correspond to roughly 4-6 average annual household incomes. At any rate, such an estimate appears very high and even the authors of these studies are somewhat sceptical about their findings.

KW (2005) 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. In fact, this paper's first result is that it is necessary to keep track of the unobservable distribution of migration incentives over time to obtain an unbiased estimate of migration costs.

This motivates us to develop a tractable microeconomic model of migration which can be aggregated and used to describe the simultaneous evolution of migration incentives and migration rates at an aggregate level. Our model picks up the general idea that migration can be understood as an investment into human capital (Sjaastad, 1962) so that the migration-decision problem is closely related to the decision problem for discrete investment projects or lumpy investment.

For the lumpy investment setup, Caballero and Engel (1999) develop a methodological framework that allows to estimate micro-level investment costs from aggregate data only. We extend their work to migration decisions. This means that we first develop a structural model of the representative microeconomic problem of migration for heterogeneous households and in a second step, this model is used to derive the evolution of the distribution of migration incentives. This evolution of incentives determines the actual aggregate migration in turn.

We simulate this model and estimate migration costs via indirect inference (see Gourieroux et al., 1993 and Smith, 1993). Particularly, we apply Smith's (1993) simulated quasi maximum likelihood method. We estimate migration costs to be about US$

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

100 000, which is within the range of two average annual incomes. This cost estimate is substantially lower than the cost estimates of previous studies. Moreover, we show that applying the techniques present in other papers, we would obtain higher cost estimates also from data generated by a simulation of our structural model with the estimated parameters. Consequently, we conclude that keeping track of the distribution of migration incentives over time has an important influence on migration cost estimates. 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, 1992, Tunali, 2000, and Hunt and Mueller, 2004).

Finding more reasonable cost estimates parallels the results of the investment literature, in which much more reasonable adjustment cost estimates where obtained when fixed adjustment costs to capital were included into dynamic models. For migration, the issue of fixed and sunk costs was emphasized in the real-options approach by Burda (1993) and Burda et al. (1998). However, these papers only look at migration as a once and for all decision, so that they preclude return migration. Moreover, the papers do not study the evolution of migration incentives to which past migration decisions feed back.

Taking into account these feedbacks, we extend the structural approaches of DGL (2001) and KW (2005) and suggest a fully structural model of migration that is based upon a dynamic optimization and hence takes into account the dynamic character of the migration decision. This allows us to track the dynamic evolution of migration incentives at the macro level, but it comes at the cost that we have to reduce the model to a bi-regional setup for numerical feasibility. One distinct feature of our model is that it allows to infer the structural microeconomic parameters of the migration decision from aggregrate data only.

Beyond the application to migration decisions, our treatment of the dynamic self-selection problem may be applicable to other important discrete choices in an economy, for example labor-market participation.

The remainder of this paper is organized as follows: Section 2 gives a brief discussion on the difficulties of estimating structural migration models when the population dynamically self selects into their prefered region. It develops the main motive of our paper and illustrates why migration costs are hard to identify by standard (discrete choice) estimation techniques. Thereafter, Section 3 presents a tractable 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 these microeconomic migration decisions. We derive the contemporaneous law of mo-

tion of the distribution of migration incentives and aggregate migration rates taking into account the heterogeneity at the microeconomic level. We provide the results of a numerical simulation analysis in Section 5 to give an idea how the proposed model actually behaves. Section 6 finally confronts the model with aggregate data on migration between US states and derives estimates of the structural parameters of the model, especially on migration costs. Section 7 concludes and an appendix provides detailed proofs as well as details on the employed data.

2 What makes migration costs so hard to identify?

Most micro studies and now also more macro studies on migration link the individual migration decision to a probabilistic model in which agents migrate if the gain in utility terms obtained by migration,

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

is large enough and exceeds some threshold value $\bar{c}$.[2] 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 the economic incentives to migrate, i.e. the gains from migration.

For example, x_{it} could contain data on remuneration, on labor market conditions and on amenities for both the home and the destination region. The parameter γ measures the sensitivity of the migration 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 due to the unobservability of the *potential* migration gains. To illustrate this point, suppose an agent only cares about the difference in income between home and destination region.

In such setting x_{it} were 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

[2] See for example DGL (2001), KW (2005), or Hunt and Mueller (2004).

income has to be proxied. Typically it is proxied by an income a similar agent realizes in the other region.[3] At a macro level, this often means replacing agent specific income differences across regions by average income differences across regions, see for example DGL (2001).

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

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

The composed error term $[\gamma\left(x_{it} - \bar{x}_{.t}\right) + \nu_{it}]$ now also includes the idiosyncratic component of income differences $(x_{it} - \bar{x}_{.t})$. Since we do not want to base our following argument on a classical measurement error or omitted variable problem, assume that the idiosyncratic component to the income difference $\eta_{it} := (x_{it} - \bar{x}_{.t})$ is orthogonal to the average income difference. For the ease of exposition, also suppose 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

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

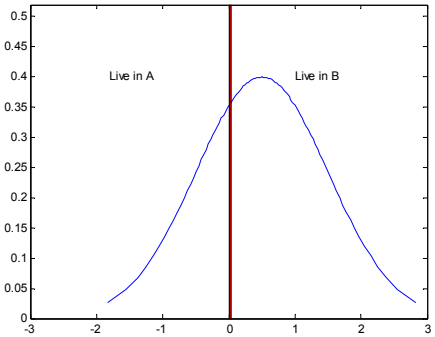
In this equation the regression residual only captures the distribution of idiosyncratic potential income differences around the mean.

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 could nonetheless estimate γ with a suitable probabilistic decision model, e.g. a logit or probit model. However, assuming one of the standard distributions for η_{it} which does not evolve over time is problematic.

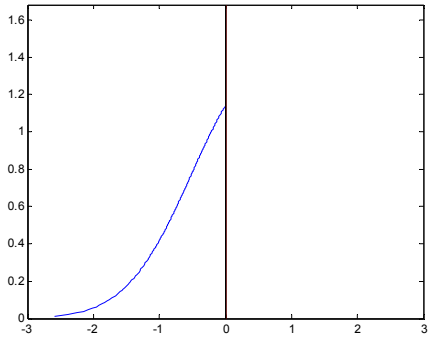
Suppose, agents are heterogeneous with respect to income potentials so that η_{it} has a non-degenerated distribution. In particular assume that η_{it} is initially normally distributed as displayed in Figure 1 (a), so that in the initial situation a probit model were appropriate. The figure displays the distribution of $\bar{x}_{.t} + \eta_{it}$. Low values of this sum imply that income in region A is favorable, high values imply better income prospects in

[3] One example is the paper of Hunt and Mueller (2004) that does a Mincer type wage regression to obtain the unobservable income potential. A similar example can be found in Burda et al. (1998) or KW (2005). For macro-data, an example is DGL (2001).

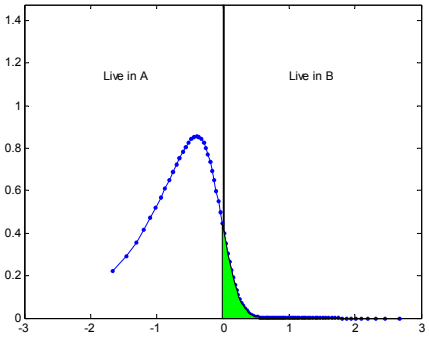
Figure 1: Distribution of potential income in region B relative to A



(a) overall population



(b) conditional on living in region A after migration



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

region B. The figure assumes zero migration costs, so that all agents with $\bar{x}_{.t} + \eta_{it} < 0$ decide to live in region A and they decide to live in region B otherwise. The agents self-select into the region that is favorable for them. [4]

As a result, 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 the migration incentives $\bar{x}_{.t}+\eta_{it}$ are no longer normally distributed

[4] This self-selection is driven directly by the heterogeneity of the agents with respect to income potentials, but is does not reflect immanent and fixed differences of the regions as in Borjas (1992).

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 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 binary choice models. The colored-in region indicates the set of agents that will migrate from A to B after the idiosyncratic shock.

Now, how does this correspond to an unreasonable estimate of migration costs? If $\bar{c}$ 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 a bias in γ directly translates into a bias in estimated migration costs. And with the distribution of migration incentives misspecified, γ will be estimated with a bias most probably. The misspecification of the distribution 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.

To put it simply: Agents are in a certain region most likely because they are better off living there. Because of this self selection, the distribution of unobserved migration incentives is most likely not symmetric (see Greenwood, 1985, pp. 533). Additionally, it displays a dynamic behavior. Accordingly, one needs to keep track of the evolution of the incentive distributions 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, which can be aggregated and used to simulate the evolution of migration incentives over time.

3 A simple stochastic model of migration decisions

We consider an economy with two regions, A and B. For simplicity, this economy is assumed to be inhabited by a continuum of infinitely lived agents of measure 1.[5] We model the economy in discrete time and at each point in time an agent has to decide in

[5] This assumption can be justified by altruism of parents to their children.

which region to live and work. First, we consider the decision problem of an individual agent. For simplicity, we drop the index i that has denoted the specific individual before, but use this index to indicate regions, $i = A, B$.

Living in region i at time t gives the agent utility $\tilde{w}_{it}$. Although $\tilde{w}_{it}$ is a catch-all variable for migration incentives, which can be interpreted as wage income, employment prospects, amenities, utility from social networks etc., we refer to $\tilde{w}_{it}$ as income for simplicity.

The agent discounts future utility by factor $\beta < 1$ and maximizes the discounted sum of expected future utility by location choice. Moving from one region to the other is not costless to an agent. When an agent moves, she is subject to a disutility c_t that enters additively in her utility function.

Hence, the instantaneous utility function $u(i, j, t)$ is given by

$$u\left(i, j, t\right) = \tilde{w}_{it} - \mathbb{I}_{j \neq i} c_t \tag{4}$$

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

Both, migration incentive (income $\tilde{w}_{it}$) and moving costs (c_t), are stochastic variables in our model. They vary over time and across individuals, but are observed by the agent before she chooses her location. The agent knows the distribution of both components of her utility function and forms rational expectations about future incomes and migration costs.

Since migration costs are stochastic and hence vary, not all individual agents who face the same income differential will actually take the same migration decision. In this sense, the individuals in our model are heterogenous and to the outside observer the migration decision is stochastic.

With both $\tilde{w}_{it}$ and c_t being stochastic, the potential migrant waits not only for good income opportunities but also for low migration costs. In her migration decision, she thus takes into account two option values. One is the value to wait and learn more about future incomes and the other is to wait and search for lower migration costs.

Migration costs themselves depend on many factors and may include both physical and psychic costs of migration (Sjaastad, 1962), but the factors that determine migration costs are not constant. For example, search costs to find a new job and accommodation evolve with market conditions, the disutility of living separated from a family or partner changes over time, just as marital status itself is neither constant nor irreversible.

We pick up the variability in migration costs c_t by assuming them to be independently and identically distributed according to a distribution function G. In the simulation and in the estimation of our model, we specify G to be a Gamma distribution.

The distribution of migration incentives, $\tilde{w}_{it}$, is assumed to be log-normal. In particular, we assume that log income, w_{it}, follows an AR(1) process with normally distributed innovations ξ_{it} and autoregressive coefficient ρ :

$$\ln(\tilde{w}_{it}) =: w_{it} = \mu_i(1-\rho) + \rho w_{it-1} + \xi_{it}. \tag{5}$$

This process holds for the whole continuum of agents and each agent draws an own series of innovations ξ_{it}. The expected value of log income in region i is μ_i. The innovations ξ_{it} have mean zero, are serially uncorrelated, but may be correlated across regions A,B and across agents (see Section 4.2).

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

$$V(j, c_t, w_{At}, w_{Bt}) = \max_{i=A,B} \left\{ \exp(w_{it}) - \mathbb{I}_{\{i \neq j\}} c_t + \beta \mathbb{E}_t V(i, c_{t+1}, w_{A,t+1}, w_{B,t+1}) \right\}. \tag{6}$$

In this equation, $\mathbb{E}_t$ denotes the expectations operator with respect to information available at time t.[6]

The optimal policy is relatively simple. The agent migrates from region j to region i if and only if the cost of migration are lower than the sum of the expected value gain $\beta \mathbb{E}_t [V(i, c_{t+1}, w_{A,t+1}, w_{B,t+1}) - V(j, c_{t+1}, w_{A,t+1}, w_{B,t+1})]$ and the direct benefits of migration $\exp w_{it} - \exp w_{jt}$. This means the agent migrates if and only if

$$c_t \leq \exp w_{it} - \exp w_{jt} + \beta \mathbb{E}_t [V(i, c_{t+1}, w_{A,t+1}, w_{B,t+1}) - V(j, c_{t+1}, w_{A,t+1}, w_{B,t+1})]. \tag{7}$$

The value difference

$$\mathbb{E}_t [V(i, c_{t+1}, w_{A,t+1}, w_{B,t+1}) - V(j, c_{t+1}, w_{A,t+1}, w_{B,t+1})]$$

may for example reflect different income expectations. Holding income expectations

[6] For technical reasons, we need to assume boundedness of ξ_{it}, so that ξ_{it} is in fact only *approximately* normal. The bounds to ξ_{it} turn the optimization problem into a bounded returns problem, which is easier to solve. But the bounds to ξ_{it} can be chosen arbitrarily wide (but finite) so that the distribution of ξ_{it} approximates the log-normal distribution arbitrarily close. Existence and uniqueness of the value function is proved in the appendix.

constant, the difference of the expected values also reflects the differences in expected future migration costs.

Since the costs of migration, c_t, are assumed to be i.i.d., expected costs at time $t+1$ do not depend on information available at time t. Moreover, the distribution of future incomes $(w_{A,t+1}, w_{B,t+1})$ is a function of only (w_{At}, w_{Bt}), because w_{it} follows a Markov-process. This allows us to summarize the expected value difference by a function $\Delta V(w_{At}, w_{Bt})$ of only (w_{At}, w_{Bt}), which is defined as

$$\Delta V(w_{At}, w_{Bt}) := \beta \mathbb{E}_t \left[V(B, c_{t+1}, w_{A,t+1}, w_{B,t+1}) - V(A, c_{t+1}, w_{A,t+1}, w_{Bt,+1}) \right]. \quad (8)$$

Substituting (8) for the value difference in (7) gives a critical level of costs $\bar{c}$ at which an agent living in region A is just indifferent between moving and not moving to region B. This threshold is

$$\bar{c}(w_A, w_B) := \exp w_B - \exp w_A + \Delta V(w_{At}, w_{Bt}). \quad (9)$$

To put it differently, a person moves from A to B if and only if

$$c_t \leq \bar{c}_A := \bar{c}(w_{At}, w_{Bt}).$$

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

$$c_t \ \leq \bar{c}_B := -\bar{c}(w_{At}, w_{Bt}).$$

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

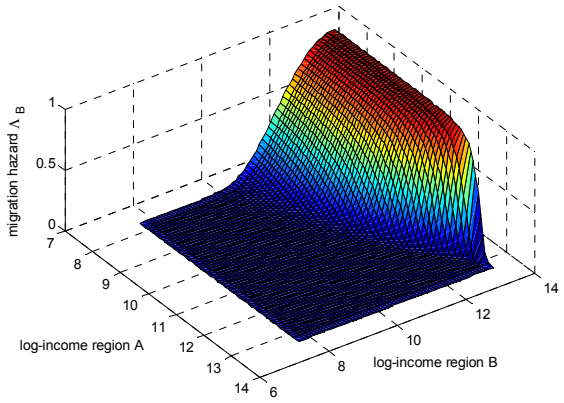
4.1 Aggregate migration

Given this trigger rationale for migration, the hazard rate

$$\Lambda_i(w_A, w_B) := G(\bar{c}_i(w_A, w_B)) \ , \ i = A, B.$$

is the probability that a person in region i moves to the other region if she faces the potential incomes (w_A, w_B). This means that the likelihood of a person to move equals the probability that her migration costs realize below the threshold value $\bar{c}_i$. Since we

Figure 2: Migration hazard for a migration from region A to region B



assumed a continuum of agents the actual fraction of migrating agents with income pair (w_A, w_B) is equal to this hazard rate, too. Figure 2 displays an example of a microeconomic migration-hazard function that stems from the optimization problem (6). The figure shows how different income combinations change the probability to migrate from region A to B.

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

$$\bar{\Lambda}_{it} := \int \Lambda_i \left(w_A, w_B\right) \cdot f_{it} \left(w_A, w_B\right) dw_A dw_B. \tag{10}$$

This means that the aggregate migration hazard, $\bar{\Lambda}_{it}$, is a convolution of the microeconomic adjustment hazard Λ_i and the conditional income distribution f_{it}. In other words, the aggregate migration hazard can be thought of as a weighted mean of all microeconomic migration hazards, 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_{it}) evolves over time and is a result of direct shocks to income just as it is a result past migration. We need to characterize the law of motion

for F to close our model and to obtain the sequence of aggregate migration hazards.

4.2.1 The effect of migration on income distributions

Since the microeconomic migration hazard depends on (w_A, w_B), different potential incomes in both regions result in different propensities to migrate. In consequence, migration changes not only the fraction P_{it} of households living in region i at time t, but also the conditional distribution of income, f_{it}. 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, are very likely to migrate. As a result, the number of those households strongly decreases after migration decisions have been taken, while the number of households facing a smaller income differential changes less.

These considerations form the backbone of our argument. The distribution of migration incentives is a result of past migration decisions, and we can express the new density of households with income (w_A, w_B) in region i after migration, $\hat{f}_{it}$, by

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

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

Analogously, the second part of the sum is constructed: $\Lambda_{-it}(w_A, w_B)$ is the probability to migrate from the other region, $-i$, to destination region i, $f_{-it}(w_A, w_B) P_{-i,t}$ is the unconditional income density for region $-i$ and dividing by $P_{i,t+1}$ conditions for living in region i after migration.

The proportion of households living in region i at time $t+1$ is itself a result of migration decisions. For the law of motion for $P_{i,t+1}$, we obtain

$$P_{it+1} = \left(1 - \bar{\Lambda}_{it}\right) P_{it} + \bar{\Lambda}_{-it} P_{-it}. \quad (12)$$

The number of households living in i is composed of those that stay, $\left(1 - \bar{\Lambda}_{it}\right) P_{it}$, and those that come $\bar{\Lambda}_{-it} P_{-it}$.

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. These shocks can be purely idiosyncratic or may effect all individuals in the economy. For a single agent, we can decompose the total shock ξ_{it} to her potential income in region i (see equation 5) into an aggregate component θ_{it} and an individual specific component ω_{it} :

$$\xi_{it} = \theta_{it} + \omega_{it}.$$

The aggregate shock θ_{it} for region i hits all agents equally and changes their potential income for region i. It is important to note that this shock does not depend on the actual region the agent lives in. For example, a positive shock $\theta_{At} > 0$ increases the *potential* income in region A for agents that currently live in region A as well as for agents that currently live in region B. They realize this potential income by deciding to actually live in region A. If both θ_{At} and θ_{Bt} are positive (negative) both regions become more (less) attractive. Income increases more in the region which has the relatively larger shock.

Hence, aggregate shocks θ measure economy wide business cycles as well as regional cycles such as local demand or supply shocks. These regional cycles could also result from different technology or industry mixes in both regions, which lead to different responses to general shocks to productivity.

Statistically, the economy wide business cycle is the common component in (θ_A, θ_B). If this business cycle component is more important relative to the regional cycles, then the correlation ψ_θ between θ_A and θ_B is large.

However, aggregate shocks are not the only source of income variation for an agent. 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 $(\omega_{At}, \omega_{Bt})$. If ω_{At} is positive, income prospects of the individual agent increase in region A. The correlation ψ_ω between ω_A and ω_B reflects economy wide demand shifts for a person's individual skills.

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 much more tractable.[7]

If the variance of idiosyncratic shocks ω is large relative to the variance of aggregate shocks θ, heterogeneity among agents plays a large role and has a big influence in

[7] If we were to include fixed effects that reflect different types of agents, the model had to be solved for each different agent type just as it is now solved for the single type of agent.

determining one's income. Since we assume aggregate and idiosyncratic shocks to be independent, the variance of the total shock to income, ξ_{it}, is the sum of the variances of idiosyncratic and aggregate shocks: $\sigma_\xi^2 = \sigma_\omega^2 + \sigma_\nu^2$.

The income distribution at the beginning of the next period, F_{t+1}, now results from adding idiosyncratic and aggregate shocks to the distribution of income after migration in period t, $\hat{F}_t$, of which $\hat{f}_{it}(w_A, w_B)$ is the conditional density, see (11). When a household has income w_{it+1} in period $t+1$, this can result from any possible combination of w_{it} and $\xi_{it} = \theta_{it} + \omega_{it}$ for which

$$w_{it+1} = \mu_i(1-\rho) + \rho w_{it} + \theta_{it} + \omega_{it} \tag{13}$$

holds. Solving this equation for w_{it} we obtain

$$w_i^* := w_{it} = \frac{w_{it+1} - (\theta_{it} + \omega_{it})}{\rho} - \mu_i \frac{(1-\rho)}{\rho}. \tag{14}$$

This $w_i^*(w_{it+1}, \theta_{it}, \omega_{it})$ is the current potential income in region i that is consistent with a future potential income of w_{it+1} and realizations of shocks $\theta_{it} + \omega_{it}$ in period t. 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 income $(w_{A,t+1}, w_{B,t+1})$ to current income $(w_{A,t}, w_{B,t})$.

The conditional density of observing the future income pair $(w_{A,t+1}, w_{B,t+1})$ can thus be obtained from a retrospective. The income pair (w_A^*, w_B^*) of past incomes corresponds uniquely to a future income pair $(w_{A,t+1}, w_{B,t+1})$. 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 $\hat{f}_{it}$. The density of the income distribution F_{t+1} conditional on the region and the vector of shocks is given by

$$f_{it+1}(w_A, w_B | \theta_A, \theta_B, \omega_A, \omega_B) = \hat{f}_{it}(w_A^*(w_A, \theta_A, \omega_A), w_B^*(w_B, \theta_B, \omega_B)). \tag{15}$$

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

$$\hat{f}_{it}(w_A^*(w_{A,t+1}, \theta_{At}, \omega_{At}), w_B^*(w_{B,t+1}, \theta_{Bt}, \omega_{Bt})) \cdot h(\omega_{At}, \omega_{Bt}).$$

Integrating over all possible idiosyncratic shocks $(\omega_{At}, \omega_{Bt})$ gives the density f_{it+1} of the income distribution before migration in period $t+1$ for a certain aggregate shock

$(\theta_{At}, \theta_{Bt})$:

$$f_{it+1}\left(w_A, w_B|\theta_{At}, \theta_{Bt}\right) = \int \hat{f}_{it}\left(w_A^*\left(w_A, \theta_{At}, \omega_{At}\right), w_B^*\left(w_B, \theta_{Bt}, \omega_{Bt}\right)\right) \cdot h\left(\omega_A, \omega_B\right) d\omega_A d\omega_B. \quad (16)$$

For given aggregate shocks, this new distribution determines migration from region i to region $-i$ according to equation (10) for time $t+1$.

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_{it} to $\hat{f}_{it}$. Thereafter, the distribution is again altered by aggregate and idiosyncratic shocks to income, moving the density from $\hat{f}_{it}$ to $f_{i,t+1}$. The latter density now determines migration decisions in time $t+1$, starting the cycle over again. In other words, migration is not only a result of past income shocks, but also a result of past migration decisions. Keeping track of the distributional dynamics of migration incentives is at the heart of our model. This is the difference to most other empirical models of migration.

5 Simulation analysis

5.1 Numerical aspects

The first step in solving the model numerically is to obtain a solution to (6) . We do so by value-function iteration.[8] For this value-function iteration, we first approximate the bivariate process of potential income for an individual agent in region A and B

$$\begin{pmatrix} w_{At} \\ w_{Bt} \end{pmatrix} = w_t = \mu\left(1-\rho\right) + \rho w_{t-1} + \xi_t \quad (17)$$

by a Markov chain.[9] Because w_A and w_B are correlated through the correlation structure in ξ, it is easier to work with the orthogonal components $\left(w_A^+, w_B^+\right)$ of (w_A, w_B) in the value function iteration.

We evaluate the value function on an equispaced grid for the orthogonal components with a width of $\pm 4\sigma_{A,B}^+$ around their means, where $\sigma_{A,B}^+$ denote the long run standard deviations of the orthogonal components. The grid is chosen to capture al-

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

[9] To save on notation we drop the regional index of a variable pair like (w_{At}, w_{Bt}) and just denote the pair by w_t.

most all movements of the income distribution F later on.[10] Given this grid, we can use Tauchen's (1986) algorithm to obtain the transition probabilities for the Markov-chain approximation of the income process in (17) .

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 of the new grid. The initial grid has $16\times16\times32$ points (income A $\times$ income B $\times$ migration costs) and the final grid has 128×128 points for income and 256 points for migration costs.[11]

The solution of (6) yields the optimal migration policy and thus the microeconomic migration-hazard rates Λ_i. 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 $(\theta_t)_{t=1..T}$ 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. This distribution is calculated by assuming that idiosyncratic shocks ω have the full variance of ξ. In the appendix, we show that the sequence of income distributions converges to a unique ergodic distribution $\bar{F}$ in the absence of aggregate shocks. 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 $\left(\bar{\Lambda}_{A,B}\right)_{t=1..T}$, we draw a series of aggregate shocks (to the orthogonal basis) $\left(\theta_A^+,\theta_B^+\right)_{t=1..T}$ from a normal distribution with variance $\phi\cdot\left(\sigma_{A,B}^+\right)^2$, $\phi\in[0,1]$. The

[10] The choice of $\pm4\sigma_{A,B}^+$ is motivated as follows. We later assume in the simulations that 95% 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 $\pm3.29\cdot\sqrt{0.98}\sigma_{A,B}^+\cong\pm3.26\sigma_{A,B}^+$ around the mean of the distribution for any given year. Additionally, the mean income for each year moves within the band $\pm3.29\cdot\sqrt{0.02}\sigma_{A,B}^+\cong\pm0.47\sigma_{A,B}^+$ in again 99.9% of all years. Since the sum of both is $3.73\sigma_{A,B}^+$, a grid variation of $\pm4\sigma_{A,B}^+$ should not truncate the income distribution.

[11] To obtain the grid for migration costs, we first discretize the [0;1] interval into an equispaced grid. Then we choose the grid points for the migration costs as the values of the inverse of the cumulative distribution function of the costs evaluated at the equispaced grid. This yields a cost grid whose grid points are equally likely to realize. By contrast to the income distribution, using such an "equally-likely grid" is possible for the cost distribution, because the cost distribution is strictly stationary. Unlike the income distribution, it does not move due to aggregate shocks. See Adda and Cooper (2003) or Tauchen (1986) for the analog case of a stationary Markov chain with normal innovations.

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 $(1-\phi)\cdot\left(\sigma_{A,B}^{+}\right)^2$.

5.2 Parameter choices

A number of parameters has to be determined to actually simulate our model numerically. The probably most important parameter choice concerns the distribution of migration costs. Throughout the rest of the paper, we assume migration costs to be Gamma-distributed, i.e. the cumulative distribution function of migration costs is

$$G(c) = \frac{1}{a^b \Gamma(b)} \int_0^c x^{b-1} \exp\left(\frac{-x}{a}\right) dx. \tag{18}$$

This distribution has two parameters a and b which determine the mean ab and the coefficient of variation $b^{-\frac{1}{2}}$. Although the mean cost is ab, one should note that the average cost paid by a migrant is smaller as she will wait and search for low migration costs. In our simulations, we try three parameter combinations (a, b) to see their influence on the dynamics of interregional migration. One parameter constellation with high, one with medium and one with almost zero migration costs. This allows us to asses the sensitivity of aggregate migration to moving costs. In particular, we are interested to see whether the high migration cost estimates reported in the literature are compatible with aggregate migration data in the light of our model.

The second important set of parameters describes the process for income and the income shocks ξ. We need to specify the autocorrelation parameter ρ and the mean μ of the income process as well as the covariance structure of the income shocks. The covariance structure is composed of the variance of income shocks σ_{ξ}^2, the correlation of income shocks between regions, ψ_{θ} (aggregate) and ψ_{ω} (idiosyncratic), and the fraction ϕ of the income shock that is due to aggregate factors, i.e. the covariance across individual agents.

We take the parameters for the income process mainly from the recent paper of Storesletten et al. (2004). They estimate the dynamics of idiosyncratic labor market risk for the US and report both income variances and autocorrelation of log household income based on the Panel Study of Income Dynamics. They find an annual autocorrelation 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 et al. 2004, Table 2). They report a frequency weighted average of 0.17 for those standard deviations in their preferred specification (Stoesletten et al. 2004,

pp. 711). Overall, their results imply a range of 0.125 to 0.192 for the average standard deviation of the idiosyncratic income shock, taking means of their point estimates for contractions and expansions. 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 simulations.[12]

Besides the autocorrelation and variance terms, Storesletten et al. (2004) also report a mean income of about US$ 45 000. To approximately match this figure, we choose the mean of the *log* income to be $\mu \cong 10.5$.[13]

Unfortunately, Storesletten et al. (2004) do not report numbers on aggregate income risk, so that we need to take this data from a different source. We estimate the variance of aggregate shocks to income from income per capita data for US states for the years 1969 - 2003 as reported in the REIS database (available online at www.bea.gov/bea/regional/reis/). We deflate the data using the US wide consumer price index.

Taking further into account fixed effects and a linear trend, the residual standard deviation of log income for US states over time is roughly 0.13. To calculate the fraction, ϕ, of income risk due to aggregate fluctuations, we compare our estimated aggregate variance with the long-run idiosyncratic variance of income that is implied by Storesletten et al.'s (2004) estimates. Their estimates of the short-run variance correspond to a long-run variance of $\frac{\sigma_\omega^2}{1-\rho^2} = 0.30$. Hence, aggregate income risk accounts only for a fraction of approximately $\frac{0.07^2}{0.3+0.07^2} \cong 0.02$ of total income risk.

Finally, we need to specify the correlations of shocks to income across regions, ψ_ω and ψ_θ. 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 specify one common parameter. As a first approximation, we measure ψ as the correlation coefficient of state average income per capita and the US average per capita income (both in logs, CPI deflated and taking fixed effects and a linear trend into account). From the REIS database we infer a partial correlation coefficient of $\hat{\psi} = 0.55$.

As we work with annual data, we choose the discount factor $\beta = 0.95$. Table 1 summarizes our parameter choices for the three specifications that we simulate.

[12] Other studies on the evolution of individual income report similar values, see the discussion in Storesletten et al. (2004).

[13] A log-normally distributed variable has mean $\exp\left(\mu + \frac{\sigma^2}{2}\right)$ where μ and σ^2 are the mean and variance of the logs.

Table 1: Parameter choices for the simulation analysis

	Storesletten et al. (2004)	REIS data	Model Specifications (1)	(2)	(3)
Gamma Parameter a	–	–	600	300	1
Gamma Parameter b	–	–	300	150	1
Fraction of aggregate shocks ϕ	–	0.02	0.02	0.02	0.02
Correlation of shocks across regions ψ	–	0.55	0.55	0.55	0.55
Long-run variance $\frac{\sigma_\omega^2+\sigma_\theta^2}{1-\rho^2}$	0.30[1]	–	0.32	0.32	0.32
Autocorrelation of income ρ	0.95	–	0.95	0.95	0.95
Discount factor β	–	–	0.95	0.95	0.95

[1] Storesletten et al. (2004) report idiosyncratic variances only. To obtain the composed variance, their estimate has to be divided by $(1-\phi)$.

5.3 Simulation results

We simulate our model for 51 pairs of regions and 26 years, but we drop the first 10 years for each region to minimize the influence of our initial choice of F_0. This generates a simulated dataset for migration data that has the same size as the Internal Revenue Service (IRS) area to area migration flow data, which we use as a benchmark. Income data is taken from the REIS database, CPI deflated and in logs. A detailed data description for both IRS and REIS data can found in the data appendix. In order to minimize simulation uncertainty we replicate each simulation 10 times and report the averages over the simulations.

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 alternatives states where she can move to. To make this comparable to our model, the 50 alternatives have to be aggregated to a single complementary region. The population weighted average income over all alternative 50 states is used as the average income of the alternative region.

To characterize the results of the simulation exercise, we have to calculate a number

of moments from the simulated dataset and compare these moments to the moments we observe from the actual IRS and REIS data. This comparison tells us how well our model is capable to replicate characteristic features of the actual migration and income data at an aggregate level. In particular, it tells us which of the three considered levels of migration costs is best compatible with the observed structure of the data. Such way of inference is frequently applied in the literature on real business cycles, see Backus et al. (1992) or Baxter and Crucini (1993) and many others.

The lines along which this literature has typically described aggregate fluctuations guides our choice of characterizing moments: variances, covariances, autocorrelations and means. We compare average migration rates, the standard deviation of annual migration rates, their autocorrelation and the cross-correlation of migration-rates. Besides, we look at the implications of the different migration cost regimes on the level and fluctuations of average incomes. To measure the cyclical behavior of migration, we calculate the mean of in- and outmigration rates and correlate this with the average income in both regions, $\bar{w}_{it} := \ln\left(\int \exp w_i f_{it}\left(w_i, w_{-i}\right) dw_i dw_{-i}\right)$.

Table 2 reports the results of our simulation exercise. The first experiment uses cost parameters close to what has been reported in the literature. We assume $a = 600$ and $b = 300$ to match an average migration cost of US$ 180 000 as reported in DGL (2001). The results of this experiment are displayed in column (1) of Table 2.

Compared to the actual data, the annual migration rates are by far too low. While we observe an annual average migration rate of 3.9%, the model predicts a migration rate of only 1.0%. With US$ 180 000, migration costs are just prohibitively high. Also migration rates fluctuate less in the simulated data than in the actual data. Simulated migration rates are too much procyclical and the cross-correlation of incomes is 0.682, while the correlation of shocks ψ was set to be 0.55.

In summary, we obtain too little migration and too little fluctuation of the migration rates, while income fluctuation is realistic. Therefore, we try a specification with lower migration costs. We set $a = 300$ and $b = 150$, so that expected migration costs are devided by four and now equal an average annual income of US$ 45 000. With these lower migration costs, migration rates more than double and are with 2.5% within a more realistic range. Migration also becomes less procyclical and the fluctuation of income decreases. At the same time, migration rates themselves fluctuate more. Consequently, the lower cost specification more closely replicates observed data. However, in- and outmigration seem to be too strongly negatively correlated. A further result of lower migration costs is an increase in average income by 3% compared to the high cost specification. With lower migration costs, the agents are more often in the region where

Table 2: Simulation Results

	Data	high costs (1)	medium costs (2)	zero costs (3)
average annual migration rate	0.039	0.010	0.025	0.102
standard deviation of annual migration rates	0.004	0.003	0.006	0.018
Autocorrelation of migration rates[1]	0.807	0.438	0.436	-0.011
Cross correlation of migration rates[2]	0.045	-0.774	-.919	-0.985
mean of log average income	10.710	10.794	10.823	10.832
standard deviation of log average income	0.071	0.073	0.063	0.062
Cross correlation of log average income [2]	0.550	0.682	0.714	0.853
Variance of household income	0.299	0.320	0.290	0.275
Correlation of $\bar{\Lambda}_i + \bar{\Lambda}_{-i}$ and $\bar{w}_i + \bar{w}_{-i}$ (procyclicality)[2]	0.225	0.757	0.596	0.037

[1] Coefficient of the autoregressive parameter in a fixed effects regression with linear time trend.
[2] Partial correlation controlling for a linear time trend.

their income is larger.

While the first scenario displayed an extreme bound of very high migration costs, the third scenario of almost no migration costs provides a lower extreme bound. It clearly shows how influential it is to keep track of the evolution of migration incentives. As KW (2005, p. 28) point out, in a model in which migration incentives are drawn randomly, we should observe migration rates of 50% in the absence of migration costs. By contrast, our model predicts a substantially lower migration rate of 10.2% when migration costs are absent. This difference to Kennan and Walker's intuitive result stems from the fact that migration incentives are not drawn purely randomly in our model. Instead, they depend on previous migration incentives and decisions.

Besides this main point, we see that the procyclicality of migration rates drops further

and in fact becomes too low. The cross correlation of migration rates becomes almost perfectly negative.

Overall, our simulation results do not yet allow a decisive assessment which level of migration costs fits the data best. The average migration rates and their fluctuations are best captured by the medium cost formulation. When we look at the cross correlation of migration rates the high cost specification seems most plausible. Finally, low migration costs imply the best match of the observed low procyclicality of migration rates.

6 Estimation

In order to find the parameters of our model that allow to match closest the observed patterns of migration that are in the data, we rely on an indirect inference procedure. Indirect inference procedures have been proposed by Smith (1993) and Gourieroux et al. (1993) to obtain estimates of structural parameters when the likelihood function of the structural model becomes intractable, as in our setting. In particular, we apply the simulated quasi-maximum likelihood (SQML) method developed in Smith (1993). This method avoids the estimation or choice of a weighting matrix and hence is arguably more robust in small samples (Smith, 1993).

6.1 Methodology

Indirect inference is the natural extension of the simulation exercise presented in the previous section. The central idea behind this methodology is to use an auxiliary statistical model to describe the observed patterns of the data, and then to calibrate and simulate the structural economic model such that the auxiliary statistical model is best replicated by the simulation.

Accordingly, we first estimate an auxiliary model that describes observed migration data x in a reduced form using a quasi maximum likelihood approach. This means we select vector of reduced form parameters $\bar{\varrho}$ that maximizes the likelihood function $L(x, \varrho)$. Thereafter, we simulate the migration model as described in the previous section for a vector of structural parameters β and thus generate artificial data y. These simulated data are then used to estimate reduced form parameters $\hat{\varrho}(\beta)$ as maximizers of $L(y(\beta), \varrho)$. Finally, we choose $\hat{\beta}$ such that the likelihood of the actual data $L(x, \hat{\varrho}(\beta))$ under $\hat{\varrho}$ becomes maximal. A comparison of the unrestricted maximum likelihood estimate $\bar{\varrho}$ and the the estimate under the restrictions imposed by the structural model, $\hat{\varrho}(\beta)$, then tells us how well our economic model is able to describe the observed data in the light of the reduced form.

In other words, we can understand the auxiliary model as a lens through which we

look at both the model and the observed data. Consequently, the choice of the auxiliary model is important (in particular in small samples). A number of studies of migration has suggested to regress either net migration or immigration on income differentials as the canonical approach to aggregate migration data (see Greenwood, 1997 for an overview). We follow this suggestion in the choice of our auxiliary model and regress aggregate immigration rates on the average incomes of both regions. In particular, we assume the following functional form

$$im_{it} = \varrho_0 + \varrho_1 \bar{w}_{it} + \varrho_2 \bar{w}_{-it} + \varepsilon_{it} \tag{19}$$

$$\varepsilon_{it} = \varrho_3 \varepsilon_{it-1} + \nu_{it}. \tag{20}$$

Immigration im_{it} to state i is depends on both the average income $\bar{w}_{it}$ in state i and the average income in the alternative region $\bar{w}_{-it}$. The constant term ϱ_0 reflects that in both our model and in reality there is an equilibrium level of migration due to idiosyncratic income risk. The parameters $\left(\varrho_{1,2}\right)$ capture not only the sensitivity of migration to economic incentives, but also the procyclicality of migration $\left(\varrho_1 + \varrho_2\right)$.

We allow for the possibility of autocorrelation in the error term to pick up a certain degree of persistency in migration rates that is both present in the actual and the simulated data. The persistency in aggregate migration rates results from the autocorrelation ρ of individual incomes, see (5). Finally, the variance σ_ν^2 of the shock ν_{it} picks up the variability in migration.

This leaves us with the following log-likelihood function for a series of observation for state i

$$\log L_i = -\frac{T}{2}\left(\log 2\pi\sigma_\nu^2\right) + \frac{1}{2}\log\left(1-\varrho_3^2\right) - \frac{1}{2}\frac{\sum_{t=1}^{T}\nu_{it}\left(\varrho\right)^2}{\sigma_\nu^2}. \tag{21}$$

The error terms ν_{it} are obtained as the residuals from the quasi-differenced data (see e.g. Greene, 2003, pp. 272).

For our simulated theoretical model all states i and j are mutually independent. For the observed data, we assume that this also holds approximately. Under this assumption we obtain the quasi-likelihood of all state data:

$$\log L = \sum_{i=1}^{N}\log L_i = -\frac{NT}{2}\log 2\pi\sigma_\nu^2 + \frac{N}{2}\log\left(1-\varrho_3^2\right) - \frac{1}{2}\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\nu_{it}\left(\varrho\right)^2}{\sigma_\nu^2} \tag{22}$$

Maximizing this likelihood provides us with the reduced form parameter estimates $\varrho_0, \varrho_1, \varrho_2, \varrho_3$ and σ_ν^2.

Table 3: Reduced Form Estimation Results

parameter	intercept (ϱ_0)	destination income (ϱ_1)	source income (ϱ_2)
estimate	-0.1037	0.0452	-0.0318
standard errors	(0.0693)	(0.0067)	(0.0082)
parameter	autocorr. (ϱ_3)	variance $\left(\sigma_\nu^2\right)$	
estimate	0.8066	3.9159×10^{-6}	
standard errors	(0.0194)	$\left(1.9492 \times 10^{-7}\right)$	
Log-Likelihood	3895.1		

6.2 Estimation results

Table 3 displays the point estimates of the parameters for the reduced form model from the IRS and REIS data.

One can see that immigration to state i responds more strongly to the wage in state i than to the wage in source state $-i$. Suppose that income equally grows in all US states, then all states experience an increase in immigration, so that the overall mobility measured by the total number of migrants increases. To put it differently, the positive sum of $\varrho_1 + \varrho_2$ reflects a procyclicality of migration.

The substantial autocorrelation reflects the dynamic structure of migration, in particular it reflects dynamically evolving idiosyncratic migration incentives. At the same time, it also reflects how well the average incomes proxy for the distribution of idiosyncratic migration incentives. Recall from Section 2 that we loose information by characterizing the unobserved distribution of idiosyncratic incentives by only the mean incomes. The dynamics of the incentive distribution implies that the lost information is correlated over time. This dynamics then manifests itself in an autocorrelation of the error term in (19) .

Table 4 displays the estimates for the structural model obtained by simulated quasi-maximum likelihood. The estimation of the model fixes all prameters besides a and b to the values we used in our simulation before. We obtain an expected migration cost of US$ 101 645 or two average annual incomes, which lies substantially below the estimates reported in other contributions such as DGL (2001) or KW (2005). With a coefficient of

Table 4: Reduced Form Estimation Results

parameter	
expected migration costs $(a \cdot b)$	101 645
	(tba)
coefficient of variation $\frac{1}{\sqrt{b}}$	0.0628
	(tba)
Log-Likelihood	3 466

Standard errors in parenthesis.

variation of 0.0628 the migration costs do not strongly fluctuate across agents in every year.

To make these estimation results more interpretable, we report the statistics calculated for the simulation exercise also for our estimated parameters, which are displayed in Table 5. Migration shows indeed a smoothing effect on incomes. The variance of income across households is about 0.304 although the long-run variance of income possibilities was set to 0.32. Realized incomes substantially correlate across states and the correlation coefficient is 0.728 for the model simulated with the estimated parameters. Overall the fit of our model to the descriptive statistics is relatively good. What remains a puzzle nonetheless is the slightly positive cross sectional correlation of migration rates and their high auto-correlation in the data, which our model cannot replicate. Moreover, our model predicts migration rates to be much more procyclical than we observe.

6.3 Comparison of cost estimates

On the basis of the latter simulation, we may compare incurred migration costs to the costs that would be estimated in the recent approaches to migration such as DGL (2001). This provides further evidence on the influence of the dynamics of the incentive distribution on the estimation of migration costs. DGL (2001) employ a random-utility conditional-logit model to describe the migration decision. Adapted to our bi-regional model, the likelihood of the conditional logit model becomes

$$\ln L = \sum_t \sum_{i=1,2} \left[\begin{array}{c} \bar{\Lambda}_{it} P_{it} \ln \left(\frac{1}{1+\exp\{\alpha+\beta(\bar{w}_{it}-\bar{w}_{-it})\}} \right) \\ + \left(1-\bar{\Lambda}_{it}\right) P_{it} \ln \left(\frac{1}{1+\exp\{-\alpha-\beta(\bar{w}_{it}-\bar{w}_{-it})\}} \right) \end{array} \right]. \tag{23}$$

Table 5: Simulation Results: estimated parameters

	Data	Estimated Model
average annual migration rate	0.039	0.016
standard deviation of annual migration rates	0.004	0.004
Autocorrelation of migration rates[2]	0.807	0.402
Cross correlation of migration rates[1]	0.045	-0.851
mean of log average income	10.710	10.82
std deviation of log average income	0.071	0.072
Cross correlation of log average income[1]	0.550	0.727
Variance of household income	0.299	0.304
Correlation of $\bar{\Lambda}_i + \bar{\Lambda}_{-i}$ and $\bar{w}_i + \bar{w}_{-i}$ (procyclicality)[1]	0.225	0.736

[1] Partial correlation controlling for a linear time trend.
[2] Coefficient of the autoregressive parameter in a fixed effects regression with linear time trend.

While DGL (2001) include a bunch of other variables, our simulated model just allows for log income as an explanatory variable. Other variables such as distance or unemployment are not prevalent in the simulation and hence cannot be included. Moreover, since we only have two regions, we cannot estimate α and β from a cross section as DGL (2001) do, but have to pool the simulated data.

Following DGL (2001), a random-utility conditional-logit approach could be motivated by assuming that utility is composed of an income component (with sensitivity $\beta > 0$) and a disutility from migration $\alpha < 0$. The money measure of this disutility is $\exp\left(\bar{w} - \frac{\hat{\alpha}}{\hat{\beta}}\right)$ and this is DGL's (2001) suggestion of a measure of migration costs.[14]

For the estimated parameters, the mean incurred migration costs are US$ 100 031 (see Table 6), while the conditional logit would suggest a cost of US$ 127 731, a number that

[14] We deviate slightly from DGL (2001) by replacing differences in relative income $\frac{\exp \bar{w}_i}{\exp \bar{w}_{-i}} - 1$ by log differences $\bar{w}_i - \bar{w}_{-i}$ for notational ease.

Table 6: Simulation Results: Comparison to cost estimate by DGL

Average incurred migration cost	100 031
Standard deviation of average incurred migration costs	152
Average annual income	50 193
Migration cost estimate based on DGL's method (2001)	127 731

is substantially higher. In terms of annual incomes this corresponds to 2 and 2.5 average annual incomes respectively. This comparative exercise shows that the estimation of the structural parameters is likely to be subject to a bias if the unobserved dynamics of the distribution of incentives is not taken into account. Besides, it is not clear what the cost estimate of a static decision model as the model of DGL (2001) exactly measures in the context of migration being a dynamic decision and $\bar{w}_{it} - \bar{w}_{-it}$ only capturing the contemporaneous gain from migration.

7 Conclusion

We have provided a tractable model of aggregate migration with a sound microeconomic foundation. It is a contribution to the recently evolving literature on structural models of migration. We explicitly deal with the problem of the unobservability of potential gains from migration and their dynamic character. The dynamic character of migration incentives has two aspects. First, the individual gains from migration evolve stochastically over time, but will typically display a high degree of persistency. Second, at an aggregate level, the distribution of migration incentives is a result of past migration decisions themselves. Starting off from the microeconomic decision problem allows us to keep track of this dynamics of the incentive distribution. This distributional dynamics may be refered to as a dynamic self-selection problem. Neglecting this self-selction problem may result in biased estimates of structural parameters. In fact, we can infer migration costs from our model and find the estimated migration costs to be substantially lower than those reported in previous studies that apply different methods. We estimate migration costs to be about 100 000 US$, corresponding to two average annual incomes at most.

These results were obtained by indirect inference, that is by means of simulated

quasi maximum likelihood. Our analysis calls once more for a careful treatment of the self-selection problem when economic incentives are not fully observable. What makes this issue in particular problematic in the migration setting is that the unobservable incentives are highly auto-correlated 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. This may be of importance not only to macro-studies of migration. Also at a micro level, income potentials are typically unobservable and have to be proxied. But such approximation regularly neglects self selection. If people live in their prefered 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 are. Besides unobservable parts of income this unobservable component can also comprise different valuations of different amenities and social networks. Also these factors can be expected to exhibit persistency. We integrated this persistency in a structural dynamic microeconomic model of the migration decision, which consequently allowed us to simulate the joint behavior of the observed migration rates, of the unobserved migration incentives, and of their observable proxies. Accordingly, simulation based inference methods prove appropriate to overcome the selectivity problem.

Future research would call for a more complex microeconomic model that allows to integrate more information into the macroeconomic reduced form regression model, for example labor market conditions and amenities. This would require some more complex general equilibrium modelling, which currently goes beyond what is numerically feasible.

Both, our treatment of the self-selection problem and the inference of microeconomic structural parameters from macro data is an attempt to overcome the dichotomy of macro and micro studies that has characterized the migration literature (see Greenwood, 1997).

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 maintext throughout this appendix, unless stated otherwise.

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

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

$$T(u)(\cdot) = \max_{j=A,B} \left\{ \exp(w_{jt}) - \mathbb{I}_{\{i \neq j\}} c_t + \beta \mathbb{E}_t u(j, c_{t+1}, w_{At+1}, w_{Bt+1}) \right\}. \tag{24}$$

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 c and have domain $D = \{A, B\} \times \mathbb{R}_+ \times \mathbb{W}^2$.

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*

$$Tu = \max_{j=A,B} \left\{ \exp\left(w_{jt}\right) - \mathbb{I}_{\{i \neq j\}} c_t + \beta \mathbb{E}_t u\left(j, c_{t+1}, w_{At+1}, w_{Bt+1}\right) \right\} \leq \exp\left(\bar{W}\right) + \beta \bar{u} < \infty, \tag{25}$$

and

$$Tu = \max_{j=A,B} \left\{ \exp\left(w_{jt}\right) - \mathbb{I}_{\{i \neq j\}} c_t + \beta \mathbb{E}_t u\left(j, c_{t+1}, w_{At+1}, w_{Bt+1}\right) \right\} \tag{26}$$

$$\geq \max_{j=A,B} \left\{ \exp\left(w_{jt}\right) - \mathbb{I}_{\{i \neq j\}} c_t + \beta \underline{u} \right\} \geq \exp\left(\underline{W}\right) + \beta \underline{u} > -\infty. \tag{27}$$

■

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

$$T(u + a) = Tu + \beta a. \tag{28}$$

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 Lemmata.* ■

8.2 Invariant distribution

We prove that without aggregate shocks migration and idiosyncratic shocks to income describe an ergodic Markov-process. Therefore there is an invariant distribution, this process converges to.

For simplicity we present the proof for an arbitrary discrete approximation of the continuous income (state-space) model.

Lemma 7 *Assume any (large and fine enough but otherwise arbitrary) discretization of the state space with n points for the potential income in the regions, each. Then we can capture the transition from f_t to f_{t+1}, which are the unconditional density of the distribution of households over both regions and (potential) income, in a matrix* $B = \begin{pmatrix} (I - D_A)\,\Pi & D_B\Pi \\ D_A\Pi & (I - D_B)\,\Pi \end{pmatrix} \in R^{2n^2 \times 2n^2}$.[15] *In this matrix, Π denotes the transition matrix that approximates the the AR(1)-process for income by a Markov-chain, see Adda and Cooper (2004, pp. 56) for details. Matrix D_i is the $n^2 \times n^2$ diagonal matrix with the hazard rate for each of the n^2 income pairs of the income grid.*

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

$$f = \Big(f(A, w_{A1}, w_{B1}) ... f(A, w_{An}, w_{B1}) ... f(A, w_{An}, w_{Bn}) f(B, w_{A1}, w_{B1}) ... f(B, w_{An}, w_{Bn}) \Big)'. \tag{29}$$

Analogous, we define the distribution after migration but before idiosyncratic shocks, $\hat{f}$. Taking our law of motion from (16) *we obtain as a discretized analog*

$$f_{t+1} = (I_2 \otimes \Pi)\, \hat{f}_t. \tag{30}$$

Here $\otimes$ denotes the Kronecker product. Now, given our vectorization of the wage grid, define d_i as the fraction of households that migrate and are in the $i-th$ wage and location triple, i.e. $d_i = \Lambda_j(w_{Ak}, w_{Bl})$, $i = 1...2n^2$, where (j, w_{Ak}, w_{Bl}) being the i-th element in the vectorized grid. Moreover, define $D = diag(d)$ 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

$$\hat{f}_t = \begin{pmatrix} I - D_A & D_B \\ D_A & I - D_B \end{pmatrix} f_t \tag{31}$$

[15] Since we work with a discretization, correctly speaking f is not the density, but the vactor of probabities of drawing a location-income possibility vector from a given element of the grid.

Combining the last two equations, we obtain

$$f_{t+1} = \begin{pmatrix} (I - D_A)\,\Pi & D_B\Pi \\ D_A\Pi & (I - D_B)\,\Pi \end{pmatrix} f_t. \tag{32}$$

■

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 income can be reached from every other pair as a result of the shock. Thus all entries of* Π *are strictly positive.* ■

Lemma 9 *For any distribution of costs with support equal to* R_+, $0 \leq d_i < 1$ *holds for all diagonal elements* d_i *of* D. *If the grid is fine enough also* $d_i > 0$ *holds for at least one* i.

Proof. *If there is no upper bound for migration costs the migration probability is strictly smaller, since* V *is bounded. This means* $0 \leq d_i < 1$. *Let* $C_{\max} = \max_{(w_A, w_B) \in \mathbb{W}^2} |\bar{c}\,(w_A, w_B)\,|$ *be the largest possible gain from migration. If the grid for costs is fine enough, there will always be a migration costs grid-point smaller than* $C_{\max}$, *since migration costs can be arbitrarily close to zero. Hence, there is some* i *such that* $d_i > 0$ *holds if the grid is fine enough.* ■

Lemma 10 *For any distribution of costs with support equal to* R_+, B^2 *has only positive entries.*

Proof. *We obtain for* B^2

$$B^2 = BB = \begin{pmatrix} ((I - D_A)\,\Pi)^2 + D_B\Pi D_A\Pi & (I - D_A)\,\Pi D_B\Pi + D_B\Pi\,(I - D_B)\,\Pi \\ (I - D_B)\,\Pi D_A\Pi + D_A\Pi\,(I - D_A)\,\Pi & ((I - D_B)\,\Pi)^2 + D_A\Pi D_B\Pi \end{pmatrix}. \tag{33}$$

Each entry of this matrix is weakly positive, since all three $(I - D_i)\,, D_i$ *and* Π *are positive. Hence we only need to argue that in each sum at least one part is always strictly positive. For the elements on the diagonal this follows directly from* $(I - D_i)\,\Pi > 0$. *For the off-diagonal elements, we may have some rows of zeros in* $D_i\Pi$. *However, at least one row of* $D_i\Pi$ *will be non-zero, because there is some non-zero* d_i *and* $(I - D_i)\,\Pi > 0$, *so that all elements of* $(I - D_i)\,\Pi D_j\Pi$ *are strictly positive.*

Proposition 11 *Under the assumptions of the above Lemmas, migration and idiosyncratic shocks define an ergodic process with stationary distribution* $F_0 = \lim_{n\to\infty} B^n e_i$.

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

■

8.3 Data

Data on migration between US states are provided by the U.S. Internal Revenue Service (IRS). The IRS uses individual income tax returns to derive internal migration between US states. In particular, the IRS compiles migration data by matching the Social Security number of the primary taxpayer from one year to the next. The IRS identifies the households with an address change from the previous year, and then totals migration to and from each state in the U.S. to every other state. Given these bilateral migration data we compute aggregate gross inmigration for the 51 US states (including District of Columbia) as the sum of all inmigrations from other US states to a particular state. Migration rates are calculated by expressing gross inmigration as proportions of the number of non-migrants reported in the IRS dataset.

Income per capita data are from the Regional Economic Information System (REIS) compiled by the Bureau of Economic Analysis. 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.

In the estimations, we remove a linear time trend from the data and express all variables as deviations from their unit-specific means, i.e. we apply a within-transformation. Table 7 reports descriptive statistics for the original as well as for the transformed data.

In order to examine the time-series properties of the employed data we performed a brief unit root analysis for the migration rates, the income per capita, and the income per capita in the complementary region. In a sample of this size ($T = 16, N = 51$) either a Breitung and Meyer (1994) or a Levin, Lin, and Chu (2003) unit root test appears 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. Similarily, the Levin, Lin, and Chu (2003) rejects the null hypothesis of a unit root taking a linear time trend into account.

References

[1] Adda, J. and Cooper, R. (2003): "Dynamic Economics: Quantitative Methods and Applications", MIT Press, Cambridge.

Table 7: 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

[2] Baxter, M. and M. Crucini (1993): "Explaining aving/investment correlations", American Economic Review, 83, 416-436.

[3] Backus, D., P. Kehoe, and F. Kydland (1994): "Dynamics of the trade balance and the terms of trade: the J-curve?", American Economic Review, 84, 84-103.

[4] Blanchard, O. and L. Katz (1992): "Regional evolutions", Brookings Papers on Economic Activity, 1, 1-75.

[5] Burda, M. (1993): "The Determinants of East-West German Migration: Some First Results", European Economic Review, 37, 452-462.

[6] 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.

[7] Borjas, G.J. (1987): "Self-selection and the earnings of immigrants", American Economic Review, 77, 531-553.

[8] 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.

[9] 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.

[10] 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(8), 898–914.

[11] Cushing, B. and J. Poot, (2004): "Crossing boundaries and borders: Regional science advances in migration modelling", Papers in Regional Science, 83, 317 - 338.

[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] Gourieroux, C., A. Monfort, and E. Renault (1993): "Indirect inference", Journal of Applied Econometrics, 8, S85-118.

[14] Greenwood, M.J. (1975): "Research on internal migration in the United States: A survey", Journal of Economic Literature, 13, 397-433.

[15] Greenwood, M.J. (1985): "Human migration: Theory, models and empirical studies", Journal of Regional Science, 25, 521-544.

[16] Greenwood, M.J. (1997): "Internal migration in developed countries", In: Rosenzweig, M.R. and O. Stark (eds.), Handbook of population and family economics 1b, Elsevier, Amsterdam.

[17] 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.

[18] Kennan, J and J. R. Walker (2005): "The Effect of Expected Income on Individual Migration Decisions", mimeo, Universitiy of Wisconsin.

[19] Sjaastad, L. (1962): "The costs and returns of human migration", Journal of Political Economoy, 70, 80-93.

[20] Smith, T. (1993): "Estimating nonlinear time-series models using simulated vector autoregressions", Journal of Applied Econometrics, 8, S63-84.

[21] Tauchen, G. (1986): "Finite state Markov-chain approximation to univariate and vector autoregressions", Economics Letters, 20, 177-81.

[22] Tunali, I. (2000): "Rationality of Migration", International Economic Review, 41, 893-920.