Testing the Real Options Approach to Migration

by

Damiaan Persyn

A thesis submitted to

Prof. Dr. Wolfgang Härdle
Institute of Statistics and Econometrics

in partial fulfilment of the requirements
for the degree of

Diplom-Volkswirt

Humboldt Universität zu Berlin
Wirtschaftswissenschaftliche Fakultät
Spandauer Strasse 1
D-10178 Berlin

Berlin, October 22, 2002

Selbständigkeitserklärung

Hiermit erkläre ich, daß ich die vorliegende Diplomarbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, 22. Oktober 2002
Damiaan Persyn

Acknowledgements

Writing this thesis would not have been possible without the help of following people:

Prof. Wolfgang Härdle, Axel Werwatz, Marlene Müller and Zdenek Hlavka
(Institute for Statistics and Econometrics, Humboldt University Berlin)

Prof. Michael Burda
(Institute for Economic Theory II, Humboldt University Berlin)

Sabine Kallwitz
(Deutsches Institut für Wirtschaftsforschung, Berlin)

The people hanging around in the Institute of Statistics and Econometrics' library, for providing helpful tips.

Merlind Harms, for her help in manipulating some of the figures.

Barbara Dietz (Osteuropa Institut München), for allowing the use of the dataset on ethnic Germans.

Special thanks to Lilo Locher (Institute for the Study of Labour, IZA) for allowing the use of her prepared version of the dataset on ethnic Germans and providing helpful comments.

Contents

List of Figures

List of Tables

List of Abbreviations

DIW	Deutsches Institut für Wirtschaftsforschung
FRG	Federal Republic of Germany
GAM	Generalised Additive Model
GDR	German Democratic Republic
GEE	Generalised Estimation Equation model
GLM	Generalised Linear Model
GPLM	Generalised Partial Linear Model
GSOEP	German Socio-Economic Panel
LME	Linear Mixed Effect model
NPV	Net Present Value (model)
PV	(Gross) Present Value
ROT	Real Options Theory
ROTM	Real Options Theory applied to Migration

Introduction

The fall of the iron curtain has brought considerable increases in earnings for most individuals living in East European countries. However, a considerable wage gap between the former communist countries and their West European neighbours still exists and rapid convergence seems out of sight.
In this light, it is surprising to see only a very small number of people decide to migrate to the West. Even in the case of German East-West migration, where there exist no legal barriers and only relatively small cultural barriers, this seemingly irrational behaviour is observed. Therefore, the main empirical puzzle of migration behaviour seems explaining immobility rather than mobility. In the course of time, different models have been developed to explain migration, but a totally satisfactory explanation for the observed immobility remains to be found and tested.
The fact we can not explain the immobility of labour seems a rather important issue, given the prominent role of factor mobility in a variety of macro-economic models. Migration theories could also contribute to the currently ongoing discussion on the enlargement of the EU and fears of mass migration from the new member states.

Burda [6] applied the real options theory, which was initially developed in the context of firm investment, to migration. As one of the main results of real options models is explaining how it can be rational for a firm/individual to wait to invest/migrate, this seems a promising route. In this thesis, I will give a survey of this theory, develop feasible tests for it, look at theoretical and practical difficulties an empirical test would face and finally apply the methods to two different datasets.

The thesis is organised as follows:

In the first section, the real options theory is developed in a dynamic programming framework, as suggested by Burda [6]. After formally deducing the equations describing the migration behaviour, a short overview of some different migration theories is given. The section concludes with a discussion on why and how different elements of these other models could be included in a real options framework, in order to obtain a functioning migration model.

The second section investigates the fact the relevant independent variable of the real options theory -the subjective estimation of the wage gain from migration- is not observable. A possible solution that is investigated, is to use wages of individuals that migrated in the past, to estimate the wage gain from migration individuals anticipate. As migrants are a nonrandom group, the problem of self-selectivity has to be addressed.

In the third section, some falsifiable predictions of the real options approach to migration are deduced. The importance of finding predictions that are unique to the real options approach is stressed. A first test uses the predicted nonlinear dependence of the migration propensity on the anticipated income gain from migration. This is the method used by Burda, Härdle, Müller and Werwatz [7]. However, by putting together the findings from the first sections, quite different results as to the exact form of the nonlinearity that is predicted by the real options approach to migration are obtained. A second test uses the negative dependence of the migration propensity on increases in the uncertainty of the anticipated income gain from migration.

In the final section, the tests are applied to real world data. Two datasets are used for this. The first set contains data from the German Socio-economic Panel from the 'Deutsches Institut für Wirtschaftsforschung', which we use to investigate German East-West migration and migration intentions. The second dataset is from the 'Osteuropa-Institut' in Munich. It contains data on migration intentions of ethnic Germans in Russia considering migrating to Germany. This dataset will be discussed only briefly.
For the first test, the results from the empirical part are in contradiction with the predictions of both the real options theory and more classical migration models under the assumption that migration wage gains are proportionally higher for individuals with a higher income. However, the addition of a model for the anticipated wage gain from migration was found to explain the observations rather well, without requiring a real options framework.
The second test showed mixed results for the German East-West migration dataset.

We conclude that with the developed methods and datasets, we find only little support for the real options approach to migration.

1 The Migration Decision in an Uncertain Environment

Burda [5] was the first to analyse migration using the 'real options' approach, as suggested by Dixit and Pindyck [11]. The Real Options Theory (ROT) has proven a successful analytic framework for some very different applications and migration decisions seem to fit the assumptions of the model quite well.

In this section, I shall try to explain how the ROT generally works and follow Burda [6] to describe how it could be applied to the individual's migration decision.

1.1 Real Options Theory

1.1.1 Background

Real options theory [31; 36; 11; 12; 13; 37] helps to explain why firms often delay decisions in situations where classic 'Net Present Value' analysis (NPV) would predict immediate action. The major difference with the older models is that in the latter, the firm is modeled as choosing just between acting now or to restrain forever. Quite often, however, a firm also has the option to delay decisions. Under specific conditions -specified later on- there can be some value attached to this option of waiting and since this value is lost upon acting, it poses an additional opportunity cost, thereby enlarging the region where a firm will delay action. This 'increased lethargy' is exactly what is observed with migration in general and with German East-West migration in particular: given the persistence of regional income differentials and the relatively low costs attached to migration, only an astonishingly small number of East-Germans moved West in the years following German unification in 1990 [1]. The ROT explains how waiting to migrate can be a perfectly rational decision in this context.

As the ROT was originally developed to describe firm investment behaviour by linking it with financial option theory and we are using it to describe migration behaviour, it is easy to get confused by the different terminology used in each literature. If a certain income differential between regions is reached, our theory

[1] see figure 1 in Burda citeBurda95

will predict 'migration'. In alternative frameworks, one would say an individual 'invests' in an uncertain future income stream, or 'exercises the call option' she has on a migration project. Table 1 illustrates how the concepts of the different frameworks connect. Let us now first study real options in their natural habitat

Migration	Investment	Financial options
PV of migration	PV of project's revenue	Stock price
Migration Costs	Investment Costs	Exercise price
Time window available for migration	Investment time window	Time to expiration
Migrant's time preference	Time value of money	Risk-free rate of return
Volatility in the income differences	Riskiness of the project's revenue	Variance of returns on stock

Table 1: Analogy between the different option approaches

of the firm's investment decision, before translating to the case of migration in section 1.2.2.

To illustrate how the ROT works, we will start with a small example, consisting of the bare minimum to let the ROT work. The assumptions we need to make are features common to a wide variety of investment projects:

- At least partially sunk costs.
- Ongoing uncertainty in the economic environment.
- The possibility to postpone the decision.

1.1.2 A simple example

As long as these conditions are satisfied, waiting has a value, as can easily be understood looking at this short example, frankly stolen[2] from Dixit [12]. Compare a risk-neutral firm facing an investment of the 'classical' type, where the investment opportunity is unique, and a 'real options' investment, where postponement is possible. With F and V we denote the sunk cost and current revenue of the project respectively. Let V follow some discrete Markov process, where the current value of V is the best predictor for future values. Then we can write the expected present value of the stream of revenues from investing in period 0 as $V_0 \frac{1+r}{r}$, with r some discount rate. Following classical Marshallian

[2] "Lesser artists borrow, great artists steal." — Igor Stravinsky

reasoning, a firm facing a now or never decision, will invest if $V_0\frac{1+r}{r} \geq F$. But suppose the firm could postpone the decision for a fixed interval of time and observe the new V, say V_1. Now assume the firm's new strategy is not to invest in period 0, and to invest in period 1 only if $V_1 > M$. As Dixit notes, it is evident this strategy is not itself optimal, but for now, we only want to prove waiting can have a value when V_0 equals the minimum expected net present value of revenue where Marshallian reasoning predicts action. We will call this level the Marshallian trigger or NPV trigger henceforth. In the first period, the net worth of the project was zero, since V_0 was equal to the marshallian trigger. If now $V_1 < M$ in the second period, the net worth also equals zero -were are not *obligated* to invest-, but if $V_1 > M$ the firm invests and receives some positive expected revenue. Therefore this strategy is always an improvement over the now-or-never case. The value of waiting appears, because by waiting the firm can avoid the downside risk, while realising the upside potential.

1.1.3 A formal expression for the value of waiting

Let us now try to describe this value of waiting more formally, following Dixit with Pindyck [11], Dixit [13] and Burda [6]. Burda tackles the problem through its analogy with option valuation. It is easy to see the connection between the investment decision the firm faces and a dividend paying American call, as we saw in table 1: An investment opportunity (call), gives a firm (the owner) the right to invest (exercise) at any time to receive some project (stock) that pays some uncertain revenue stream (dividends). While this approach is much easier and reasonable for firm investment, it is much less justifiable in the case of migration: to get a valuation for the option to invest, we would have to use some arbitrage argument. For financial options, this is a natural approach, given the existence of well developed derivatives markets. For the case of firm investment, the assumption seems acceptable, as a firm would, at least in principle, be able to sell its 'option to invest'. Alternatively one could argument that in a reasonably functioning market, a firm not valuing its projects correctly would be competed out. But for migration, arbitrage is not possible, as individuals can not sell options to migrate.

The approach we will take, dynamic programming, is described in Dixit and Pindyck [11]. It does not require such assumptions. With dynamic programming, the firm's maximisation problem is tackled by breaking the sequence of decisions the firm has to make over time into just two components: the immediate decision and a valuation function, valuing all subsequent decisions, starting with the position that results from the immediate decision. We will now extend our first example using this framework. Before turning to continuous time, we examine a model in discrete time. Uncertainty enters the model through changes in the price W of some commodity which determines the return of the investment project; eg. the price of copper for an investment in a copper mine or the wage differential for migration decisions. For now, suppose W in period 0 is some W_0 and changes to $(1+u)W_0$ with probability q or $(1-d)W_0$ with probability $(1-q)$ from period 1 on to eternity: W changes only between period 0 and 1, remaining constant on the period 1 level for all folowing periods. Again, look at the firm, facing a now or never decision in period 0. The NPV of investing in period 0 is

$$\begin{aligned} NPV_0(W_0) &= W_0 + \Big[q(1+u) + (1-q)(1-d)\Big]W_0\Big[\frac{1}{1+r} + \frac{1}{(1+r)^2} + \cdots\Big] - F \\ &= W_0 + \big[1 + q(u+d) - d\big]\frac{1}{r}W_0 - F \qquad (1) \end{aligned}$$

which gives us for the value of the project in the now or never case:

$$\Omega_0(W_0) = \max[NPV_0, 0]. \qquad (2)$$

Again, this value is always positive, because the firm does not have to invest. Ω is called the *stopping value*, since it is what a firm gets when it stops waiting and invests (getting the NPV) or abstains (and gets no payoff). The NPV of investing in period 1 is

$$\begin{aligned} NPV_1(W_1) &= W_1\Big[1 + \frac{1}{1+r} + \frac{1}{(1+r)^2} + \cdots\Big] - F \\ &= W_1\frac{1+r}{r} - F \qquad (3) \end{aligned}$$

which gives us for the net payoff in period 1:

$$V_1(W_1) = \max[NPV_1, 0]$$

This is the *continuation value*, the value of the future optimal decision. Since

W_1 is unknown in period 0, we have to take expectations:

$$\begin{aligned}\mathcal{E}_0[V_1(W_1)] &= q\max[(1+u)W_0(1+r)/r-F,0] \\ &\quad + (1-q)\max[(1-d)W_0(1+r)/r-F,0] \end{aligned} \tag{4}$$

And from this, we can calculate the net present value of the investment project with the possibility to wait, in period 0:

$$V_0(W_0) = \max\left\{\Omega_0, \frac{1}{1+r}\mathcal{E}_0[V_1]\right\} \tag{5}$$

If a firm wants to maximise its profits, it chooses whether to invest or not by comparing the two arguments of this maximisation problem. The choice made in period 0 will depend on the values of NPV_0 and $\mathcal{E}_0[V_1]$ and therefore essentially on the value of W_0 and knowledge of the stochastic process determining W_1. Comparison of V_0 and Ω_0 makes clear the former will always be larger or equal than the latter, illustrating in a more general context the point made in the very first example, that waiting *always* has a value. What also becomes clear, looking at (5), is that sometimes not waiting will have an even larger value and acting becomes the optimal decision. Equation (5) is called a Bellman equation, where we can clearly see the decomposition principle of dynamic programming at work; Bellman's principle of optimality states that *an optimal policy has the property that, whatever the initial action, the remaining choices constitute an optimal policy with respect to the subproblem starting at the state that results from the initial actions.*

If we now change the underlying stochastic process from a one-time jump from period 0 to period 1, to a simple discrete time random walk process, we can write the value function V_t in some period t as

$$V_t(W_t) = \max\left\{\Omega(W_t), \frac{1}{1+r}\mathcal{E}_t[V_{t+1}(W_{t+1})|W_t]\right\} \tag{6}$$

If we assume an infinite horizon, which we will from now on, the 'absolute time position' is no longer relevant; eg. it is not important how much time passed since $t=0$, to describe the optimisation problem in some time period, since the problem now looks exactly the same for all time periods. Hence we can drop the time labels. This gives us for (6):

$$V(W) = \max\left\{\Omega(W), \frac{1}{1+r}\mathcal{E}[V(W'|W)]\right\} \tag{7}$$

where W' denotes the level of W in the next time period: $W' = W + \Delta W$. In this example and also in the application to migration later on, there will be only one value of W, W^* that separates the areas where optimal policies differ [3]. For $W < W^*$ the left side of (7) will maximise V, for all other values it will be the right side. For compatibility with Burda [6] and Burda, Härdle, Müller and Werwatz [7] we will write H for W^* from now on.
Therefore we know that

$$\begin{aligned} \forall\, W \geq H: &\qquad \Omega = V \\ \forall\, W < H: &\qquad V(W,t) = \frac{1}{1+r}\mathcal{E}[V(W'|W,t)]. \end{aligned} \tag{8}$$

Now let us finally move to continuous time! Reasoning in continuous time greatly facilitates calculations, but we should remember that continuous time probably is a bad approximation of reality for firm investment and even more so for migration decisions. People will not evaluate their migration opportunity every split second, quite on the contrary. Nevertheless, we will continue in continuous time for ease of use and make remarks on the validity when necessary [4]. Looking at V in the 'continuation area', where $W < H$, the length of the time interval we are considering, say Δt, makes that total discount rate over this interval now becomes $(1 + \Delta tr)^{-1}$.

$$\begin{aligned} V(W) &= \frac{1}{1+\Delta tr}\mathcal{E}[V(W, t+\Delta t|W)] \\ rV(W,t)\Delta t &= \mathcal{E}[V(W+\Delta W, t+\Delta t|W) - V(W)] \\ rV(W,t) &= \mathcal{E}[\frac{1}{\Delta t}(V(W+\Delta W, t+\Delta t|W) - V(W))] \end{aligned} \tag{9}$$

where $V(W)$ can enter the expected value because it is known in t. Letting

[3] See appendix B of Dixit and Pindyck [11] to learn more about this; for us, the necessary condition is that $\Omega(W)$ is increasing in W -which is natural to assume for our application to migration later on, cf. the definition of Ω- and that the underlying stochastic process is persistent, which seems a reasonable assumption for the income differential and we will therefore always assume to hold.

[4] There is an important issue hidden here: in the discrete-time model it was evident decisions in period t are independent of knowledge in future periods; in continuous time, however, we need the process for the state variable to be continuous from the right and strategies to be continuous from the left to assure this.

$\Delta t \to 0$ and dropping the W argument, gives:

$$rVdt = \mathcal{E}[dV] \tag{10}$$

Equation (10) nicely illustrates the equivalence of the dynamic programming approach and option valuation theory: if we would interpret V, the value of our investment project, as a call option, a rational and risk-neutral agent would only be willing to hold the option if it yields an expected return equal to the 'normal return' over the period dt: $rVdt$. We could have obtained equation (10) much faster through the financial options approach, but, as mentioned before, with dynamic programming we do not need to make assumptions that would seem unnatural in the application to migration.
Substituting (10) in the value function (7) yields

$$V = \max\left\{\Omega, \frac{1}{rdt}\mathcal{E}[dV]\right\} \tag{11}$$

which is exactly what we were looking for: a functional form, describing the value of the investment project with the possibility to wait, when it is optimally managed.

1.1.4 Solving for the value function

To solve for the value function, we have to define what stochastic process drives the state variable W. Following Burda [6], we model it as a simple brownian motion:

$$dW = \mu dt + \sigma dz \tag{12}$$

where dz is an increment to a Wiener process and μ is the trend. Unfortunately, V is now no longer differentiable using standard techniques. We have to make use of Itô's lemma, which gives for dV

$$dV = V'dW + \frac{1}{2}V''dW^2 \tag{13}$$

and since we need $\mathcal{E}[V]$ this becomes:

$$\begin{aligned}\mathcal{E}[dV] &= V'E[dW] + \frac{1}{2}V''E[(dW)^2] \\ &= V'\mu dt + \frac{1}{2}V''[\sigma^2 dt + \mu^2 dt^2] \\ &= V'\mu dt + \frac{1}{2}V''\sigma^2 dt + O(dt).\end{aligned} \tag{14}$$

If we substitute this in (11), we obtain for the value function:

$$V = \max\left\{\Omega, \frac{1}{rdt}[V'\mu dt + \frac{1}{2}V''\sigma^2 dt]\right\} \tag{15}$$

and therefore, in the continuation area ($\forall\ W < H$:)

$$rV = V'\mu + \frac{1}{2}V''\sigma^2 \tag{16}$$

Which is equation 14 of Burda [6]

Now we should point to the fact we have ignored the effect on the functional form of NPV when we changed the underlying stochastic process for W in previous steps. But now we are trying to solve for V, we have to recalculate it explicitly: in continuous time, the present value of expected future returns becomes $\int_0^\infty e^{-rt}\mathcal{E}(W_t)dt$. With $\mathcal{E}(W_t) = W_0 + t\mu$ we get $\int_0^\infty e^{-rt}\mathcal{E}(W_t)dt = (W_0 + \mu/r)/r$ and therefore

$$NPV = (W + \mu/r)/r - F \tag{17}$$

and the marshallian trigger:

$$M = rF - \mu r. \tag{18}$$

Solutions of (16) will be of the form

$$V(W) = Ae^{\beta_1 W} + Be^{\beta_2 W} \tag{19}$$

with β_1 and β_2 solutions of $\frac{\sigma^2}{2}\beta^2 + \mu\beta - r = 0$:

$$\begin{aligned}\beta_1 &= \frac{1}{\sigma^2}(-\mu + (\mu^2 + 2\sigma^2 r)^{\frac{1}{2}}) > 0 \\ \beta_2 &= \frac{1}{\sigma^2}(-\mu - (\mu^2 + 2\sigma^2 r)^{\frac{1}{2}}) < 0\end{aligned} \tag{20}$$

To solve this equation, we need to make additional assumptions. For our partic-

ular application, one of them turns out to be $\lim_{W\to-\infty} V = 0$. Indeed, it seems natural to assume the value of the investment project goes to zero if the state variable W makes the NPV very negative, even if all procrastination possibilities are included. This is only possible with $B = 0$. Since we do not need β_2 any longer, we will simply write β for β_1 from now on.
Additionally, we know from the Bellman equation that in the stopping region $V = \Omega$, so by continuity we also have $V(H) = \Omega(H)$.
Since H is an unknown endogenous variable, we need to impose a third condition. For economic applications, we need to require not just the value of the functions V and Ω to be equal at H but also their first derivatives.[5]
To begin with the 'equal derivatives' condition (called 'smooth pasting' or 'higher order contact' condition), we get for the continuation region:

$$V'(H) = A\beta e^{\beta H} \tag{21}$$

And in the stopping region (where we know $\Omega = (W + \mu/r)/r - F$):

$$V'(H) = \Omega' = \frac{1}{r} \tag{22}$$

And therefore $A = \frac{e^{-\beta H}}{r\beta}$. If we substitute this in V's definition for the continuation region we get:

$$V(H) = \frac{e^{\beta H}}{e^{\beta H} r\beta} = (H + \mu/r)/r - F \tag{23}$$

From this, finally, we find for H

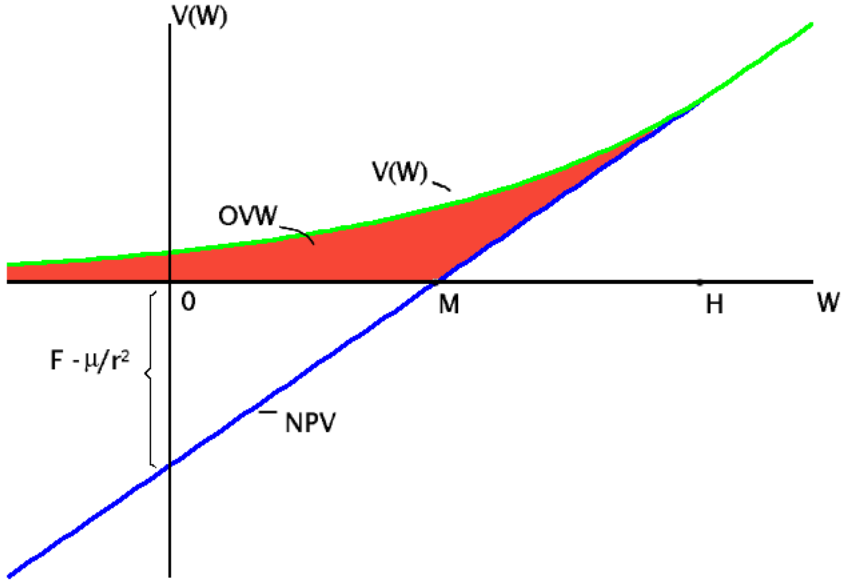
$$H = \frac{1}{\beta} - \mu/r + rF \tag{24}$$

and also the solution for the value function:

$$V(W) = \begin{cases} \frac{1}{r}(W + \mu/r) - F & \text{for } W \geq H \\ \frac{1}{r\beta} e^{\beta(W - rF + \mu/r) - 1} & \text{for } W < H \end{cases} \tag{25}$$

Figure 1 shows the value function in function of W. Here too, the analogy with

[5] An oversimplified explanation for this is that if it would not be the case, the value function would 'kink' at $W = H$ and this would allow gains over V. This should not be possible, given V's definition. Consult Dixit and Pindyck [11], appendix C of chapter 4 for a more technical treatment.



source: own calculation

Figure 1: The value function; beyond the trigger level H, V equals the NPV. The red surface is Burda's OVW.

option theory is clear: people familiar with the subject will recognise $V(W)$ as the value of a continuous dividend paying American call option in function of the underlying spot price (see, for example, Franke, Härdle and Haffner [18], p. 116). Perhaps the mapping between the different approaches in table 1 is easier to understand after this introduction to the dynamic programming approach of the ROT.

As noted above, there are sound arguments for using the dynamic programming approach and ignore option valuation theory in the ROTM, since there are no markets a migrant could sell his 'option to migrate', in contrast with financial options and the real options approach to firm investments for a wide variety of applications. The results, however, are the same.

1.2 How real options theory could be used to understand migration

What makes the ROT suited for analysing migration, is the fact the minimal assumptions that we made for the ROT to work in the context of firm investment;

- At least partially sunk costs.
- Ongoing uncertainty in the economic environment.

- The possibility to postpone the decision.

seem equally valid in the case of migration. One could then use the mapping suggested in table 1 to obtain a functioning model. But before argumenting *how* the ideas we've summarised in the previous subsection can be used to build a migration model, it is worth considering *why* real options theory can offer a contribution to this field of research. As written above, an attractive feature of the real options approach to migration is that it would offer an answer to a major empirical puzzle of migration behaviour: the prevalence of immobility, despite the existence of huge wage differences between regions. In section 3 we will see the ROT makes other interesting predictions about migration behaviour and intentions (we will call the ROT, when applied to migration, ROTM from now on), but apart from explaining phenomena, one could argue the theory is interesting on its own; indeed, it sounds evident that, in general, migrants do not consider migration as a now-or-never decision, but know procrastination is an option, in a wide variety of situations. So on first sight, it surprises to see only relatively little attention for the ROTM in the migration literature. In fact, the only articles and working papers I know of that (some more, some less) discuss the theory itself are Burda's 1993 and 1995 articles [5; 6], Burda et al.[7], O'Connell [34], Locher [29], Parikh and Leuvensteijn [35] and the not so new approach of Khwaya [26]. In chapter four of his well written book, Peter A. Fisher [15] discusses different microeconomic approaches to explaining (im)mobility. Only one paragraph (pp. 144) briefly describes the ROTM.[6] So why is the theory so neglected? One reason could be that there are quite a few other migration models that can produce the observed insensitivity to income differentials; actually almost every remotely realistic migration theory deviates from the portrait of neo-classical economics of '*the marginal man juggling with a pocket calculator to compute present values of investment moves to alternative locations'* (Böhning quoted in Fisher [15]), that would predict immediate action on positive NPV as in some of the examples we looked at before. There exists a large number of microeconomic migration models that make sense and offer very reasonable explanations for the observed lethargy and other observed migration

[6] there is also a mistake in the paragraph: Fisher is obviously wrong when he states that *[...] For risk-neutral decision makers it can be reasonable to delay the 'go' decision in order to reduce the uncertainties involved in migrating by collecting more relevant information.[...]. Such behaviour is only beneficial, though, [...] as long as the income differences between origin and destination are not increasing.* The reader can verify that even with $\mu > 0$ there exists a value of waiting.

characteristics which do not require knowledge of dynamic programming or option valuation theory. Therefore, these theories probably are more attractive for economists in this field.
In the next section, 1.2.1, I shall briefly describe the key concepts of some important relevant migration theories, before looking at what real options theory could contribute to the discussion. The first two columns of table 1 illustrated the analogy between migration and firm investment, but to obtain a functioning model we will need to give content to terms like the cost and revenue of migration, time preference,... Other migration models also give hints on what variables could be used for this and in subsection 1.2.2, I will discuss how a functioning real options model for migration could be constructed, combining our findings from the ROT in the context of firm investment of section 1.1 and elements of the migration models we will now discuss.

1.2.1 Relevant alternative migration theories

Since it would be both impossible and quite useless to give a complete overview of the migration literature in this thesis, I decided to look only at those microeconomic theories relevant to our discussion. This section borrows heavily from Fisher [15] and Davanzo [9], who give an overview of the different microeconomic approaches that were developed over time to describe migration behaviour. I will give some remarks on whether it is desirable to try to integrate the main idea behind a specific model into the real options approach of migration and our applications of it to German East-West migration and migration of ethnic Germans.
Perhaps it is interesting to follow Fisher's approach and look at the different models as deviations from a super-neoclassical model. The assumptions of this model are:

- Migration is cost free
- Migration is risk free
- Potential migrants are a homogeneous group of people
- Potential migrants have access to perfect and free information
- Potential migrants behave in an unconditionally rational manner

- The potential migrant is an autonomous human being with no social context

The simple NPV-calculation from the first subsection already included *costs* and *inhomogeneity* of the subjects. Apart from some real fix cost attached to migration, one could include more intangible factors like cultural differences, political or linguistic barriers,... as is often done in migration models. This could help explain low international migration flows between certain countries or regions, but seems to contradict low German East-West migration, since in this case all those costs are probably rather low. Accounting for overall migration costs alone cannot explain why someone migrates while somebody else does not. Therefore one should also consider migration costs that apply to a single individual. For example: apparently it makes sense that an increasing number of children in a household accounts for higher migration costs.

This brings us to the homogeneity assumption: any reasonable microeconomic migration model will recognise *the group of possible migrants is inhomogeneous*, apart from differences in costs: It is clear that most other elements of the model, such as preferences, information, risks, costs and so forth will differ among different subjects. Therefore, detailed information is needed to construct proxies for costs and benefits attached to migration on the individual level.

As Davanzo [9] notes, the benefits (and perhaps some of the costs) of migration may accrue over a period of time. This is explicitly modeled in human capital models of migration. They can help explain how different time preferences can influence migration behaviour or why migration rates decline with age (for example by modeling the revenue stream to stop at some time T or by assuming people accrue location specific human capital over time which is lost upon migration). We could have developed the real options approach in a finite time framework, which would produce these effects automagically; but to keep things as simple as possible, one could simply include proxies for variables which connection with the migration propensity would be affected by setting a time limit, in the costs, as in Burda et al. [7].

The *risk* associated with migrating to another region was introduced in the migration literature in Harris and Todaro's [22] famous 1970 article; they developed a model where migrants do not only care about the level of achievable income, but also about the probability of realising it. In this context, vacancy rates and unemployment levels in both the region of origin and destination become important. In the context of our application this makes clear why we cannot simply model migration intentions as a function of the income difference between regions, as people will base their behaviour not on existing wage differences, but rather on their subjective anticipations. And these will depend on their individual estimation of the wage differential, the probability of success in finding a job in the region of destination, the wage development in the two regions, probability of job loss and so forth. We will denote the subjective estimation of the wage gain to migration of some individual, as her 'anticipated' wage gain or simply 'anticipations' to distinguish them from the statistical expectations operator. (as in Tunali [46]). To estimate these anticipations is quite an impossible task, but unfortunately at the same time, it also seems vital to successfully apply the theory, so this is probably the Achilles heel of any real options approach to migration. In Burda et al. [7] this problem was recognised, but since the paper concentrated on technical side of the semi-parametric estimation it was decided to include income from the East only, as the determinant of the anticipated wage gain, 'rather than producing spurious findings based on biased estimates of the West-East income differential[...]', which, of course, makes perfect sense. A few alternative approaches and the problems related to them were discussed by Burda et al.; we will take a closer look at them in section 2. We will also investigate if our results sufficiently deviate from the assumption made in Burda et al. to justify the addition of a (probably rather untrustworthy) model for the anticipated wage gain in section 4.

So far we only considered risk-neutral subjects, eg. uncertainty was coped with by reasoning with expected values in stead of real values, but the associated utility levels remained the same. Of course one could suggest the observed reluctance to migrate is simply caused by risk averseness of the possible migrants, but once again this seems to apply less to German interregional migration: it is probably less of a gamble moving from Dresden to Munich than, say, from Marakesh to New York. Also the legislative system is the same for the whole country, which for example makes that a migrant knows in advance how much unemployment benefits he can receive in the destination region, would he not be able to find a job there. For our application to the Russian ethnic Germans, this issue seems more relevant, but we will ignore it.

In order to estimate the migration determinants we discussed, like income differentials, job opportunities and the risks involved with migration, an East-German would need lots of *information*. A lot of empirical research on migration with incomplete information has been done. Generally speaking, since information collection and processing is costly, including such costs in migration models will make them predict less migration. There are important issues connected to information flows and migration we will ignore here; incomplete information also challenges the assumption of *unconditionally rational behaviour* on which all economic analysis is based. The addition of a model to describe how possible migrants make anticipations about migration in section 2 will allow for some degree of 'irrationality' in our approach, as migrants are allowed to make wrong anticipations of the gains from migration. But conditional on these subjective anticipations, we will assume rational behaviour. Also related to information costs are labour market matching models, which can explain how larger immigration and emigration flows between regions without larger net flows can exist. These incomplete information models are such a different class of models -using game theory, stopping rules,...- that they would be hard to include in a real options framework. But once again, in the context of German East-West migration, we probably have more important issues to tackle, since information is probably relatively easy to obtain for migration candidates if one considers the overall presence of national media and the absence of language and other barriers... Once again, this seems less so for the ethnic Germans dataset.

Since the *social independence* assumption of the neoclassic migration model seems both unrealistic and highly relevant for migration decisions, it received considerable attention in the literature. Individuals strongly attached to someone else are likely to have additional opportunity costs related to migration. Moreover, the information search problem becomes more complicated since now two or more matches have to be found in the region of destination. On the other hand, it can be a risk diversifying strategy to spread different household members over different geographical locations, or it can be a way to access some capital for investments, but this seems to be an issue more relevant to third world migration. Family re-unification can also be a drive for migration, this could be dealt with in our context by inclusion of a proxy for the number of relatives a migration candidate has in the region of destination. Also the household could be considered as the decision unit instead of the individual.

Wealth may also influence the migration decision: since most costs of migration must be incurred before returns can be reaped, some level of wealth will be needed to be able to migrate. Banks are unlikely to lend for migration, since hardly any collateral can be offered. This would cause only the local income level to matter for some low levels of income; income would then have a positive effect on the migration propensity for these levels [7].
So far we have looked at pure and simple comparative-static micro models only. There are, however, important dynamic aspects connected to migration. In economics, migration is commonly seen as a reaction to income differences between regions and in this spirit it is also an instrument of arbitrage: labour flows should correct disequilibria in labour markets. Burda [6] considers this issue in his 1995 paper. Even without looking at the macro level, one could/should think about the effects migration has on the determinants of migration itself; one could argument that migrants make cheaper information about the destination region available to possible migrants at home. This would cause migration to accelerate more migration. We will simply ignore macro-effects, since the total number of migrants is so small it is unlikely to have important feedback effects on the macro level. On the micro level there can be feedback effects through, for example, agents with relatives or friends in the West having better information about the region of destination (network models of migration).

1.2.2 Elements of alternative models we retain

It should be clear that including too many elements from these migration models in a ROTM framework would be neither technically possible nor desirable. It would also greatly complicate the interpretation of empirical tests of the model. Therefore, we should strive to include only those elements we think would greatly bias our findings if left out.

Since Ravenstein's early work on the determinants of migration [39], income differences between regions have been widely recognised as an important force

driving migration. So instead of the decision problem of the firm evaluating her investment decision in function of the current revenue W, we could now look at the decision problem of a possible migrant deciding in function of the current income difference between her current location and some other region. Unfortunately, neither we nor the possible migrants can readily observe how much they would earn in the region of destination. Therefore, we have to assume possible migrants reason in function of their anticipations of the individual income difference between the location, in stead of real or estimated differences.
In the previous subsection we assumed a simple Brownian motion with drift for the statistical process driving the 'underlying': $dW = \mu dt + \sigma dz$. For income differences this seems an acceptable assumption. One could argue a geometric process to be more suitable, if one believes income differences (and therefore probably anticipations individuals make of them) between East and West to be bounded by zero; or perhaps a jump process if one believes some macro-economic shocks influence the regions asymmetrically. All this can be dealt with within the real options framework by appropriate recalculations. The main findings, however, like the shift of the trigger value, remain valid as long as the process exhibits positive first order correlation. Therefore we will continue to use the simple Brownian motion with drift as the underlying statistical process.
Since the subjects can not observe W, they somehow will make some subjective estimations of it. Because this cannot be handled in a ROT model (for investment, W will often be observable to both the firms and the researcher), we discuss some methods to construct an artificial W in section 2.

From the discussion in the previous section 1.2.1, we remember *costs* (sensu stricto) are a natural explanation for migration deterrence in a variety of migration models. These costs should be included in the F variable of our model for the real options approach to work: without some sunk cost there is no value of waiting. Apart from an approximation of 'real costs', more intangible costs could also be included in F, making it an index of all costs attached to migration; of course we will often be able only to add proxies of the variables suggested by alternative migration theories, eg. proxies of the costs of information or social dependence to the region of origin and destination. In the empirical part we will take a close look at which variables in the GSOEP seem adequate to construct such proxies. With X a vector of cost proxies, we now model F as a linear index of these costs: $F = X\beta$ with β some coefficients.

We developed the net present value and real options approach without setting some termination date T; modeling as if the income stream from migration would continue forever. This contradicts the *human capital* approach of migration we discussed before and it explains why our model would fail to produce some well established results of these models; worse, it would return biased estimates of the effect of W on the migration propensity, since some of these variables will be correlated both with migration propensity and (variables determining) the expected income differential. To keep things as simple as possible, we will include proxies for those variables which the human capital models predict to be connected with the migration propensity, in the model's cost variable index F, as in Burda et al. [7]. For example: The human capital prediction of diminishing returns of migration with rising age will very probably be reflected in the data we will use later on. If we now include age as a cost variable in F at least this effect of human capital considerations our subjects make will not disturb the results. The importance of personal (time)preferences from the human capital models is somewhat been taken care of through a random-effects panel data approach.

This is the approach we will take with other variables the different migration theories propose to explain migration. We can try to include those variables in our real options model of migration without altering it, by turning F into an index collecting all influences other than the income difference. However, a rather important issue hidden here is that probably some of these costs accrue over time and are at least partially stochastic in nature, which would demand inclusion in W rather than F. Unfortunately, in general, this would require us to numerically solve a partial differential equation and the additional conditions needed to solve for the boundaries of optimal policy regions (H) would become much more difficult to find and solve, see Dixit and Pindyck [11] pp. 207-210 for details. Therefore, we will assume all costs are incurred only once. $F = X\beta$ should therefore contain some proxies for:

- real costs
- location specific human capital
- preferences (time,risk aversion,...)
- remaining time on the labour market

- cost of information
- social dependence
- liquidity constraints
- ...

Burda et al. [7] handled this by using detailed personal information available from the GSOEP as proxies for costs and benefits attached to migration on a personal level. We will adopt this approach, investigate some more variables and also try to make use of the panel property of the data to improve estimation. More on the variables that were chosen from the GSOEP in chapter 4. For the ethnic Germans dataset, the variables included in F, are the same as in Locher [29].

2 The Gains from Migration and the Problem of Self Selection

As mentioned in the previous section, the income gain from migration is an important determinant of migration in a wide variety of models and this is equally so for the real options approach to migration. In a real options framework, migrants behave in function of their anticipatition of the gain of migration, which is a nonlinear function of the anticipated income gain. Unfortunately, in practice, this quantity will not be observable[7] and in this section we will try to address some theoretical issues on how it could be estimated from real world data.
There are a number of problems connected to determining the anticipated income gain from migration for an individual. A typical approach would be to construct some model to estimate what income a possible migrant would earn in the region of destination and add some (sometimes implicit) assumptions on how the migrant builds anticipations of the income gain and other gains and costs to migration, when she decides on whether to migrate or not.

Assuming the migrant has perfect knowledge on how much she will earn in the region of destination, or at least makes an on average correct estimation of it, would greatly simplify matters, but unfortunately, in our case, this could be unrealistic. This is the question of the rationality of migration behaviour and we will address later.
A naive approach would then be to look at the income in the destination region, of those who migrated in the past and use this as a predictor for comparable subjects in the region of origin. Under the assumption that possible migrants make on average correct anticipations, one could tend to believe the measured income of similar migrants could be a good approximation. But no matter how detailed information we have on the similarity of the subjects, '*(...) the question is how similar these other individuals are, since for an identical set of observed characteristics they chose not to move, whereas the person in question decided to move*', as Davanzo [9] notes. So even within a 'rational an-

[7] It is even doubtful that data on answers to direct questions like 'what is your wage anticipation would you migrate to the west' would constitute unbiased observations on this variable, would such data be available. In practice this will rather be the exception. Nevertheless it would probably be a better approximation of the anticipations as those obtained through methods we propose later on.

ticipations' framework, we would still struggle with the eternal problem of social sciences, compared to for example physics when trying to address the effect of some policy, phenomenon,...: the fact that it is impossible to compare the same subject, or at least a subject similar in a sufficient degree, in both a 'treated' and a 'non-treated' state. The average causal effect would then be the expected value of the outcome variable when all individuals in a given population receive treatment minus the expectation of the outcome when no individual receives treatment (Holland-Rubin definition). Since individuals self-select their treatment status -in our case by migrating- this probably happens in a non-random fashion; therefore alternative methods had to be developed to approximate the ideal experiment.
We can write this more formally in the notation of Rubin [42], adapted to our migration application:

$$W^o = W^w M + W^e(1 - M) \tag{26}$$

where M is a binary variable equal to one if the subject migrated, W^O is the observed wage and W^w and W^e are the wages the subject would earn (earns) in the West and East respectively. Comparing average earnings for migrants and non-movers does not generally return the desired answer, as can be seen from the following decomposition, which can be found in any work on the subject, for example Angrist [3]:

$$\begin{aligned} E[W^o|M=1] - E[W^o|M=0] &= E[W^w|M=1] - E[W^e|M=0] \\ &= E[W^w - W^e|M=1] \\ &\quad -\{E[W^e|M=1] - E[W^e|M=0]\} \end{aligned} \tag{27}$$

The first term of the right hand side of equation (27) gives the average 'treatment effect' of migration. The second term represents the bias caused by the endogeneity of the migration decision. In general, this bias is different from zero because people will anticipate potential outcomes and this will affect selection (the determinants of selection do not have to be related to the outcome, but in our case they will probably be, at least partially).

The classic approach to this problem in a Roy [41] framework, was popularised by Heckman [23] and Lee [27]. Davanzo with Hosek [10], Nakosteen with Zimmer

[33] and Borjas [4] apply their approach to the migration problem. The migration decision is modeled in a separate probit model and a simple wage equation handles the income determination of the subjects: Denote with U^* the non-observable utility an individual anticipates to obtain from migrating and assume

$$\begin{aligned} U^* &= Z\gamma + \epsilon^* \\ M &= 1 \text{ if } U^* \geq 0 \\ M &= 0 \text{ if } U^* < 0 \end{aligned} \tag{28}$$

where γ contains some coefficients and Z is a vector containing some variables thought to influence the migration decision. ϵ^* contains some random errors. Together with a model of income determination for migrants and non-migrants

$$\begin{aligned} W^w &= X\beta_w + \epsilon_w \\ W^e &= X\beta_e + \epsilon_e \end{aligned} \tag{29}$$

we now have a complete model with (26), (28) and (29).

In (29) differences between β_w and β_e would account for some discrimination the migrants undergo, or unmeasured differences in the variables X thought to determine the income in the regions (or a combination). X and Z may contain common variables. ϵ_w and ϵ_e are error terms. The error-terms may be correlated; in fact this causes the whole self-selection problem in this framework. The error-terms are assumed to be jointly normally distributed:

$$\begin{pmatrix} \epsilon_w \\ \epsilon_e \\ \epsilon^* \end{pmatrix} \sim \mathbf{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_w^2 & \sigma_{we} & \sigma_{w*} \\ & \sigma_e^2 & \sigma_{e*} \\ & & 1 \end{pmatrix} \right)$$

which is equivalent to Nakosteen and Zimmer's [33] equation (12). To see how applying OLS directly on the observed wages for migrated individuals to obtain estimates for non-migrants would produce biased estimation of the income effect in this framework, consider

$$\begin{aligned} E[W^w|U^* \geq 0] &= E[X\beta_w|U^* \geq 0] + E[\epsilon_w|U^* \geq 0] \\ &= X\beta_w + E[\epsilon_w|Z\gamma \geq -\epsilon^*] \\ &= X\beta_w + \epsilon_{w*}\epsilon_w^2\lambda(Z\gamma) \end{aligned} \tag{30}$$

with $\lambda(z) = \phi(z)/\Phi(z)$ the rate of the normal distribution to its cumulative distribution, as can be shown by integrating the expected value (λ is known as the hazard rate or the inverse Mill's ratio). The second term in the right hand side of (30) would not be accounted for in a simple OLS model and cause biased estimation of β.
The Heckman two-step estimation of the parameters of this model, involves a probit regression on all individuals, to obtain some estimator of γ and with $\lambda(Z\gamma)$ and X known, we can estimate (30) by simple OLS on the migrants, to obtain estimators for β (and $\epsilon_{w*}\epsilon_w^2$).

Over time, it became clear the normality assumption of these models is over-restrictive in a number of cases, including the migration-income application as noted by Tunali [46]. Some less restrictive models have been developed, Tunali applies an approach by Lee [27] to migration. There exists, however, a class of models, that do not make any distributional assumption on the ϵ's apart from:

$$f(\epsilon_w\epsilon_e\epsilon^*|Z) = f(\epsilon_w\epsilon_e\epsilon^*|Z\gamma) \tag{31}$$

with f some smooth function. That is: the dependence of the error-terms on Z is only through the index $Z\gamma$. Now our estimated effect of migration on the income of non-migrants from (30) becomes $E[W^w|U^* \geq 0] = X\beta + \lambda(Z\gamma)$ where λ is some smooth function that needs to be estimated. Powell [38] and Andrews and Schafgans [2] constructed estimation methods for this. Luckily copy-pasting formulas is rather impossible in LaTeX, so I will not admit that the following is an exact copy of the compact summary on this from the self-selection tutorial of the XploRe documentation [47]:
For the estimation of γ -the probit part of the Heckman approach- a semi-parametric single index model is used. Suppose we have two observations with

$X_1 \neq X_2$ but $Z_1\gamma = Z_2\gamma$, then we could calculate β from:

$$E[Y_1|X_1] - E[Y_2|X_2] = X_1\beta - X_2\beta + \lambda(Z_1\gamma) - \lambda(Z_2\gamma) \quad (32)$$
$$= (X_1 - X_2)\beta \quad (33)$$

In practice, of course, we will hardly see any observations with equal $Z\gamma$ values. Powell's idea is to regress differences in observations on differences in the X variables, weighing strongly for those observations where the $Z\hat{\gamma}$ values are close. $\hat{\gamma}$ is estimated in the first (single index) step.
The flexibility of the semi-parametric approach comes at a price: estimation is less efficient than with their parametric counterparts, if the parametric specification is correct. Other non -and semi-parametric approaches to the self-selection problem exist, but we will ignore them here. Perhaps one could consider using instrumental variables estimation; it could be argued the very first waves contain relatively more individuals that moved primarily for non-wage reasons and this would make them not self-selected on the wage expectations. But since the East German panel was started in exactly the same period, only few of those individuals will be contained in the database. Furthermore, it is probably unrealistic to assume independence of these non-wage considerations of variables influencing income in the west. I should also mention a reason for not considering IV models is that I am not at all familiar with them. In that view, I chose not to further investigate this approach.

After discussing some different self-selection approaches it could be useful to recapitulate the initial aim. The real options approach to migration has the additional problem that, compared to the typical application to firm investment, the underlying stochastic variable W driving the value function, is not observable. Where a firm (and the investigator) will often be able to observe the current price of some underlying commodity, we ignore how much an individual is expected to gain (in earnings), or anticipates to gain upon migration. So far we have discussed two approaches which would allow to estimate the expected (as opposed to anticipated) income gain for migration. The expected gain to migration measures how much a randomly chosen non-migrant would gain from migration, conditional on some wage variables X.

In equation (28), the anticipations of the gain to migration were assumed to

drive the selection into migration. This gain probably contains wage and non-wage components (remember that X and Z were allowed to contain correlated or even the identical elements). If we could needly separate these components in Z, the probit-style estimated coefficients of the relevant variables thought to be correlated with the anticipated wage gain (a subset of γ from (28)) could be used to construct a proxy for the wage gain anticipations. Additionally, taking this approach would allow to test the rationality of migration behaviour. This is related to the method used by Tunali [46].
It is hard to conclude which measure would be best: assuming perfect foresight, the migrant knows what his expected wage gain will be and therefore expected wage gain and anticipated wage gain coincide. The advantage of this approach is that the income gain can be estimated in a separate model, using real income data. By dropping the 'rational migration' assumption, we allow migrants-in-spe to make wrong anticipations of the wage gain from equation (28). This could be caused by bad information availability or some other cause of seemingly irrational behaviour, as proposed by herd behaviour or relative deprivation models,... (cf. section 1.2.1). This approach has the advantage seemingly irrational migration behaviour can be described (eg. negative wage gains, in the absence of large enough non-wage advantages to migration). A drawback is that we have to estimate the sub-vector of γ in a probit model, inferring only from the binary information contained in M, ignoring available wage information. Additionally we know we shall have only few observations with $M = 1$ in our application to the German East-West migration. Also wage and non-wage considerations in $Z\gamma$ may be correlated or not nicely separable. Furthermore, the real options theory predicts the relationship between the anticipated wage gain W and U^* to be nonlinear [8] and this could be only partially compensated for by including transformations of the relevant variables in Z; all this makes our estimates will probably be quite unreliable.

However, if irrational anticipations of wage differences are allowed, one could argue there are alternatives to inferring from a probit-like model describing the migration behaviour: if we assume East Germans (somewhat naively) anticipate to earn West German wages upon migration, we could use the database on West

[8] For the exact form of the theoretically predicted dependence of U^* on W, see to figures 2 and 3 below.

German wages estimate those anticipations, independently from the question if they are rational or not. By using the West German wage database to estimate them, we would obtain more reliable estimates, since the number of observations is that much larger. Furthermore, this would help us to avoid the problem of the nonlinear dependency of the latent variable on the anticipations. In practice, one would regress West German wages on some variables thought to determine income and have a corresponding variable in the East German panel. The resulting estimators could then be used to construct predictors for each individual in the East German Panel, or define some 'counter-factual wage-equation' for the whole panel.

Yet another alternative 'irrational anticipations' specification, could be to assume East-Germans ignore the fact the migrant are self selected and anticipate to earn the wage of comparable individuals that migrated in the past.

After the discussion of some possible approaches one could tend to believe it is simply impossible to determine what East Germans believe about their wage prospects in West Germany, and even when we do not know which approach will deliver the correct answer. But by comparing the results we may exclude some hypothesis. If, for example the self-selection models predictions are sufficiently different from the simple West-wage anticipation models, it is reasonable to reject the latter, since, apparently, they do not describe the migration considerations people make when migrating, assuming the models are correctly specified.

Due to space-time constraints, we will look only at the Heckman self-selection-corrected estimates for the 'rational migration' assumption. For the 'irrational migration' models, we consider only the case where East Germans are assumed to anticipate to earn the same wages as similar migration-predecessors (thus ignoring the selection-bias). The semi-parametric approach proved infeasible due to the limited number of migrants and the fact a large number of explanatory variables is binary (dummy variables).

3 Developing Tests for the Theory

In this chapter we investigate what falsifiable predictions the real options approach to migration makes, before confronting it with real world data in section 4. For reference, the most important equations from the first section are repeated here: equation (25) for the value function

$$V(W) = \begin{cases} \frac{1}{r}(W + \mu/r) - F & \text{for } W \geq H \\ \frac{1}{r\beta} e^{\beta(W - rF + \mu/r) - 1} & \text{for } W < H \end{cases}$$

the definition of β from (20)

$$\beta = \frac{1}{\sigma^2}(-\mu + (\mu^2 + 2\sigma^2 r)^{\frac{1}{2}}) \tag{34}$$

the old (NPV-based) trigger level M

$$M = rF - \frac{\mu}{r} \tag{35}$$

and from (24), the new (ROTM) trigger level H:

$$H = \frac{1}{\beta} - \mu/r + rF. \tag{36}$$

Figure 1 illustrated the differences between the classic and the real options approach to investment. From section 1.2 we know it makes sense to use real options theory to understand migration and also what elements of different migration models could be included in the real option model: one-time costs and gains through F and the anticipated wage gain of the migrants through W. Unfortunately, as we saw in section 2, the latter variable is not directly observable and indirect estimation will be difficult. In section 4 we will try to estimate these anticipations from real data. For now we will assume this quantity to be known and focus on how empirical tests of the real options approach to migration could be constructed.

3.1 Testing through shifts in the trigger level

An interesting prediction of the real options theory for migration is the large shift in migration trigger level of W. This was illustrated in figure 1. To develop some empirical test for the ROTM, one could investigate with real-world data, whether this trigger level behaves as predicted by the ROTM . Of course, we should be more interested in predictions that are not made by other models. For example: it is clear from (36) that $\frac{\partial H}{\partial F} = r > 0$, but since this is a property of a large class of migration models, it will hardly be of any use as a test for the ROTM.

More interesting in this respect, is the dependence of H on σ^2. To calculate it, one could take the derivative of H, but it is easier to note that H depends on σ^2 only through β and evaluate the sign of $\frac{\partial \beta}{\partial \sigma^2}$ indirectly, by totally differentiating the quadratic function we calculated β from (in (20)), and evaluate it at β (as in Dixit and Pindyck [11], pp 143-144). With $Q = \frac{\sigma^2}{2}\beta^2 + \mu\beta - r = 0$ we get:

$$\frac{\partial Q}{\partial \beta}\frac{\partial Q}{\partial \sigma^2} + \frac{\partial Q}{\partial \sigma^2} = 0 \tag{37}$$

Since $\frac{\partial Q}{\partial \sigma^2} = \frac{\beta^2}{2}$ is always positive, the left term of (37) must be negative for the equality to hold. With

$$\begin{aligned}
0 &< \frac{\partial Q}{\partial \beta} \\
0 &< \sigma^2\beta + \mu \\
0 &< \sigma^2/\sigma^2(-\mu + \frac{\partial \beta}{\partial \sigma 2}(\mu^2 + 2\sigma^2 r)^{\frac{1}{2}}) + \mu \\
0 &< \mu^2 + 2\sigma^2 r
\end{aligned} \tag{38}$$

we know $0 < \frac{\partial Q}{\partial \beta}$ will always hold and therefore, finally, $\frac{\partial \beta}{\partial \sigma^2}$ must always be negative. With $\frac{\partial \beta}{\partial \sigma^2} < 0$ we know from (36) that $\frac{\partial H}{\partial \sigma^2} > 0$: increasing variance in the underlying process driving the income differential will *raise* the trigger level. Since this is rather counter-intuitive, the positive defence of H on σ^2 offers an opportunity to test the ROTM.

Unfortunately, σ^2 will be hard to define and estimate in our case. For financial option valuation, estimation of σ^2 is equally important. In this context, one would try to estimate the volatility from the variance in the stock price in the past or, alternatively, estimate σ^2 from prices of related options on the same stock. The latter method seems irrelevant for the ROTM, since there exists no market for migration options. Estimating the variance with some historic data on wage differences may be feasible, but since -as with the wage differences

themselves- for our application it is the subject's anticipation that is relevant and not the observed variance, results could be severely biased.
Even under the assumption that individuals anticipate the correctly estimated historic volatility, we should consider how large a shock in variance should be to result in a 'testable' shift of H. For some reasonable parameter specifications I tried, a ten percent change in σ^2 caused a shift of about four percent in H.
Moreover, it will probably be rather difficult to find some separate measure for σ^2 in micro-data. For example: for individuals in the GSOEP we know whether they are temporarily employed and how long they have been working in the same place. It seems reasonable to assume σ^2 will be higher for temporary employees that have acquired relatively little specialised human capital, since they are relatively more likely to be fired (and earn less), obtain a full-time position (and earn more), change jobs, not be member of a trade union,... At the same time, F is likely to be lower for them, since they will not lose a lot of locality-specific human capital upon migration. A significant negative coefficient on some temporary-employee-dummy in a regression of migration propensity would therefore offer great support for the ROTM: we could explain it by the effect on H through σ^2. However, this is not very likely to happen, given the effect through F and to interpret a relatively high but negative or insignificant coefficient as support for the ROTM seems risky.

Changes in the μ parameter, the trend of the Wiener process, influences both the NPV value (and therefore the marshallian trigger M) and the ROTM value function (and therefore the ROTM trigger level H). Therefore, we must look at the influence μ has on both triggers. For H, we have

$$\frac{\partial \beta}{\partial \mu} = \frac{1}{\sigma^2}(-1 + (\mu^2 + 2\sigma^2 r)^{-\frac{1}{2}}\mu) < 0 \tag{39}$$

and from this, we obtain

$$\frac{\partial H}{\partial \mu} = -\sigma^2(-\mu + (\mu^2 + 2\sigma^2 r)^{\frac{1}{2}})^{-2}(-1 + (\mu^2 + 2\sigma^2 r)^{-\frac{1}{2}}\mu) - \frac{1}{r}.$$

Noting the whole first term is positive, we see the sign of the derivative is still a rather complicated function. Dixit and Pindyck look at the influence of the difference of μ and r, in a model with a geometric brownian motion. We do not take this approach since measurement of such a variable would be quite hard in the case of migration. We could make assumptions on the other parameters, to obtain some sign for $\frac{\partial H}{\partial \mu}$, but by taking the limit of H, we can go without this and obtain some asymptotic results:

$$\lim_{\mu \to -\infty} H = \sigma^2(+\infty + (+\infty + 2\sigma^2)^{\frac{1}{2}})^{-1} + \frac{\infty}{r} + rF = +\infty$$

This is evident: if convergence is very strong, the current level of W becomes irrelevant and people will not migrate. Recall that a negative μ means convergence and, in time, higher wages in the region of origin.
Turning to the effect of μ on M:

$$\frac{\partial M}{\partial \mu} = -\frac{1}{r} \tag{40}$$

we see the effects of μ on M and H are of equal sign and therefore this variable will be less useful to test the ROTM. [9]

As the variable r will be practically unmeasurable in our case, given its definition as personal time preference, and the limit of H and the derivative of M turn out to be of equal sign here too, we conclude that the only really testable prediction the ROTM makes about shifts in the migration trigger level, are through the dependence of H on changes in the subjective assessment of the variance in the underlying process driving the anticipated wage differences: σ^2. Unfortunately, this nice theoretical property will probably be hard to be used in a real test, due to measurability problems and these problems are pronounced in the application of the ROT to migration.

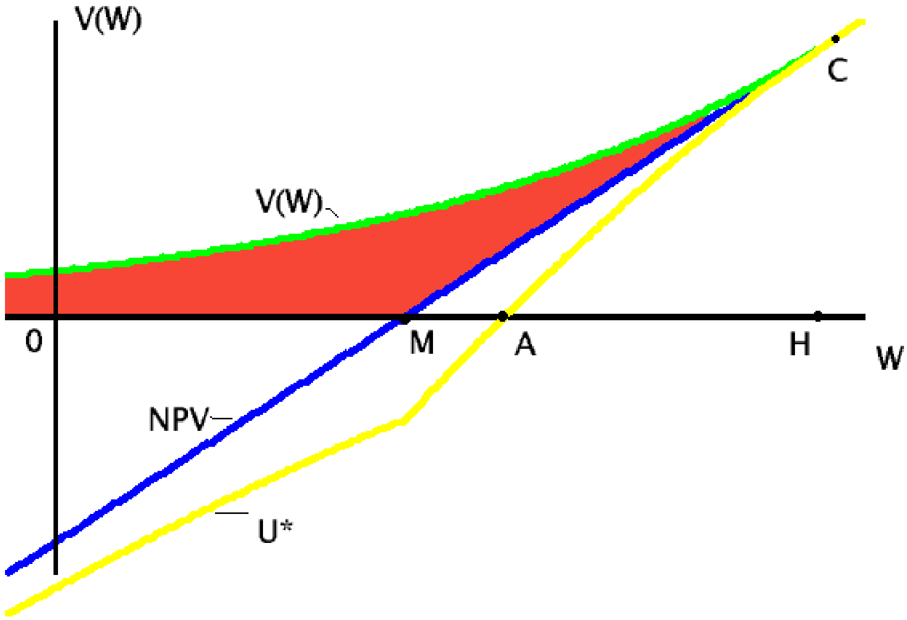
3.2 Testing the nonlinear dependence on the income differential

Looking at figure 1 or equation (3) makes clear the value function is highly nonlinear in W. To understand how this property could be used to test the ROTM, I mention here that in both datasets we will use to test the ROTM, people are asked (or at least give information indirectly) about their migration intentions. The approach, taken by Burda et al. [7], Locher [29] and also by us, is to *in-*

[9] Of course, since we take the limit and not the derivative, this is not entirely correct, but for $\frac{\partial H}{\partial \mu}$ to be less interesting as a test for the ROTM, we only need to know there are *some levels* of μ where the derivatives of both trigger levels are of equal sign.

terpret these migration intentions as being driven by the unobserved utility level an individual would obtain from migrating: people with a utility level between some critical levels are thought to give a certain answer. Of course, migration behaviour itself would also indicate an individual thinks to deprive some utility from migrating. The simplest approach (as in logit and probit models), that is also investigated by Burda et al. [7] and Locher [29], is to assume this unobserved utility level to be a linear index of some variables.

If data is (made) binary (comparing just 'low' or 'high' intention to migrate, or 'migrant' vs. 'non-migrant' status), the parameters of the index driving U^* could be estimated in a simple probit or logit model, as in the probit part of the self selection models we discussed (see equation (28)). In the classic Marshallian model, it is clear that it is the NPV that should be used as the unobserved utility level: in this model people with positive utility (NPV) of migration are predicted to migrate. Alternatively, using data on migration intentions, one would say people with a utility level larger than some (negative) level of utility have high migration intentions. Since the NPV is a linear function of W and assuming we have some proxies for the variables contained in it, such a model would allow to test the 'human-capital'-NPV theory. Burda et al. and Locher note some results for such a simple logit model and indeed, the coefficients on the different variables, thought to reflect elements from F and W as described in 1.2.2, are significant and have the predicted signs.
In order to use this 'latent variable approach' with the ROTM, we have to answer exactly the same question: what utility would an individual obtain from migrating. The whole argument of the ROTM was that the NPV-calculation was wrong because it ignores the opportunity cost of making a commitment now and giving up the option to wait. Burda et al. correct the NPV latent variable to include this opportunity cost. Remarkable is, however, that they do not use the value function -that is the value of the option of migrating later- as the opportunity cost, but instead the difference between the value of the project with all procrastination possibilities calculated in and the value of the project in the classic now-or-never case is used. This is what Burda calls the 'Option Value of Waiting'. Figure 2 illustrates the specification of Burda et al.: the yellow curve U^* is the difference of the NPV-curve and the opportunity cost used by Burda et al. -the OVW- which is shaded red. This specification of the latent variable

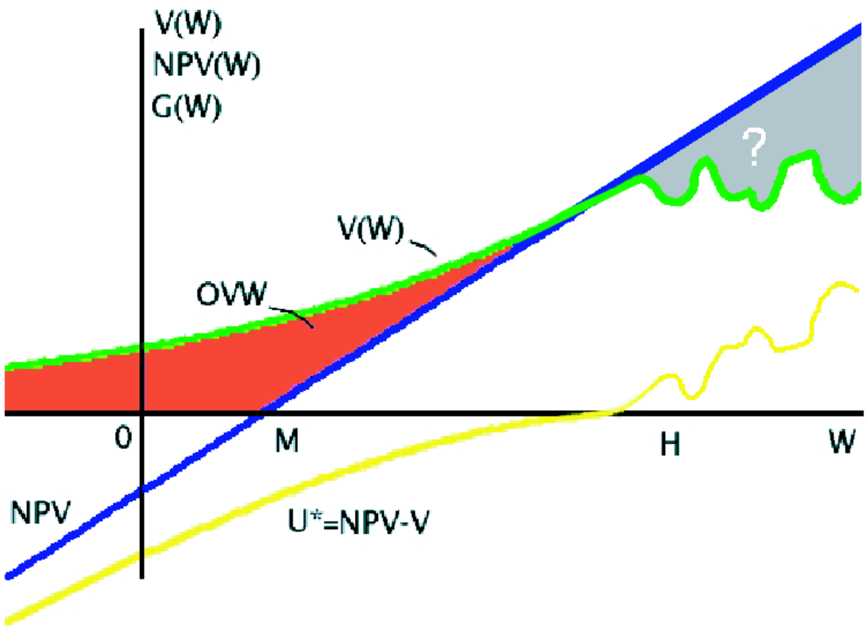


source: own calculation

Figure 2: The gain from migration

is given by equation (11) in Burda et al. As can be seen in the figure, taking the OVW as the opportunity cost, predicts positive utility from migration for individuals to the right of point A. But as we saw in section 1.1 the ROTM predicts migration only to the right of the trigger level H! In principle this could be corrected for by redefining the utility level that has to be reached from migration to the level of point C, but apart from 'repairing' the trigger level, it would be hard to justify such a change.

As an alternative, I propose to take as the option opportunity cost, not the OVW, but the whole value of the option, the value function V. Figure 3 illustrates this alternative specification. The form of U^* and V in the figure was drawn somewhat strange to the right of H, to illustrate an important issue we have so far ignored. Normally, U^* would be equal to zero for $W \geq H$: in figure 2, the value function was equal to the NPV, beyond H. This was a logical consequence of the assumptions of our model. For levels of W where immediate migration would be optimal, the value of the option to migrate was exactly equal to the NPV of migration. This means that for $W \geq H$, the utility from migrating would be exactly zero, as the individual loses exactly as much as she gains. Note that this is rather counter-intuitive: an individual contemplating to migrate to a region where she would earn infinitely more then she does now ($W = \infty$), would



source: own calculation

Figure 3: The gain from migration; to the right of H the exact form of U^* is unclear.

still obtain zero utility from migrating.

This is, however, a consequence of the unrealistic continuous time framework we used. If the individual decides to postpone migration for another infinitesimal time-span, the present value of the utility that is lost because the income stream would start to flow a bit later is very small. Therefore, in continuous time, the individual reaps only an infinitesimal utility gain when migrating, for levels of $W \geq H$.

In discrete time, however, or when the option to migrate can only be exercised in certain time periods or otherwise deviates from the theoretical option, there exists a clear net gain of migration compared to waiting for $W \geq H$. An individual with a large value of W would now loose a positive amount of utility when postponing migration. The exact form of the difference between the NPV and V (the red and grey area in figure 3) and therefore the form of U^*, will be highly dependent on how exactly the migration option differs from the theoretical continuous-time and always exercisable option. The NPV-case could be understood as the most extreme deviation from this ideal option: here, V(W) equals zero, as there is absolutely no value of waiting. Therefore, the opportunity cost is always zero and U^* coincides with the NPV in this case. I choose not to draw 3 for some specific model (which would also demand recalculating

the form of V and U^* to the left of H), but rather leave the the form of V and U^* for $W \geq H$ undefined.
Note that we should only observe individuals with $W < H$ when estimating with migration intentions, so the alternative specification of the form of the estimated m in the GPLM would be the part of U^* to the left of H in figure 3. As our model specification will never be perfect, we will also observe individuals with W-levels that are predict to gain from immediate migration. As the migration option will always deviate to some degree from the theoretical option, we can assume U^* to be rising to the right of H. Therefore, the probability of observing a non-migrant (or low migration propensity individual) should approach zero, for $W \to \infty$. To calculate V for a discrete time - partially exercisable option should be possible using numerical methods.
Correcting the trigger level of the latent variable is an *effect* of this alternative specification for the opportunity cost. The argumentation, however, is simply the consequent application of the ROT: The opportunity cost of migrating now is not to be able to wait any longer and V measures exactly this.

From this it is clear the specific nonlinearity of U^* (be it in the specification of Burda et al. from figure 2 or the alternative in figure 3) is something we could use to test the ROTM and this is the approach Burda et al. take. In stead of accommodating the simple logit model to allow for some nonlinearity (e.g. by including higher order transformations of the W variable) they use a GPLM model: $U^* = X\beta + m(W)$. Where $m(W)$ is a smooth function to be estimated. If the estimated m would show the specific form illustrated in figure 3, this would offer strong support for the validity of the ROTM.
Since the simple logit and the GPLM can not handle ordered data one would have to transform the data and this would cause valuable information to be lost. Since data on migration behaviour is already binary, this will not be an issue, but the variable on migration intentions from the datasets is ordered. A way to conserve the information using this data in a test for the nonlinearity of U^* could be to construct different datasets with binary data, containing only observations from 'neighbouring' classes of the dependent variable. If we would measure a lower slope for those models comparing 'more enthusiastic pairs of migration intentions', this would prove the slope of U^* is decreasing over all the answers and therefore could also offer support for the ROTM, in the specification illustrated

in figure 3. However we will not investigate this approach. Another approach could be to simply ignore the fact the dependent variable is not continuous and use, for example, a PLM to investigate the nonlinearity in U^*. We will do this in section 4.

Locher [29] uses ordered and multinomial logit models to test the ROTM. Unfortunately, although her approach nicely describes the data on migration intentions and provides support for the predictions of the model about the effects of costs, age,..., it fails to be a real test for the ROTM, since all these effects are equally predicted by simpler models of migration. The fact that the signs of the coefficients on the cost variables are different for the different classes of responses in the multinomial model offers no prove beyond the fact that on average, in the dataset she uses, people with relatively higher (lower) migration costs, are less (more) probable to be highly enthusiastic about migrating, which is also predicted by e.g. the NPV model.

3.3 The effect of changing the model specification

As I wrote above, our specification of the underlying stochastic process for W is not critical for the ROTM to work, as long as the process is persistent. But this is true only for shift in the trigger value relative to the marshallian trigger; under a wide variety of specifications an increase in σ^2 would still cause a increase of H, but the specific form of the nonlinearity in U^* changes with the underlying process. In general, these effects are not dramatic: for a geometric brownian motion we would get almost the same picture. For mean reverting, jump processes and mixed processes the reader is invited to take a look at Dixit and Pindyck [11] pp. 161-173. So perhaps it would be best to first apply the GPLM approach of Burda et al. and simply look at the form of the nonlinearity of U^* before applying some specific model or tests.

Another test of the ROTM could be to look if migration (intentions) change, when people change their estimation of the expiration date of the option. We will not investigate this issue formally, as we constructed the whole model assuming a never expiring option on migration. For the application to German internal migration, this is probably realistic. For other applications, this could be less so. Evidently, people will value an option value to wait less if the expiration date is (thought to be) closer. Lilo Locher [29] found proof of these effects in international migration flows from East Europe to Germany.

4 Testing the Real Options Approach to Migration

We will use two datasets to evaluate the two different tests of the real options approach to migration from section 3: the German Socioeconomic Panel (GSOEP) from the 'Deutsches Institut für Wirtschaftsforschung' (DIW) for German internal East-West migration and a dataset from the 'Osteuropa-Institut' in Münich, containing data on migration intentions of ethnic Germans in Russia. In the next two subsections, I will discuss the two datasets separately. For each dataset, I first give some background, look at some interesting statistics and finally apply some statistical models to test the falsifiable predictions of the ROTM (cf. section 3): the shift in the trigger level caused by changes in the volatility σ^2 and the nonlinearity of the dependence of the migration propensity on the wage difference W. For the GSOEP data we will consider the possibility to construct estimated W^W as described in section 2. For the ethnic Germans data this was not investigated.

4.1 The German Socio-economic Panel dataset

4.1.1 Background

This section bases on Haisken-DeNew and Frick [19].

The GSOEP was started in the FRG in 1984 and several subpanels were added in the course of time. For us, subsample C, containing data on German residents of the former GDR is the most interesting. It was started right after the wall fell (even before actual German unification) and contains information on several variables that could be linked to the ROTM. As we saw in section 1.1, for the ROTM to be relevant, three assumptions need to hold: there has to be uncertainty in the evolution of the wage differences; there have to be substantial sunk costs attached to migration and the decision to migrate has to be postponable. All three assumptions are fulfilled in the case of German internal migration. There are large wage differences between the two regions, and the evolution of this wage gap is all but certain, especially on the individual level. There is no 'expiration date' for the migration option in the foreseeable future. However, a significant part of the costs is probably recoverable in the case of German

East-West migration, therefore the ROTM could be less relevant [10]. Since the ROTM does not need all costs to be unrecoverable, the ROTM will probably remain valid, at least to some extent.

The data in the GSOEP is gathered on a yearly basis, through questionnaires which are filled out in a face-to-face interview with the subject. These subjects are all household members older than 16 years, which live in the surveyed household at the time of interview.
The households were randomly chosen in june 1990, by following procedure: first, using the central residents' database of the former GDR, representative subsets of households were drawn with respect to its stratification according to community size. From these samples, some households were chosen, using systematic with random start selection. In the next step, a random household member older than 16 was chosen from these households; the address of this individual now becomes the start adres of the next procedure, where the interviewer lists the households on a clearly defined literally 'random route' from the starting adress. From these adresses every third was chosen.
This way, 2179 households were selected. This represents a sampling probability of about 0.0004. Foreigners, approximately 1.7% of the population, were excluded from the sample.

The DIW puts a lot of effort in avoiding dropouts in the panel. The subjects are given some nonpecuniary rewards for participating (eg. they take part in a lottery and receive some small gift every year). Also the interviewer is changed as little as possible over time, to create some personal relationship. Apart from refusals to further participate, there will be 'natural' dropouts, when the individual dies, moves abroad,... Individuals may also enter the sample by moving into the surveyed household. Children in the household enter the sample when they reach the minimum survey age of 16. There are also split-offs, when a member leaves the household; if this happens, a new household identifier is created in the sample. The aggregate effect, however, is a diminishing sample size over time (cf. table 5), mainly due to people declining to further participate. As dropouts are likely to be nonrandom, not accounting for them could bias findings. However, the bias caused by the manipulations we have to make to adapt the dataset to

[10] Lilo Locher pointed me to this fact.

our framework would make the relatively small corrections for the dropouts a bit ridiculous, so we will ignore this issue.

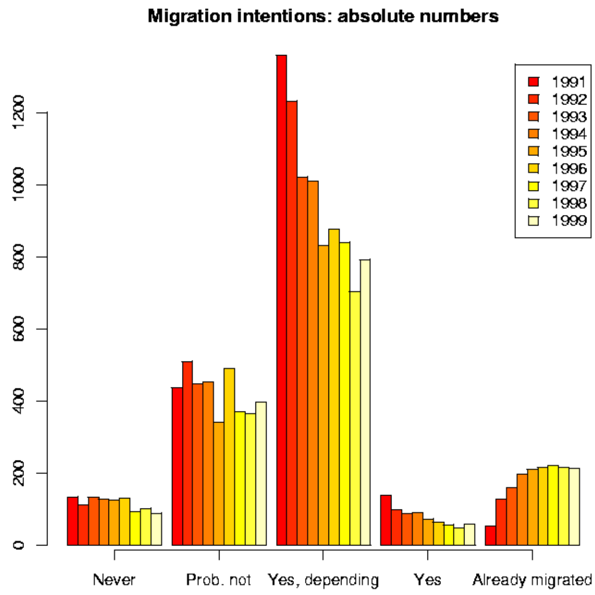
4.1.2 Data description

Because the East German panel of the GSOEP was started so soon after the wall fell, the data in the first wave(s) probably contains influences on migration intentions and behaviour other than those we are interested in (eg. politically motivated migration). Inclusion of this wave could bias our results. Therefore, the first wave included in the models is the second wave of the East German panel, collected in the spring of 1991. This second wave is the one used by Burda et al. [7]. Unfortunately, the last wave available at the time of writing (collected in 2000), lacks some of the key variables relevant to the ROTM analysis, most notably the variable we will use as the dependent variable. Therefore, only the nine 1991-1999 waves are investigated here.
To test the ROTM through the nonlinearity in the value function, we need some variable that relates to this function. In the dataset, one variable contains the answers to the question 'Could you imagine to move to West Germany, that is to the old FRG or to West-Berlin?'. Possible answers were:

- Yes, very much
- Yes, depending on situation
- Probably not
- Never

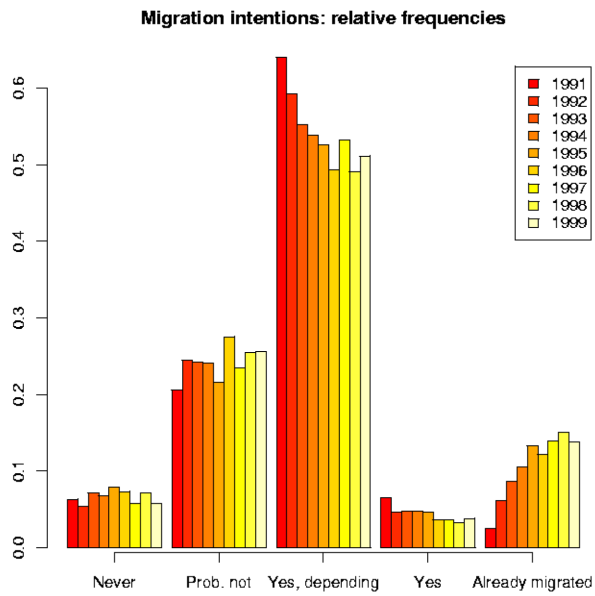
A key assumption in Burda et al., which we will also (have to) make in all following models, is to assume this answer to be driven by the individual's subjective anticipation of the gain to migration. This is the yellow curve U^* in figure 2 or the alternative from figure 3.

As a first attempt in understanding the structure of data, it is always interesting to look at some graphical representations. For the dependent variable, the migration intentions, figures 4 and 5 show barcharts with the absolute and relative frequencies of the answers over the different waves. Differences between the two charts (apart from the scale) are caused by the dropouts: the sample size diminishes over time. Again, if these dropouts are not random this could cause bias in



source: own calculation from GSOEP data

Figure 4: Migration intentions: absolute frequencies

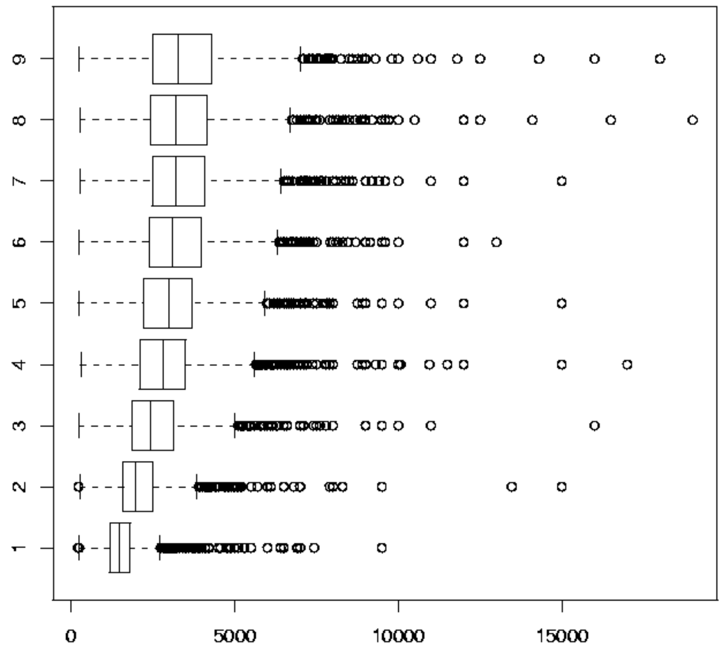


source: own calculation from GSOEP data

Figure 5: Migration intentions: relative frequencies

our estimation later on. For some summary statistics of the dependent variable, the reader should consult table 5, where summary statistics for all variables from the GSOEP we use are given for the different waves.

Turning to the independent variables, we start with the variable which connec-



source: own calculation from GSOEP data

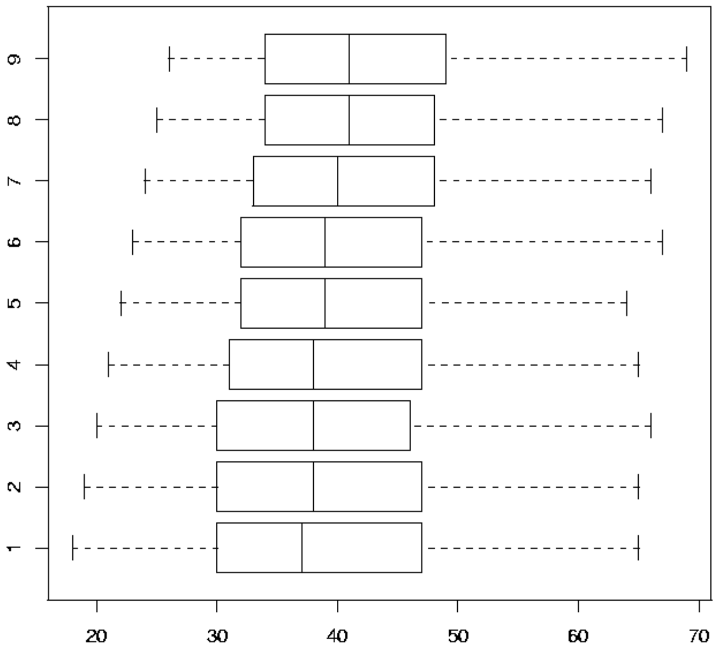
Figure 6: Boxplots of the household income

tion with the dependent variable will hopefully allow to test the ROTM: the expected income gain from migration, W. From now on, we take the approach from Burda et al. [7] and model $W = W^W - W^E$. Of course W^W is not observed for people that did not migrate, so assumptions will have to be made (cf. section 2).

The GSOEP contains a wide variety of income information one could use to model W^E. I considered using the reported gross monthly household income, as in Burda et al. [7]. Household income seems more relevant, since migration is often a family decision. However, the household income will probably be highly correlated to other variables influencing migration behaviour (below we take a closer look at this issue). Also, it is questionable if there exist large differences in income between East and West Germany for voluntarily unemployed family members, as unemployment benefits -and social benefits- are equal in both regions. Therefore, I chose as independent variable of interest the reported gross *labour* income.

Figure 6 shows boxplots [11] for the income for waves 1991 to 1999. For the sake

[11]The boxplots include only observations with $20 \leq AINC \leq 20000$. It may seem a bit strange to censor observations for boxplots, but by consoring from above, only ten observations



source: own calculation from GSOEP data

Figure 7: Boxplots of the respondents age

of completeness, density estimates are included in figure 17 in the appendix, but perhaps the boxplots illustrate more clearly some important characteristics of the evolution of the income distribution: there has been a large increase in the distribution of income over time, but this process has slowed down significantly for the last waves. From the boxplots, we also clearly see the income distribution has become more unequal, which could mean the uncertainty an individual faces with respect to his home income increased. This would cause the ROTM-effect to become more pronounced over time. Another variable suggested as a major determinant of migration (propensity) by human capital models of migration, was age. Here too, we see a (relatively small) shift in the distribution over time, as is illustrated in figure 7. From this we learn a simple cross-sectional approach using data from all waves could be biased. i.e. If we assume the underlying functional form of the dependence of the migration propensity remains constant over time, but the average income rises and, at the same time, (through some

(in total, for all nine waves) with extremely high income are removed. When included, these observations made the plots hardly readable. One could also question the validity of some of these reported income levels (typo's, unwillingness to report real income . . . ?). Also, only by censoring from below, a true image of the sample used in the models later on is obtained, as we will only include individuals with a positive labour income.

influences we do not model) the average migration propensity diminishes, this would cause more high income-low propensity observations to enter later waves, thereby biasing estimation of the dependence. In the next subsection a simple dummy variable approach is used to somewhat filter out such time-effects.

Table 5 in the appendix gives some summary statistics on the other independent variables that will be used in the different models (only observations with a reported gross personal income between 200 and 7000 were included).
Compared to Burda et al. [7], I considered some more explanatory variables. The GSOEP contains an enormous amount of variables, of which quite a few could be used for explaining an individual's migration decision/intentions. A first selection was made from a broad set of variables, suggested by the different migration models we discused in section 1.2.1. Using statistical selection criteria, the number of variables was further reduced. It must be noted, however, that for the majority of variables that were not included in the model, the sign on the coefficients was as predicted by theory. The variable that proved the most significant throughout the different models was ATT, measuring to what degree the individual feels a part of the local community.

In order to construct a functioning model from the raw data from the GSOEP, certain manipulations were necessary: observations that had no entries for some wave were removed. Some individuals, however, have an entry indicating some other specific reason for the missing value: the question was not relevant for this person, was left blank,... Since there is a large amount of observations that have such an entry for some variable, deleting them would drastically reduce our sample size (and probably cause bias). To avoid this, I calculated the individual mean over all waves where we have information on this variable for each individual, and assumed this value for the wave where the answer fails. For individuals without a valid observation for some variable in any wave, the mean over all individuals was used. It must be noted that this approach will probably cause little bias, since for most entries only on one or two variables fail for some waves. This prevents the large amount of valuable information contained in all other variables to be lost for that entry. This was only done for variables where it makes sense, for example for ENV, the environmental satisfaction. For other variables some other value was assumed for invalid entries; eg. people with an invalid entry for UNI are assumed not to have a university degree. People that once had a university degree are assumed to have one for all following waves. Observations without valid entries for the most important variables of our analysis, the reported income and the migration intensity or had too many missing entries for some wave, were not included for that wave.

4.1.3 Estimation using the reported income W^E

We know the reported migration intention is an interesting variable because it relates to the anticipated gain of migration U^*, and therefore could be used to test the specific nonlinear dependence of the migration propensity on W. We will denote the dependent variable simply with Y and the 'true' dependence of Y on W by $f(W)$. We assumed $W = W^W - W^E$ and will now turn to estimating the dependence of Y on W^E, say $m(W^E)$, ignoring the fact that $m(W^E) \neq f(W)$ for now.

Independently from the assumptions we make about the dependence of W on W^E, estimation of $m(W^E)$ will not be straightforward, since in our dataset the dependent variable is non-continuous. There exists a wide variety of interesting models that can handle such data. But the fact we want to test the ROTM through the nonlinearity of the dependence of the latent variable driving Y on the income, makes that ordered or multinomial logit approaches would not contribute to this aim. These models would return some estimated influence of some variable on the probability for the dependent variable to take on *each possible* value compared with some base category (odds ratio's), which is not what we need.

There are essentially two solutions to this problem: since the variable takes on four values, we could simply ignore the ordinal character and use a standard linear model. Of course the assumption of normal distributed error terms will be severely violated, but this approach could work nevertheless. By inclusion of higher order terms, some nonlinear influences could be accounted for, but, as became clear from the theory (cf. Burda et al. [7] and section 3), the relation we are looking for is highly nonlinear and would probably be rather hard to include parametrically in a robust way. Therefore, we will investigate some non -and semiparametric models, which can handle this. For the ordinal response variable models, I look at a nonparametric and a partially linear model. Alternatively, one could use a binary transformation of the data, discriminating only between high and low migration intentions. This way we could estimate the coefficient(s) on the latent variable(-index) driving the migration intentions. We will investigate the estimation of GAM and GPLM models which allow estimation of nonlinear dependencies for binary response variables.

Binary response

Parametric approach Using a binary transformation of the dependent variable, distinguishing only between high (answers one and two) and low (answers three and four) migration propensity, I fitted a number of parametric models. A first attempt to build a model for the whole sample could be to use the traditional approach to fixed-effects panel data, as described in, for example, Judge et al. [25] pp 468-491. These models are used for balanced panels and make use of the block-linear structure of the covariance matrix of such data. Three different approaches are possible, depending on the characterstics of the data.
For a small number of subjects, observed for a large number of time periods, the so-called within groups estimator can be used. This estimator can be understood as a simple OLS regression of the differences of the variables with their within group (subject) average. Since we have a large number of subjects and only nine observations on each of them, this approach probably will not work. But the estimator could be interesting to test a remark made in section 4.1.2 and also considered by Burda et al. [7]: the fact that both the ROTM and the NPV model of migration predict a *negative* dependence of the migration propensity on W^E, under the assumption that W^W is constant. Although this assumption is clearly violated looking at the whole panel or at the different individuals in a single wave, it is probably much less so for a single individual over the different waves. Indeed, it seems quite probable most characteristics determining the possible W^W's (or alternatively, the individual anticipations of this) remain relatively stable over time for a single individual. Some changes likely to influence these anticipations (eg. completing university, becoming children) could be controlled for. Estimation of this panel model in XploRe [47] revealed the sign on W^W is significant and negative. It must be noted that, as our data does not have an equal number of observations on each subject, one could hardly use this model as a complete model for the data data, apart from the fact we have relatively little observations on the each subject.
The other extreme approach to panel data is to regress on the averages of the

variables for each subject, taken over the different waves. This estimator ignores all within group information and uses only between groups variance, hence it is called the between groups estimator. This estimator should show a positive dependence on W^W, if the assumptions $W = W^W - W^E$ and $W^W = aW^E, a > 1$, or some other specification where W^W is positively correlated with W^E holds and indeed this was found to be the case.
For balanced panels, it could be usefull to construct a generalised least squares estimator, as described in Judge et al. [25], which is an efficient combination of the within and between groups estimators. However, since it is unclear how the results should be interpreted, given the above arguments and the serious problems of applying it to unbalanced data, I did not attempt it.

The *gee* package [40] of the statistical software R [16] allows to handle the binary response data from the different waves in a single model, with unbalanced data. GEE models are likelihood based, and should be unbiased in the binary unbalanced case if the missing observations are completely at random (cf. Liang [28]). As mentioned above, this assumption does not hold completely for the GSOEP and we will simply ignore this issue.
Since the income variable we are interested in has only relatively small explanatory power, ignoring the nonlinearity in this variable will probably not greatly bias the estimated coefficients on the other covariates. This panel model includes information from all waves and therefore could point us to which variables could be included in the other models we will use to actually test the ROTM. Table 2 shows the call and output of the GEE function in R. The variable names and their descriptions are listed in table 5. The 'd.' variables are dummies for the different waves (eight dummies and the intercept for a total of nine waves in the sample). With the 'corstr=unstructured' option, the GEE function estimates the correlation structure for observations within a group (individual). Recall that the dependent variable BIN is a binary transformation, discerning only between high and low migration propensity for each individual. With 'id=nr', we estimate a simple random effects model: for each individual, the intercept is modeled as consisting of a fixed part (estimated as 'Int.') and a random component that is different for each individual. Note a random effects approach is well suited for this dataset, as we have relatively little observations on a large number of subjects. (cf. Judge et al. [25] pp. 489-491)

```
Call:
gee(formula = bin ~ SEX + PARTNER + OWNER + FF + ENV + UNI +
    BIL + QUALW + KID + SAME + WHL + WHH + ATT + AGE + AINC +
    SATI + SATD + SATJ + d.2 + d.3 + d.4 + d.5 + d.6 + d.7 +
    d.8 + d.9, id = nr, data = D3, family = binomial(link = probit),
    corstr = "unstructured", silent = FALSE)

Summary of Residuals:
       Min         1Q     Median         3Q        Max
-0.8525825 -0.3405287 -0.2020815  0.4905943  0.9265504


Coefficients:
             Estimate  Naive S.E.    Naive z  Robust S.E. Robust z
(Int.)   9.518846e-01 2.098133e-01   4.5368169 1.8729e-01   5.08237
SEX     -1.204040e-01 3.356402e-02  -3.5872934 3.9360e-02  -3.05902
PARTNER -1.129863e-01 4.065387e-02  -2.7792262 4.4969e-02  -2.51250
OWNER   -1.783019e-01 2.455063e-02  -7.2626182 2.3173e-02  -7.69433
FF       2.440763e-01 4.314095e-02   5.6576476 4.6165e-02   5.28699
ENV     -1.760478e-02 7.609371e-03  -2.3135654 7.7300e-03  -2.27745
UNI      1.892631e-01 3.891194e-02   4.8638832 4.6478e-02   4.07202
BIL     -2.527476e-02 1.784092e-01  -0.1416673 1.4189e-01  -0.17811
QUALW    2.376582e-02 7.380913e-03   3.2199022 7.9442e-03   2.99156
KID     -1.727875e-01 5.046870e-02  -3.4236572 5.6349e-02  -3.06635
SAME    -1.359878e-01 5.089747e-02  -2.6717981 6.6199e-02  -2.05420
WHL      1.461819e-01 3.031086e-02   4.8227573 3.2383e-02   4.51408
WHH      1.111231e-01 4.493881e-02   2.4727638 4.7326e-02   2.34802
ATT      5.767193e-01 2.921689e-02  19.7392417 3.3119e-02  17.41324
AGE     -2.797759e-02 1.661566e-03 -16.8380813 1.8533e-03 -15.09605
AINC     5.577744e-05 1.389838e-05   4.0132346 1.5560e-05   3.58444
SATI    -1.298507e-02 6.265913e-03  -2.0723350 6.1531e-03  -2.11031
SATD    -2.284901e-02 5.356340e-03  -4.2657876 5.3764e-03  -4.24982
SATJ    -1.474830e-02 5.628745e-03  -2.6201750 5.7784e-03  -2.55230
d.2     -7.674905e-02 2.975818e-02  -2.5790904 3.0147e-02  -2.54577
d.3     -1.898655e-01 3.480270e-02  -5.4554811 3.4724e-02  -5.46772
d.4     -1.592155e-01 3.812023e-02  -4.1766659 3.7356e-02  -4.26208
d.5     -3.149186e-01 4.119942e-02  -7.6437625 4.1603e-02  -7.56946
d.6     -2.049438e-01 4.362470e-02  -4.6978848 4.3790e-02  -4.68011
d.7     -1.950173e-01 4.620355e-02  -4.2208284 4.6062e-02  -4.23374
d.8     -2.558771e-01 4.851668e-02  -5.2740025 4.8447e-02  -5.28154
d.9     -9.844280e-02 4.930181e-02  -1.9967383 4.9744e-02  -1.97898

Working Correlation
           [,1]      [,2]       [,3]       [,4]       [,5]       [,6]
 [1,] 1.00000000 0.4669600 0.35096999 0.27112749 0.22097254 0.17813363
 [2,] 0.46696000 1.0000000 0.42416491 0.32062967 0.23532943 0.20591346
 [3,] 0.35096999 0.4241649 1.00000000 0.36873410 0.26953837 0.23775691
 [4,] 0.27112749 0.3206297 0.36873410 1.00000000 0.31945068 0.24017097
 [5,] 0.22097254 0.2353294 0.26953837 0.31945068 1.00000000 0.26238571
 [6,] 0.17813363 0.2059135 0.23775691 0.24017097 0.26238571 1.00000000
 [7,] 0.11406446 0.1194291 0.15098146 0.16962212 0.18829027 0.21290164
 [8,] 0.07252956 0.1021717 0.11389335 0.12399409 0.15331552 0.17380145
 [9,] 0.04453767 0.0559195 0.06570037 0.07342149 0.07696599 0.08909114
          [,7]       [,8]       [,9]
 [1,] 0.1140645 0.07252956 0.04453767
 [2,] 0.1194291 0.10217168 0.05591950
 [3,] 0.1509815 0.11389335 0.06570037
 [4,] 0.1696221 0.12399409 0.07342149
 [5,] 0.1882903 0.15331552 0.07696599
 [6,] 0.2129016 0.17380145 0.08909114
 [7,] 1.0000000 0.15780073 0.08664480
 [8,] 0.1578007 1.00000000 0.10735881
 [9,] 0.0866448 0.10735881 1.00000000
```

Table 2: Results of the GEE estimation

The GEE model seems to fit the data rather well and confirms the findings of the migration models from section 1.2.1: all variables have the expected signs. Some of the variables not included were a measure for the required training the individual needs for his job (proxy for human capital), a variable indicating part-time workers (dependency on the current job), a variable indicating the amount of labour-training the individual received (human capital) and an indicator for individuals working on the same job for a long time (job-specific human capital). These variables were insignificant in the model, but mostly had the expected signs. All this illustrates the quality of the dataset and confirms the fact the dependent variable contains information about the value the migrant attaches to the migration project.

For testing the ROTM through the nonlinearity in the dependency on income, the GEE model seem only useful as indicator to which variables are important determinants for the migration propensity: to test the ROTM one could try to add some higher order terms. This returned following coefficients and z-statistics on these variables (AINC2 and AINC3 denote a quadratic and cubic transformation of AINC respectively):

	Estimate	Naive S.E.	Naive z	Robust S.E.	Robust z
AINC	1.449664e-05	1.481355e-05	0.97860727	1.582300e-05	0.9161755
AINC2	4.738402e-09	2.826731e-09	1.67628295	3.180996e-09	1.4895968
AINC3	-1.688228e-13	9.899792e-14	-1.70531658	9.456771e-14	-1.7852055

Apparently, the relationship is nonlinear, but the specification seems to fit the data rather poorly.

Another important issue is to asses whether the assumptions of the independence of the explanatory variable holds. If there exists some relation between the independent variables, the model will wrongly attribute correlation of these variables and the dependent variable and return biased estimates, if one or more of these independent variables is left out of the model. Even when included, multicolinearity would cause estimation to be less efficient. Given the signs and significance levels of the coefficients, we could hope this problem not to be too pronounced in our case. However, since we are interested in the influence of income only for the first test and we will later use a GPLM to estimate this dependence, we should should pay extra attention to dependence between the reported income and the other independent variables. Also, since we allow for nonlinear influences, verifying if correlation between variables exists will not suffice: we should also look for nonlinear dependencies. For this, it could be usefull to have a look at figure 8, which shows a local linear kernel regression estimation of the fitted values of the GEE model we just discussed, against the reported household income (only the linear term was included). The results are drawn for three different bandwidths: blue includes 0.11% of the observations for each local regression. Green and red use 0.22 and 0.33 percent respectively. Note that this scheme will be used throughout the rest of this thesis. If the inde-

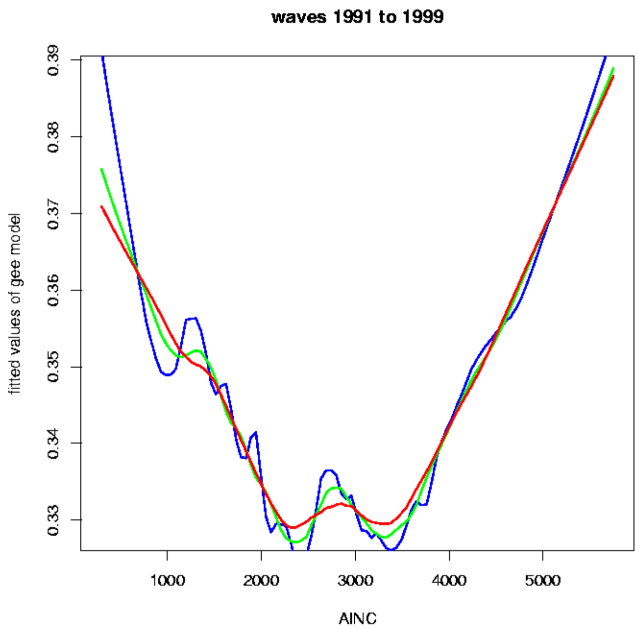


Figure 8: The fitted values plotted against income.

pendent variables would be uncorrelated, or have a simple linear dependence on income, this figure should show an approximatively straight line with a slope equal to the estimated coefficient on income. Note that even if the reported migration propensity depends nonlinearly on income, this would still hold. As the distribution of AINC changes greatly over the different waves, I could be more interesting to look at the earlier and later waves separately. Figure 9 again shows the fitted values of the GEE model, against AINC. In the left panel only waves 1991 to 1993 were included, in the left panel waves 1994 to 1999. There is obviously some nonlinear structure and this is caused by strong dependecies between the other covariates with both income and the migration propensity. Indeed, it is not hard to find some obvious relations between them:

- Having a partner has a negative influence on the migration propensity, but

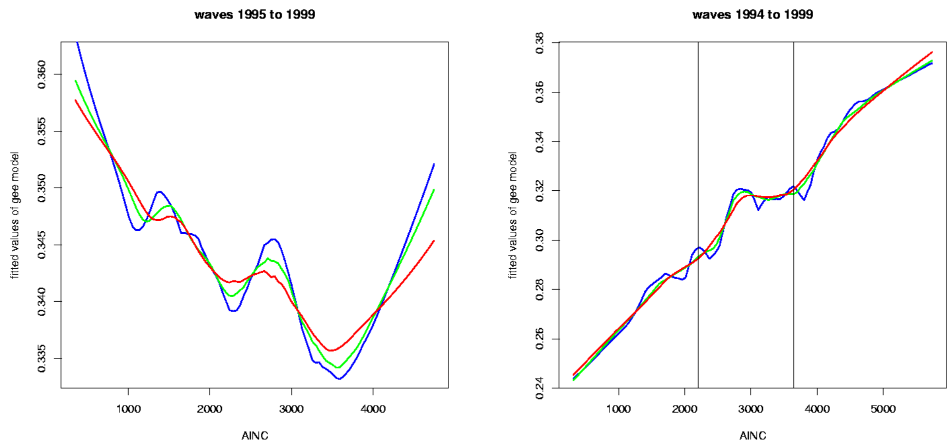


Figure 9: The fitted values plotted against income.

when using *household* income, reported low income is much more probable for singles, as there is no way for them to have a double income household. Also younger people on average earn less and are more likely to be single.

- Since men are, on average, socially less dependent (for example on children), they are more likely to migrate. At the same time, it is a well established fact that, on average, labour income for males is higher.
- Elder people are less likely to migrate, and at the same time there exists a clear (nonlinear) dependence of age on income.

To investigate these issues it is interesting to regress the suspect variable on income, preferably using some nonparametric method which allows the data to freely determine the shape of the estimated dependecy. I estimated several GAMs in XploRe. As an alternative, I repeated the GEE approach, adding variables one by one and noting those that cause a nonlinear form as in figure 8. Where adding these covariates could help constructing a model where an unbiased estimator of the dependence of the migration propensity on income could be found, it is evident it would be even better to have a measure of income that is less correlated with other covariates. Therefore, I chose to use the reported personal labour income as the variable of interest. This has the disadvantage we will no longer be able to describe the migration propensity for non-working household members, since we trimm the observations at some minimum level of income. This causes our number of observations to drop rather drastically from, for example, 3110 to 1916 in the 1994 wave. This will also allow us to use the same income variable in all models, as we will have to use personal income to model W^W later on.

Semiparametric approach As the nonlinear form we are looking for is clearly hard to approximate by including higher order terms (or other simple transformations) of income into the regression equation, I followed the approach of Burda et al. [7] and fitted a semiparametric GPLM model.

Model Description

These models have a simple linear parametric part for some variables, while allowing inclusion of some nonparametric dependence for other variables. More formally, the GPLM can be written as

$$E[P(y=1)] = G(X\beta + m(T)) \tag{41}$$

where G is some known link function, X is a matrix containing known regressors and T is a continuous explanatory variable. Unfortunately, this setup is not compatible with a panel-approach, since this would necessite a more complicated structure for the X variables (or assumptions on the error-terms). Therefore, I repeated the approach of Burda et al. and estimated the GPLM for each wave separately. To estimate m, different approaches are possible. I briefly sketch here the approach of Severini and Wong [45] and Severini and Staniswalis [44], making use of the script of Härdle et al. [21]. This approach is implemented in XploRe in the *gplmest* quantlet. To estimate β, m is assumed to be known, so that we can maximise the parametric likelihood function:

$$\mathcal{L}(\beta) = \sum_{i=1}^{n} l(G\{X_i^T\beta + m_\beta(T_i)\}) \tag{42}$$

for estimation of m, β is assumed to be known and is found by maximising the smoothened likelihood function:

$$\mathcal{L}^S(\eta) = \sum_{i=1}^{n} \mathcal{K}_H(t - T_i) l(G\{X_i^T\beta + \eta, Y_i\}) \tag{43}$$

at each point t, with $\mathcal{K}$ some smooth kernel function. The nonparametric specification could be replaced with a local polynomial one (or some other nonparametric form?). Since we assume G to be known, these likelihood functions are clearly defined and can be optimised analytically or by standard numerical methods. To find the ultimate estimators of β and m, an interative procedure has to be set up; for example the Newton-Raphson algorithm as described in Härdle et al. [21]. As initial values the resulst from a GLM fit could be used. The resulting estimator reaches $\sqrt{n}$-consistency, a level typically reached by parametric estimators. For a more complete summary of this estimation procedure, I refer to Burda et al. [7] and the Script of the non -and semiparametric methods course of the Institute of Statistics and Econometrics [21]

Three models are considered here: A GAM model containing only the reported personal income AINC as the dependent variable, a GPLM using the covariates used by Burda et al. [7] and a GPLM using a larger set of regressors.
As comparing different GPLM's for each wave, for both income variables, for different bandwidth, censoring choices and model specifications would easily double the size of this thesis, I decided to include only the estimated m for these model model specifications in the appendix.
To correctly interpret the results, some remarks have to be made. They are valid for all semi -and nonparametric models we will consider. As always with semi -and nonparametric models, the bandwidth choice can greatly influences results. The fact the distribution of the independent variable we include nonparametrically changes so drastically over the different waves (cf. figure 6) makes it undesirable to use the same censoring scheme and bandwidth choice for all waves. eg. Including observations with $AINC \geq 5000$ in the GPLM for the 1991 wave would cause a very small number of observations to greatly influence(bias?) our estimation of m. On the other hand, exclusion of these observations from all waves would make we cannot estimate m for income levels which carry a significant part of the distribution for the last waves.
Therefore, I decided to use a 'dynamic' scheme for determining the censoring and bandwith choices for each wave. *Exactly the same scheme was used for all models.* For the 1991 wave, observations were censored at $200 \leq AINC \leq 4750$. For each subsequent wave, the lower and upper bounds were increased with 25 and 250 respectively. For the 1999 wave this implies censoring with $400 \leq AINC \leq 6750$.
Estimations were done for different bandwidths, for each of the three model

specifications. As the range of $AINC$ changes with every wave, a fixed bandwidth choice would be unwise. For each wave and model, three estimations were made, using bandwidths which are fixed proportions of the range of $AINC$ for that wave: The blue curve shows $\hat{m}$ for $h = 0.11(max(AINC) - min(AINC))$. The green and red curves give the results of estimations with proportions 0.22 and 0.33 respectively. We will now look at the different model specifications and their corresponding $\hat{m}$ [12] before interpreting the results.

The first specification includes only $AINC$:

$$E[BIN = 1] = G(m(AINC))$$

this is a GAM with a single nonparametric term. Estimation was done using the *gintest* quantlet. Figure 18 shows the estimation of $m(AINC)$ for each wave.

The second model was constructed to be comparable with the model used by Burda et al. [7]. Unfortunately, the highly significant 'city size' variable could not be included, as the DIW changed its policy and imposed additional restrictions for the use of regional data [13]. Figure 19 shows the estimation of m for following model specification:

$$\begin{aligned} E[BIN = 1] &= G(SEX + PARTNER + OWNER \\ &\quad + FF + JEOP + ENV + UNI + AGE + m(AINC)) \end{aligned} \tag{44}$$

The last model includes more regressors:

$$\begin{aligned} E[BIN = 1] &= G(SEX + PARTNER + OWNER + FF + JEOP \\ &\quad + ENV + UNI + AGE + ATT + QUALW \\ &\quad + KID + SAME + WHH + SATI + SATD \\ &\quad + SATJ + REQTR \end{aligned} \tag{45}$$

<u>The Results</u>

Comparing the estimated form of m for the three different models, over the

[12] we denote an estimated variable with a hat (ˆ) above it

[13] I thank Sabine Kallwitz from the DIW for her effort in trying to find a solution for this problem.

nine different waves reveals a remarcable common structure. The green cuve (22%, medium bandwidth), seems the most 'reasonable' bandwidth choice: the blue curve (11%, small bandwidth) is rather erratic. The red curve (33%, large bandwidth) ignores much of the local bahaviour of m. Together(but especially in the medium bandwidth case) these estimations suggest m starts rather high, decreases in $AINC$ and reaches a first local minimum somewhere between 1000 and 2000.(recall that the scale is different for the different waves, when comparing the images!). Then m rises (although not so clearly for the last model specification), before a second kink appears somewhere between 3000 and 4000. Sometimes there are no 'local minima' around 1500 and 3500, but in almost every case there is a clear 'downward deviation' from the global trend of m around these values. Note that we could narrow down the bounds on the regions where this specific local behaviour is observed: obviously, the local minima are reached at somewhat lower income levels in the first waves compared to the later ones, however, I will simply refer to 1500 and 3500 from now on.
Furthermore, it appears that adding more covariates to the model is not very productive: where observing some consistent and fundamentally different $\hat{m}$ for the large GPLM specification would invite to be further investigated, the fact that the estimated relations for the different waves are quite different is probably caused by interdependencies between the covariates. Especially the ATT variable has a strong influence on BIN and is also strongly related to income. As we saw before, excluding such a variable could cause bias. On the other hand, inclusion will make estimation less efficient and this seems to have caused the sometimes bizar results in the third model specification. It should be noted that here too, the local downward deviation from the trend at income levels 1500 and 3500 is observed.

Comparing the results with the ROTM predictions from figures 2 and 3 reveals the data does not reflect the predicted dependence of m on W. On the other hand, the data neither offs support for the theory we treated as the sole competitor of the ROTM: the NPV [14].
However, if we ignore observations with $AINC$ below the first 'kink' at 1500 the picture greatly changes: the resulting estimation would be quasi identical to the

[14] It is evident that when the data prove the NPV wrong, this does not imply the ROTM must be valid or vice versa.

ROTM prediction in the specification of Burda et al. [7]! Of course, simply discarting these observations would be disastrous, as $AINC < 1500$ carries enough observations (cf. figure 6) to exclude the local behaviour of $\hat{m}$ is only due to some border-effect. This should also be seen in the light of the size of local deviation and its persistence over the different waves, bandwidths and model specifications. Also $AINC$ has already been censored from below at reasonable values (200 to 400), to avoid outliers to greatly influence estimation. I have tried quite a number of changes to censoring, bandwidths, and model specifications not mentioned here, hoping to find some model that would return the ROTM-predicted form for $\hat{m}$. The deviation from the ROTM-prediction for lower income levels turned out to be very robust over a broad class of specifications not included here.

Ordered response In this section, we will treat the response variable as continuous and apply models that were developed for such response variables, ignoring the violation of the assumed normal distribution on the error-terms. The model would still be described by (41), with G the identity function. By taking this approach, we could use all information contained in the dependent variable, and would still be able to derrive coefficients on our dependent variables (and a possibly nonlinear estimation of m). On the other hand, it is hard to say how reliable these estimates will be. Therefore all results from this section have little value beyond use for exploratory purposes.

As in the binary case, parametric estimation of the panel model is possible using the GEE function in R. Alternatively, the LME function could also be used (it can not handle binary response variables). As the results turned out to be almost the same for LME and GEE and are quite similar to the binary response case (apart from the z-values, which have no valid interpretation given the severe violation of the normality assumption), I will not repeat the results here. Here too, inclusion of higher order terms turned out to be unproductive.

What seems more interesting for our purposes, is repeating the GPLM approach we also used with the binary response, specifying G as the identity link. This would then be called a partial linear model, or PLM. Estimation was performed using the *intestpl* quantlet in XploRe, using a local linear nonparametric specifi-

cation. Figures 21, 22 and 23 in the appendix show the estimated m [15], for the different waves and the three bandwidths described above. Recall the censoring scheme is the same for all models. Note that as always, the figures are ordered rowwise for the different waves.
The PLM estimated dependence of the migration propensity looks very similar to the binary response GPLM estimates: a steep descent to a minimum around 1500, then a clear upward trend, interrupted by a kink around 3500. We conclude that the form of $\hat{m}(AINC)$ offers support neither for the ROTM nor for the NPV models of migration, under the assumptions we made on $W(W^W(W^E), W^E)$.

Testing the nonlinearity Up to this point we have seen that there seems to be some persistent nonlinear dependency of the migration propensity in income through the different waves of the GSOEP, which seems to contradict the predictions of the ROTM from figures 2 and 3. It must be noted, however, that a GPLM model in general will not produce linear estimates, so we must investigate if the deviation from the linear case is significant enough to reject a linear dependency. This is possible using a likelihood-ratio test, developed by Härdle, Mammen, and Müller [20]. This test is used by Burda et al. [7]. The procedure is summarised in the non -and semiparametric methods script [21]. A simple Likelihood ratio test would not work for a GPLM, since we would be comparing a smoothened with a non-smoothened estimate and the former will always be biased. To correct for this we have to use a bias-corrected parametric estimate for the likelihood test. As the test-statistic converges to normality (as proved by Härdle, Mammen, and Müller [20]), but the rate could be rather slow (as shown by Müller [32]), bootstrap methods could be used to calculate an approximate distribution for the test-statistic under the null hypothesis (the GLM specification). This is implemented in XploRe in the *gplmbootstraptest* quantlet.

Using this quantlet gave mixed results. As the bootstrap is computationally intensive, I only estimated the test statistics for some waves, bandwidths and model specifications.
Figure 10 shows the GLM fit and the (0.33% bandwidth) GPLM estimate for

[15] As in the binary case, the first model is not really partial linear, as there is no linear part. Thus we actually have a simple nonparametric model. Estimation of this model was done using the *lpregxest* quantlet in XploRe. Note that this model is sometimes refered to as a PLM, like the GAM is sometimes refered to as a GPLM to avoid pedantic repetition.

wave 1994 and 1999. The bootstrap test rejected the GLM on the 0.001 level

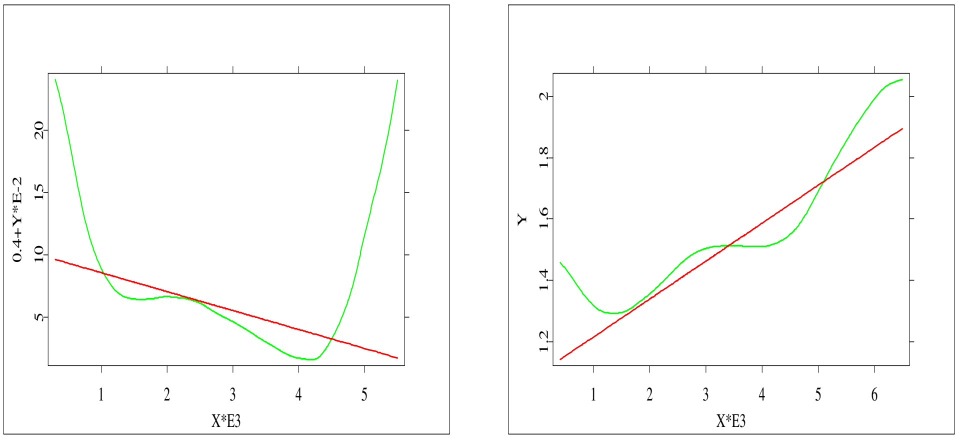


Figure 10: Comparing the GLM and GPLM fit

for the 1999 wave, but not at all for the 1994 wave(only on the 0.64 level). This illustrates how dangerous it can be to interpret a large deviation from linearity as a strong proof for a true underlying structure per se.

4.1.4 Estimation using generated W

As mentioned above, the choice of the exact variable used from the GSOEP dataset to proxy W in the ROTM is an important matter. In this section we will estimate W^W, the anticipated wage in West Germany, from the data, as described in section 2. With this results we will investigate the possibility to construct W from this data, using some specification for $W(W^W, W^E)$.

Methodology Burda et al. [7] consider some specifications of W^W as a function of W^E. To see how crucial the exact assumption on $W^W(W^E)$ [16] is to the model, one could consider the influence on the migration propensity of a change in W^E alone -keeping W^W constant-, and the repercussions of this on the migration propensity under the asumption that $W = W^W - W^E$: it would cause higher W^E to imply a lower migration propensity! This is clearly not what we see in the data, as we saw in figure 19 for the GSOEP, and will become clear for the ethnic Germans dataset, as illustrated in figure 14. If $W^W = aW^E$ with $a > 1$

[16]Note that although we notate $W^W(W^E)$, we do not explicitely estimate W^W from W^E, but rather investigate the estimated W^W, for each level of W^E.

is assumed, this would imply the dependence between the migration propensity and W^E to be exactly like in figure 2 and 3, with W on the X-axis, replaced by (a rescaled version of)W^E.
Whatever the approach taken, we should try to define exactly what form of the estimated dependence of the migration propensity on W or W^E, would offer support for the ROTM. We saw how the different models estimate the relationship between W^E and the migration propensity from the data; now we could evaluate what assumptions would be needed on W^W and $W(W^W, W^E)$ to produce this dependence. This approach is not entirely satisfactory from a methodological point of view, since it is no longer clear what estimations would falsify or verify the ROTM. Therefore we will try to estimate $W^W(W^E)$ from the data.

The different models we looked at, the GEE, GAM and GPLM for the binary response variable and the LME, GEE, nonparametric and PLM model for the ordered response, fit the data rather well, so we can assume the positive coefficient on AINC, or the on average rising estimated m, reflects a truely existent positive relation of the migration propensity and the reported personal income. Furthermore, we know from the migration theory discussion that any reasonable migration model predicts a positive dependence of the migration propensity on the wage difference W. Combining these two findings makes clear that under the $W = W^W - W^E$ assumption, any reasonable specification for W^W should be 'on average' rising in W^E, with a rate higher then unity. The 'on average' is rather important, since the data suggests a nonlinear dependency cannot be excluded. Also, since income is all but uniformally distributed, predominance of observations of some income region with some different local bahaviour would severely bias any linear estimator of $W^W(W^E)$. This effect would be aggravated by dropouts correlated with the reported income. Therefore, I will try some methods which allow to analyse the estimated $W^W(W^E)$.
As mentioned a number of times above, to estimate W^W, we will use the observed wages of migrated East-Germans in the database and their number is rather limited. Therefore, drawing hard conclusions from this section would be unwise. In this light it seems reasonable to use results from other research on this subject. Unfortunately (for our purposes), attention has gone mostly to comparing immigrants with the local population and researching the causes of differences of the labour renumeration between these groups. Interesting approaches in this

direction, using data from the GSOEP are, for example, Dunn, Kreyenfeld and Lovely [14] and Burda and Schmidt [8] (which use a Mincer approach).
Schwarze [43] recognises the importance of the model for W and constructs a model where the uncertainty in W^E is explicitely modeled through 'discounting' of W^E for those individuals that report a high subjective estimation of the probability of job loss in the foreseeable future. We will ignore this issue and take an approach more similar to Davanzo and Hosek [10] or Tunali [46].

First, we will look at the classic two-stage Heckman [23] estimator, then the use of the semiparametric estimators of Powell [38] and Andrews with Schafgans [2] will be discussed. The reader should consult section 2 and the references therein for a discussion on the how and why of these estimators. Here I will only briefly summarise that self- selection models try to assess what the effect of some measure would be on some randomly chosen subject, given that data is only available on self-selected individuals. Note that in a migration context it is unclear if this 'expected treatment effect' is relevant for the migration decision, since it can not be excluded migrants behave irrational (or, alternatively, make wrong anticipations of their possible wage in the region of destination).
The first step of the Heckman estimator consists of a simple probit model, describing the 'selection' into the 'treatment'; for us this would be the migration decission. The results could than be used to correct for self selection in estimating the treatment effect.
What I did not discuss in section 2, but should have become obvious now, is that for our application of the theory to the GSOEP data some questions arrise about the relevant treatment effect of migration on income. One approach could be to compare pre-migration and post-migration income for the migrants in the panel, but what should be taken as the treatment effect? The earned wage one year after migration occured? Two years? An average of all available years? This is an important issue since human capital models of migration predict steeper income curves for migrants, as they (have to) invest more in new human capital upon arrival. Also, one could ask if increases in the general price level and wages should be accounted for, since we would be comparing (differences in) income levels of different years.
For these reasons I decided to look at the two groups (migrants and stayers) for each wave separately and define the treatment effect as the measured in-

come difference between them, ignoring the exact point in time where migration occured. This approach should be less sensitive to the time-wage curve differences, since we would be looking at wages of people that have been living in West Germany for different lengths of time. Of course this would not solve this problem completely. It could be hoped this will not cause severe problems in the case of German East-West migration, as the measured differences in earnings are probably not primarily caused by necessary investment in human capital, but rather reflect purely regional differences that are independent from migrant characteristics, or alternatively, by a regionally different renumeration of some human capital characteristics, which can be accounted for in a self-selection setup. Nevertheless, it should be emphasised that by taking this approach, we are not constructing a completely satisfying approximation of W^W, which would be the expected immediate wage gain from migration (controlling for the time-curve problem), but rather the expected value of what earnings the possible migrant would have, at this point in time, had she migrated some time ago, where the 'time ago' definition is unclear since the subjects used for estimation migrated at different times in the past. Again, I chose to mention these consideration, but the problem will probably be less pronounced in the case of the GSOEP data.

Turning to the estimation itself: for the household income case, 3669 people had a valid entries in the 1991 wave, this diminishes to 2794 for the 1999 wave (excluding migrants). By changing the independent variable of interest to personal income, the sample size diminishes rather drastically: from 2618 in 1991 to 1572 in 1999. For the self-selection approach, we are interested in what causes people to migrate and how much they gain from it. Unfortunately, the number of migrants is very limited in the household income case (33 to 146) and even more so in the case of personal income (26 to 126) (cf. table 5). These numbers make clear that estimators will suffer from bias due to nonrandom drop-outs and be rather unreliable because of the very limited number of observations. Despite all this, I bravely continued estimating since, as noted before, we could limit ourselves to look if the results greatly contradict the assumption that $W^W = aW^E$ with $a > 1$, or similar assumptions on W^W and W^E.

estimation All estimations in this section were conducted in XploRe. The first approach we discussed in section 2 was the two-step Heckman estimator, which is implemented in XploRe in the *Heckman* quantlet. To estimate the treatment effect, we need to construct a model for both the outcome and selection part. Variables may be part of both the outcome and selection equations. For the selection (migration) part I choose as independent variables SEX, PARTNER, OWNER, FF, JEOP, ENV, UNI, AGE, INC, ATT, QUALW, KID, SAME, WHH, SATI, SATD, SATJ and BIL [17]. Using this model for the probit part, a $\overline{R}^2$ of 0.25 was obtained. For the outcome (income), the variables SEX , AGE , UNI , JEOP, TREIMAN, PART, BLUE, SAME, AGESQ, HOURS, SELF, WHL and WHH were used. AGESQ is AGE squared. Figure 11 illustrates the

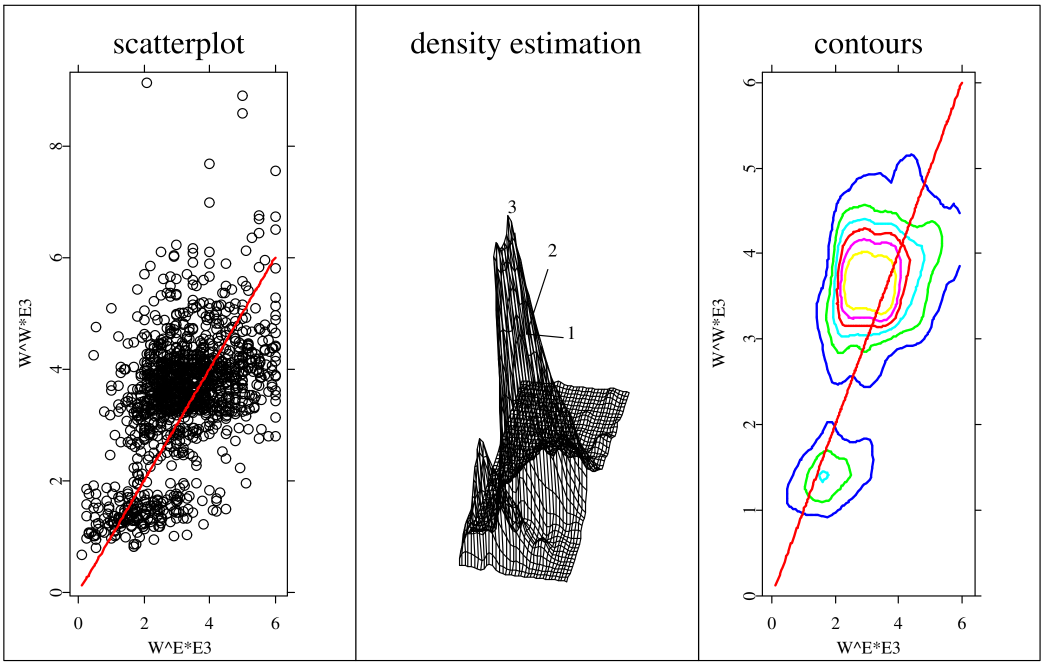


Figure 11: Heckman self-selection-corrected estimates for W^W for the 1999 wave plotted agains the observed income W^E;scatterplot, density estimation andcontour plot. The red curve shows $W^W = W^E$

results, for the 1999 wave. The left side plots the expected W^W given the covariates, against the current W^E for each individual. The red curve is $W^W = W^E$: individuals with an estimated W^W above this curve are predicted to gain (in income) from migration. In the middle, a density estimation (using a quartic

[17]see table 6 in the appendix for an description of these variables

kernel, bandwidth 600), of the distribution is given. The right panel gives a contour plot of this density, where the relation between the two variables can be seen more clearly. Note that, as always, the observations have been censored with respect to $AINC$. The figure makes clear that even in 1999 a majority of individuals is predicted to gain from migration in terms of income. For earlier waves, the point cloud is more concentrated and more points lie above the red curve: with a larger average wage gap, more people are predicted to gain from migration. (cf. figure 25 in the appendix). Note that the estimated relation is positive, but the spread in predicted W^W for some income (class) is quite large. For the lower income levels, an almost horizontal relation is estimated. In the left panel of figure 12, a kernel regression (local linear, bandwidth 1500) is added to the contour plot. As there are a lot of outliers in the estimated W^W for each level of W^E, this estimation could be rather unreliable. The right panel shows a

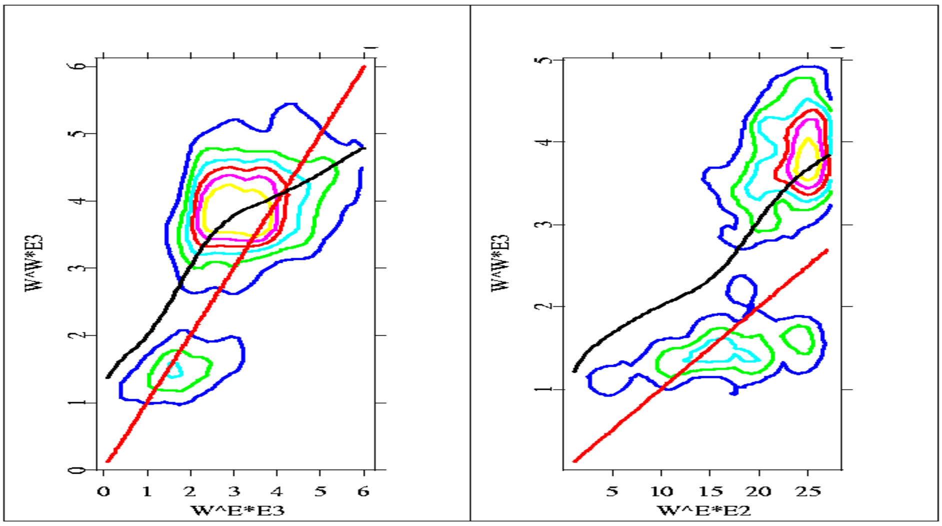


Figure 12: Contour plot of the estimated $W^W(W^E)$

more detailed view of the contour plot from figure 11 for income $AINC < 2700$. For the 1994-1999 waves contour plots have been added in figure 24 in the appendix for the whole $AINC$ range. A more detailed view on the distribution for $AINC \leq 2700$ is also given there, in figure 25. As allways, the different waves are ordered rowwise. Estimation for the first waves proved unreliable, or even infeasible, due to the low number of observations. Therefore the first wave included in the figures of the appendix is 1994. The most important observation is that the estimated income gain first diminishes in W^E, before rising. Also there seems to be a kink in the trend of the predicted wage gain around $AINC = 3000$. This correspondents to the AINC levels where we observed 'anomalities' in the different estimations of m (cf. figures 18 and related figures in the appendix). This could be due to a deficiency of our approach: for extreme income levels, our predictor variables for income for the migrants are probably a bad approximation for what the individual could really earn in West Germany. For example: a individual with an above normal feeling for business oportunities will have a high W^E, but our limited set of explanatory variables will probably not predict a high W^W for such individuals. Also, since (quasi-) [18] unemployment does not only hit people with a low educational background or other characteristics we control for that are highly correlated with a low income, there will be a significant number of individuals with a relatively low income, which our model predicts would benefit greatly from migration. This is because the assumption that the person would find a job in the West is also reflected in the prediction. Or, a bit more absurd, that some dissabilities or other causes or reasons of unemployment, would simply dissapear upon migration. On the other hand, if (quasi-)unemployment is mostly involuntarily, the specific form of the general trend in the dependence of the predicted W^W on the current earnings could reflect low-income individuals really have a lot to gain from migration. As this is certainly the case for some (most?) of these individuals in the sample, we conclude the estimated high income gain for individuals with a very low W^E probably at least partially reflects a true underlying structure. As a similar effect (involuntarily high income?) can not be assumed for higher income levels, the estimated $W^W(W^E)$ is probably underestimated for high income observations. This could also explain why high migration propensity for lower income levels was not observed for the first wave, if involuntary unemployment was less frequent in this very first wave.

All in all, this greatly contradicts the assumption of $W^W = aW^E$ with $a > 1$. Therefore, under the $W = W^W - W^E$ or similar assumption, if the estimated expected gain is the gain people think about when they answer the questions about their future migration plans, both classic NPV models and the ROTM would predict the migration propensity to be negatively related to the current W^E for lower income levels. The form of $W^W(W^E)$ therefore could explain the

[18] note that there are no observations on unemployed individuals included in our samples, as we censor at positive labour income values. However, we will call individuals with extremely low labour income levels quasi-unemployed or simply unemployed.

first observed deviation of $\hat{m}$.

As it is unclear whether the 'extreme income' results are caused by the limitations of the model we just discussed or rather reflect a true underlying structure, I decided to take another approach and compare the generated W^W with generated W^E, rather then the true W^E. In this way, effects caused by the limited capabilities of our income models should be less pronounced: we now compare the predicted wage in the West, based on a regression on some covariates in the migrant group, with the predicted wage in the East, based on a regression of the same covariates in the non-migrant group and account for the fact the migrants are a nonrandom group. Note that in the simple Heckman model from figure 11, a bad model specification would cause W^W to be less correlated with W^E. Therefore, a bad model specification would make us draw the conclusion that the obtainable income gain diminishes in the reported W^E. This is a very undesirable property (also given the known limited available observations). The alternative approach with a constructed W^E should be less sensitive to this, since it imposes the exact same 'handicap' on both variables. Figure 13 illustrates the estimated W^W using this technique, again for the 1999 wave. Again, it is clear from the figure that $W = W^W - W^E$ first diminishes, and then rises for higher levels of W^E. Figures 26 and 27 in the appendix give the contour plots and a close-ups for $AINC \leq 2700$ for all waves. Recalling the remarks made above about the underestimation of the income gain for higher income levels, these figures seem to confirm the hypothesis in Schwarze [43], that 'the procentual increase of the wage by migrating to West Germany(...),follows an exponential curve with respect to the wage in East Germany' [19], *but only for higher income levels*. In all later waves, the income gain first decreases in W^E for low income levels. We should probably pay more attention to the last waves, as they contain more observations on migrants and will also suffer less from the 'time-wage' problem discussed above. Also our wage equation is probably not well suited to explain W^E for the first waves as the economic structures of both regions were still very different for those waves.
As emphasised in section 2, it is all but certain the migration candidates consider the fact migrants are not a random sample when making anticipations on the

[19]'Der prozentuale Zuwachs des Lohnes bei der Migration nach Westdeutschland folgt also, bezogen auf den Lohn in Ostdeutschland, in etwa einem exponentiellen Verlauf.'

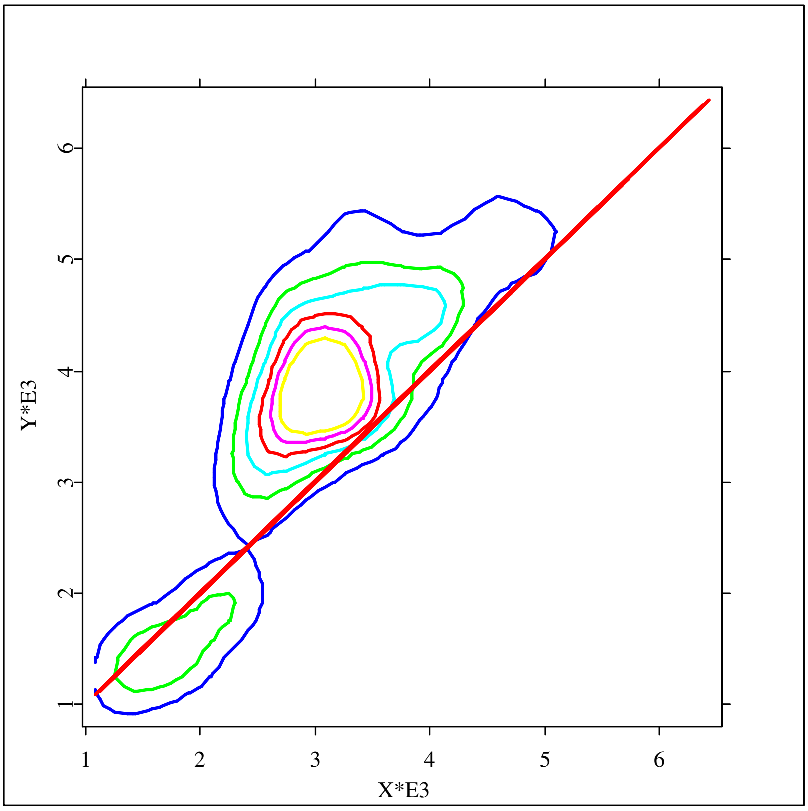


Figure 13: Contour plot of the estimated W^W plotted agains GLM estimated W^E

possible income gain from migration. Therefore I repeated the estimations without the self-selection corrections. The main results turned out rather similar to the self-selection approaches: there is a significant deviation from the assumed increasing trend in $W^W(W^E)$ for low income levels. Figures 30 and 31 in the appendix show the results. Note that we do not further investigate whether this model specification is usefull, as the results turn out to be quite similar to the other models. It could be done in a setup as Tunali [46] uses.

Interpreting the results A common characteristic of the estimated $W^W(W^E)$ is the remarkable deviation from the estimated increasing trend for low income levels: for most models there exists a range of income levels between 1000 and 2000 that have, on average, a negative predicted income gain from migration. Combining this finding with the predicted form of the dependence of U^* on W (cf. figures 2 and 3) and the estimated dependence of U^* on W^E (cf. eg figure 19 shows that the first 'dip' in $\hat{m}$ could very well be caused by the strange form of $W(W^E)$, rather than reflecting an underlying structure in the dependence of U^* on W. The fact the deviation from the global trend in $W^W(W^E)$ and $\hat{m}$

occurs at almost exactly the same $AINC$ levels over the different waves, offers support for this hypothesis. Note that this holds for a wide variety of model specifications for both the $W^W(W^E)$ and $\hat{m}$ estimations.

A logical next step would be to use the estimated $W^W(W^E)$ to calculate W under some assumption for $W(W^W, W^E)$ for each individual and repeat the estimation of $\hat{m}$ with this new variable as the nonparametric part of a GAM, GPLM, PLM or nonparametric model. As can be clearly seen from the contour plots, there is a lot of variation in $W^W(W^E)$ for each $AINC$ level. Note that this does not imply our model is bad: the contour plots compare the *fitted values* of the W^W model with observed or estimated W^E. Therefore it may very well be that the large spread in $W^W(W^E)$ for some W^E would still be observed if we had perfect models and an infinite amount of information and observations. One would tend to believe this large spread would mean the nice stucture in $\hat{m}(W^E)$ is totally lost upon transforming to W. However, the estimated $W^W(W^E)$ for an individual contains more information than contained in W^E and therefore the $\hat{m}$ using a calculated W based on individual information on W^W and W^E should be superior to using a simple transformation of W^E, to obtain W^W and W, as investigated by Burda et al. [7].
Unfortunately, estimation of GAM, GPLM, PLM and nonparametric models turned out to be unproductive: $\hat{m}$ is quite different for the different model specifications, waves and bandwidth choices. Quite often $\hat{m}$ was downward sloping over the whole W range. In principle, this could have many causes, but given the convincing different migration theories predicting a positive dependence of m on W, this is much more likely to indicate our self-selection models are simply too bad to predict W^W for each individual. As it is implausible $W(W^W, W^E)$ changes drastically between waves, the problem must be related to the estimation of W^W (nevertheless, I tried some different specifications of W, including differences in logarithms, without success.)

Alternatively, some parametric specification of $W^W(W^E)$ could be investigated, as in Burda et al. [7]. We could consider what happens to $\hat{m}$ when regressing on $W = W^W - W^E$, with W^W specified as a function of W^E as suggested by the different self-selection models: increasing in W^E, but with a significant downward deviation around $W^E = 1000$ to 2000 (perhaps the not-so-clear effect around $W^E = 3500$ could also be included). Equally important is the specification of $W(W^W, W^E)$. So far we only considered $W = W^W - W^E$, but what if $W = ln(W^W) - ln(W^E)$, as in Schwarze [43]? It is easily veryfied that both specifications would comply with the observed low income behaviour of $\hat{m}$. With the specification of $W^W(W^E)$ and $W(W^W, W^E)$, W could be calculated for each individual and the semi -and nonparamtric estimation could be repeated. It is, however, quite utopic to believe a sufficiently correct parametrisation of both functions could be found that would needly 'filter out' all bias caused by the fact $W \neq W^E$ and would allow to estimate $m(W)$ using data on W^E alone. Therefore, I chose to assume $W^W = aW^E + b$, $W^W = a * exp(W^E) + bW^E + c$, or some other simpler specification holds on average and consider where the estimated $W^W(W^E)$ suggests a deviation from this specification.

As mentioned a number of times above, the most important deviation from the trend in $W^W(W^E)$ seems to occur for the lower income levels (cf. figure 24 and related figures in the appendix). This could explain the observed first kink in $\hat{m}$ under a variety of specifications for $W(W^W, W^E)$. Especially $W = ln(W^W) - ln(W^E)$, would produce the very steep descent of $\hat{m}$ for lower income levels that is observed in the data, as the predicted *relative* wage gain is very high for extremely low income levels, and decreases very fast as the estimated absolute gain becomes negative. Under not-all-to-crazy specifications for $W^W(W^E)$ and $W(W^W, W^E)$, nothing special happens at $W^E = 3500$. However, the second main result of the semi- and nonparametric models was that there is a significant (but less pronounced) dip in $\hat{m}$ at this income level. Note that this local deviation of $\hat{m}$ from the trend was not as impressive as the low-income dip, but proved robust over the different model specifications.

It would be a great conclusion for this thesis to simply state that the ROTM-predicted specific form in the specification of Burda et al. [7] is present in the different estimates for m, if one knows the large dip for lower income levels is caused by some effect due to the strange form of $W(W^W(W^E), W^E)$ for these income levels, as suggested by the models for W^W. Indeed, if one 'filters out' this effect (eg. by simply laying ones hand over the lower income-level observations) the estimated $\hat{m}$ truely is remarkably similar to the predicted ROTM-function from figures 2.

However, it must be noted the self selection models sometimes also show a deviation from the trend for income levels around 3500. Therefore it can not be excluded this second 'dip' in $\hat{m}$ can also be explained by a local deviation in $W^W(W^E)$, rather then a local deviation in $U^*(W)$ as suggested by the ROTM. There also seemed to exist some interdependencies between the covariates possibly causing more difficult estimation around exactly this values of AINC, cf. figure 9. Unfortunately, I ignore how this could be further investigated.

This all illustrates how vulnerable the ROTM is without a complete model for the dependence of W on W^E: if the estimated local behaviour of $W^W(W^E)$ of the different models is real, we would observe a similar dependence of the migration propensity on W^E as predicted by the ROTM, in the simple NPV model without the ROTM! Indeed, it is not difficult to find some waves and model specification where the contour plot (illustrating the estimated $W^W(W^E)$) more or less has a form that would produce the estimated structure in $m(W^E)$, under the NPV assumption of a linear relation between U^* and W.

The reader hopefully noticed that we have not estimated $W^W(W^E)$ using the semiparametric methods proposed in section 2. The *adedis* XploRe quantlet implements a method developed by Horowitz and Härdle [24], that would allow to estimate the first-step single index model with binary covariates. However, this method uses differences between the results from separate single index models for each value of the dummy variables and therefore suffers from the fact we have so little migrants in the dataset. This caused the estimation to be infeasible for larger sets of covariates and made it very sensible to the model specification. Therefore, we will not look at the results. Given the remarks in Tunali [46], the fact we only look at models with the strong normality assumption could have biased our estimates.

4.1.5 Testing the shift in the trigger level

Testing through the predicted positive dependence of the trigger level on the parameter driving the variance of the underlying stochastic process of W is also not straightforward. As noted in section 3, the main challenge will be to identify variables that should be highly correlated with σ^2, as direct information is not available.

The variables I considered are SELF, PART and LOSS. These are dummy variables measuring whether the individual is self-employed, is a part-time worker and fears losing his current job in the foreseeable future (cf. table 5). As σ^2 is probably higher for individuals with these characteristics, a estimated negative dependency of the migration propensity would offer support for the ROTM. As with the test through the nonlinearity in the value function, it is important to consider if other migration models would make the same prediction. This seems not to be the case, as self-employed individuals and part-time workers are likely to be less dependent on their current job-specific human capital, the classic human capital model predicts exactly the opposite. For people fearing job-loss, this would also hold if, like in Schwarze [43], the individual 'calculates' some $E[W^E]$ using the probability of job-loss and would therefore have a higher W and therefore a higher migration propensity.
As we only need to investigate the sign on these variables, I simply repeated the GEE estimation summarised in table 2 with additional inclusion of these variables. This gave following result:

```
        Estimate Naive S.E.    Naive z Robust S.E. Robust z
SELF -8.3805e-02 5.8908e-02  -1.422648 6.3182e-02  -1.32640
PART  2.6985e-02 3.4347e-02   0.785646 3.5581e-02   0.75841
LOSS  1.9361e-03 1.2271e-03   1.577831 1.1498e-03   1.68387
```

Our dependent variable, BIN, differentiates only between low and high migration intententions. This is not desirable for testing the trigger level effect, because it is probably more pronounced for higher levels of W and U^*. The ultimate test, therefore, would be to look to the signs of the GEE model, using the migration decission MIG as the dependent variable. The resulting estimates were

```
SELF -0.372110  1.2130e-01   -3.06755  1.5430e-01  -2.41150
PART  0.371717  6.2624e-02    5.93568  7.2123e-02   5.15391
LOSS -0.002356  1.8269e-03   -1.28997  2.4034e-03  -0.98052
```

We can not draw strong conclusions on the validity of the ROTM from these estimates, as the estimated signs are mixed. It should be noted, however, that the fact that the sign on some of these variables is negative and relatively significant, greatly contradicts the predictions of the human capital models. Unfortunately, the significance levels are not impressive. Therefore, we conclude the GSOEP data offers only weak support for the predicted shift in the trigger level.

4.2 Data on Ethnic Germans

The second dataset we will use to test the nonlinearity of the dependence of the latent variable U^* on W, is a dataset provided by the Osteuropa-Institute in Munich [20].

4.2.1 Background

The dataset we will use is the first dataset in Locher [29], containing data collected in 1991. Since the second dataset Locher discusses entails no information on income, it can not be used for the fist test. The individuals on which the dataset contains information, are ethnic Germans living in Russia, a minority that has been present in different locations in this country for centuries, most notably in the Saratov-oblast near the river Volga. There even existed an 'Autonomous Socialist Soviet Republic of the Volga Germans' from october 1918 until august 1941. Already before the outbreak of WWII, but especially in 1941, Germans were exterminated as 'counter revolutionaries' and ennemies; the majority of the German population was deported to Kazachstan, Siberia and other remote areas in Russia. In 1972, ethnic Germans were allowed to return to the Volga-region. Many have tried to return to the cities from which they were banished only to find hostility and despair (source: [30] and [17]). With the event of perestojka and the fall of the iron curtain, return migration to Germany suddenly became possible and this is what makes migration intentions (and movements) of this population interesting for us to test the ROTM. As the ROTM assumption of sunk costs seems more valid in the case of migration between Russia and Germany than for internal Russian migration (and German internal migration, for that matter), the 'ROTM-effect' should be more pronounced in this case.

4.2.2 Data description

In 1991, restitution of the autonomous republic of Wolga-Germans was considered a real possibility by a significant part of the ethnic Germans (for an interesting discussion on this, see [30]) and a lot of questions in the dataset concentrates on this issue. One question, however, is whether the subject wants

[20] I thank Barbara Dietz of the Osteuropa Institute in Munich for providing the data and Lilo Locher for letting me use the transformed dataset she used in her papers [29]. Without the availability of this ready-to-use dataset from Lilo Locher, and her helpfull comments, I would not have been able to investigate the dataset in this thesis.

to migrate to Germany and if so, what steps she has taken to so far. As with the GSOEP data, we will assume a binary transformation of this variable to be driven by the value U^* of the ROTM, depicted by the yellow line in figure 2 or 3. Note that this is the same variable as Lilo Locher uses. For the dependent variables too, I simply used the same set as Locher. Summary statistics on these variables are given in table 4. A description of the variables is provided in table 3. The choice of the income variable is much simpler compared to the first dataset, as we only have information on the reported average monthly income. The dataset contains only a single observation for each individual, therefore we do not have to account for a panel-structure of the data as in the case of German internal migration. Since I did not have any information on the earnings of migrants, setting up a self-selection or other model to construct anticipated or expected wage gains was not possible. It must be noted that here too, an alternative approach could have been to use data on wages of West Germans for this. As explained in section 2, this would be equivalent to modeling a specific (probably irrational) anticipated wage gain model, in contrast to the expected wage gain model using self-selection corrected data on migrants. Such an approach would be difficult to implement, as renumeration was probably still based on quite different characteristics in Russia and Germany in 1991. This would make it difficult to construct inidividual specific predicted wage gain predictions. Therefore, I chose not attempt such an approach.

4.2.3 Estimation

The properties of the dataset make the choice of the model to use for estimation much easier compared to the first dataset. As above, we could estimate some refined GLM and add some higher order terms of the income variable and test if they offer proof for the specific form of the dependence on income we are looking for.

But, as became clear when testing with the first dataset, a GPLM offers a much more flexible solution. Therefore I will consider only the GPLM estimation and also look at some changes in the bandwidth. As with the first dataset I also look at three different specification of the parametric part to investigate the robustness of our estimates. The results largely confirm the specific form of U^* as predicted by the ROTM, in the version illustrated by figure 2 (or some version of 3) *assuming* $W = W^W - W^E, W^W = aW^E, a > 1$ *or similar*. However, it should

Variable name	Description
y	Response variable
AGE90	Age of the respondent in 1990
AGE90SQ	Age squared of the respondent in 1990
EDU	Levels of education degrees in the Soviet system, increasing from 0-6
MARR	Married (1) or not (0)
KID	Dummy for having children
INC	Average monthly labor income(censored at $60 <= INC <= 750$)
MATCH	Match Measure in how far education and experience can be used at present job, increasing from 1 to 5
MRUS	Married to a person who is not of German nationality(a Russian in the overwhelming majority of cases)
GER	Being German native speaker
RELIGION	Being member of a church, excluding Russian orthodox(mainly protestant churches)
WGEB	Member of Wiedergeburt (union of ethnic Germans, tried to reestablish the autonomous Volga republic)
RELG	Respondent has relatives in Germany
CITY	Respondent lives in a city or urban area
RUSS	Respondent is from Russia
BIN	Binary transformation of the dependent variable $BIN = 1$ if $y == 3$ for an individual

Table 3: Variable description for the ethnic Germans dataset (after Lilo Locher [29])

be noted that this analysis is made with the observed wages in Russia only, and therefore the remarks made in section 2 are still valid: it is unclear whether the nonlinearity is caused by the ROTM-effect or by some form of the relation of the anticipated wage gain for people and the reported income. Furthermore, we observe only one deviation from the trend in $\hat{m}$, at INC=400. A possible explanation could be that the income gain is higher for all income levels for the Russian-German migration case (as income differences are so much larger) and also possibly a steeper function of W^E. This would make the form of the income gain for 'low-income' levels relatively less pronounced and less important (depending on the utility specification $W(W^W, W^E)$). In combination with the fact that the ROTM should be more pronounced because of the higher sunk costs, one could than interpret the measured nonlinearity in the dependence of U^* on W for this dataset as a reflection of the predicted form in figure 2. Yet another explanation could be that the measured 'dip' is nothing else than exactly this

[1,]						
[2,]		Minimum	Maximum	Mean	Median	Std.Error
[3,]		----------	----------	----------	----------	----------
[4,]	y	1	3	2.3343	3	0.76432
[5,]	AGE90	22	60	36.445	36	8.6457
[6,]	AGE90SQ	484	3600	1402.9	1296	671.35
[7,]	EDU	0	6	3.6322	4	1.1414
[8,]	MARR	0	1	0.84196	1	0.36504
[9,]	KID	0	1	0.74266	1	0.43748
[10,]	INC	60	750	272.58	250	127.54
[11,]	MATCH	1	5	2.9524	3	0.7065
[12,]	MRUS	0	1	0.34266	0	0.47493
[13,]	GER	0	1	0.58392	1	0.48574
[14,]	RELIGION	0	1	0.29231	0	0.45514
[15,]	WGEB	0	1	0.15385	0	0.36105
[16,]	RELG	0	1	0.69371	1	0.46128
[17,]	CITY	0	1	0.55804	1	0.49697
[18,]	RUSS	0	1	0.5007	1	0.50035
[19,]	BIN	0	1	0.51469	1	0.50013
[20,]						

Table 4: Summary statistics for the ethnic Germans dataset

'low-income' effect. Given the argumentation for the second specification of the ROTM, that was illustrated in figure 3, this seems the most plausible explanation. This shows once again how difficult it is to draw conclusions on the ROTM from some estimation on W^E without a model for $W(W^W(W^E), W^E)$.

Figure 14 compares the GPLM estimated nonparametric dependency of U^* on the reported income, $\hat{m}$, for three different bandwidths. This clearly shows that the estimates are quite robust to changes in the bandwidth parameter. Note that these bandwidths correspondend to 14, 29 and 43 percent of the support of the reported income. Only observations with an income between 60 and 750 and age between 22 and 60 were included. The sample size was 643. I give here the output of the *gplmout* XploRe quantlet in figure 15. *Gplmout* gives some interesting statistics from the GPLM estimation [21]. Here too, the coefficients on the variable are as expected and are comparable with Locher's, as can be seen from the left panel. Again, this is not surprising given the relatively little

[21]For the GSOEP dataset the estimated coefficients were as expected in the vast majority of cases. the coefficients on the variables, are given in table 2 for the panel model

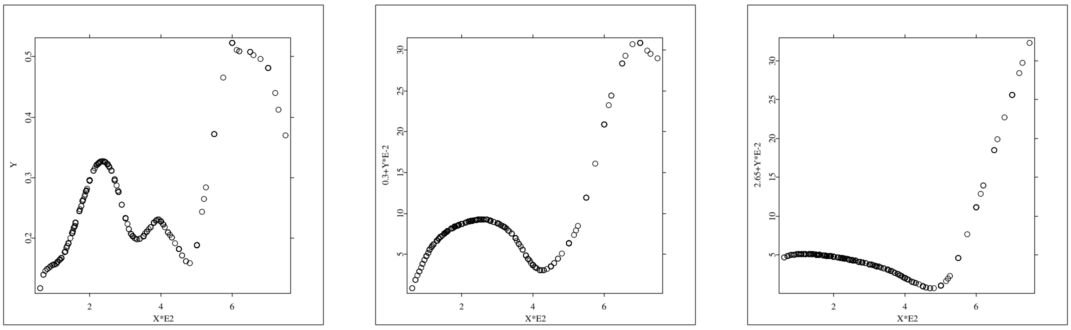
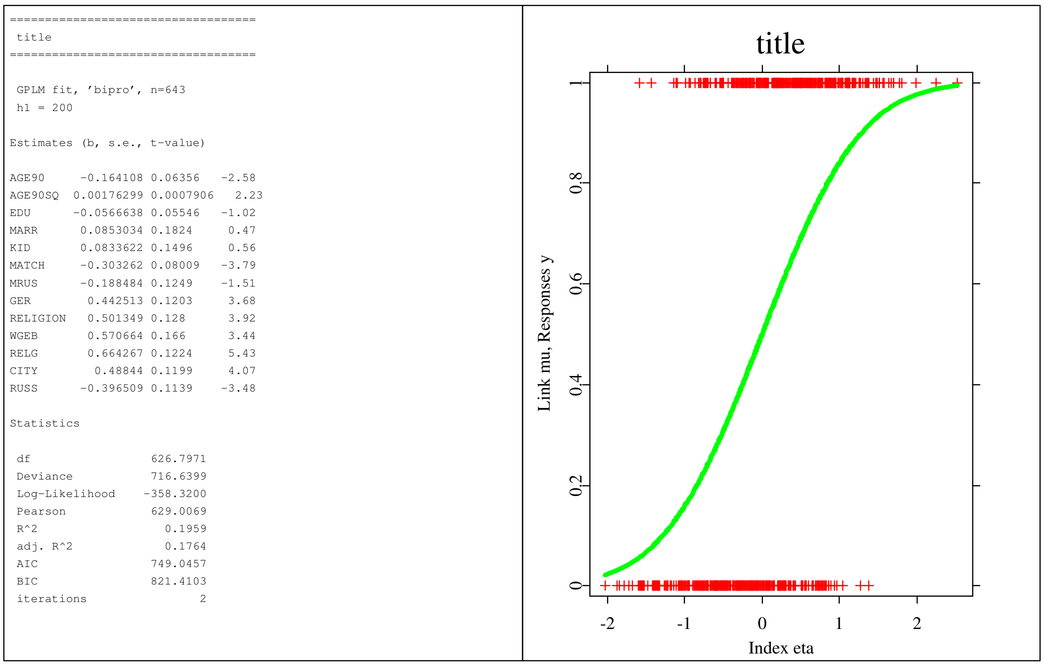


Figure 14: The GPLM estimated dependence of U^* on income, for bandwidths 100, 200 and 300

influence of income on the migration propensity. In the left panel of figure 16 the



```
===================================
 title
===================================

 GPLM fit, 'bipro', n=643
 h1 = 200

Estimates (b, s.e., t-value)

AGE90      -0.164108 0.06356   -2.58
AGE90SQ   0.00176299 0.0007906   2.23
EDU       -0.0566638 0.05546   -1.02
MARR       0.0853034 0.1824     0.47
KID        0.0833622 0.1496     0.56
MATCH      -0.303262 0.08009   -3.79
MRUS       -0.188484 0.1249    -1.51
GER         0.442513 0.1203     3.68
RELIGION    0.501349 0.128      3.92
WGEB        0.570664 0.166      3.44
RELG        0.664267 0.1224     5.43
CITY         0.48844 0.1199     4.07
RUSS       -0.396509 0.1139    -3.48

Statistics

 df                    626.7971
 Deviance              716.6399
 Log-Likelihood       -358.3200
 Pearson               629.0069
 R^2                     0.1959
 adj. R^2                0.1764
 AIC                   749.0457
 BIC                   821.4103
 iterations                   2
```


Figure 15: Output of the GPLM quantlet

variables MARR, KID, MRUS, RELIGION, WGEB, RELG, were dropped to see how sensitive our estimation is to the exact specification of the parametric part. As argumented in the context of the German East-West migration, it is a well known fact that dropping variables (or not including them in the first place) that are correlated to both income and the migration propensity will bias the (non - semi or completely parametrically) estimated dependence on income. Therefore, in the left panel, only variables we could assume to be relatively uncorrelated with income were dropped. [22] In the right panel, the parametric part was left empty, so the panel shows a simple nonparametric (local-linear, quartic kernel) estimation of the influence of income on the migration propensity.

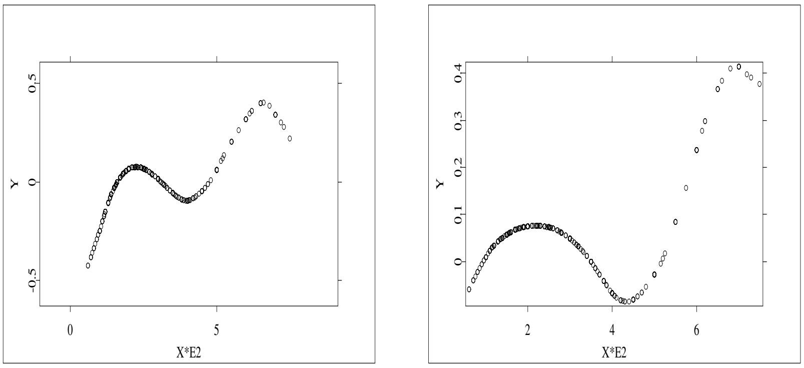


Figure 16: Dropping covariates: a smaller and a single covariate model

4.2.4 Testing the shift in the trigger level

Due to time constraints I investigated the predicted positive dependence of the trigger level for the ethnic Germans dataset only briefly. There are some variables in the dataset that could possibly be used as proxies for σ^2. Eg. we know whether an individual is self-employed and if she is a member of some trade union. As with the GSOEP dataset, it could be questioned if such variables really offer an approximation of σ^2. But for the data on ethnic Germans, estimation additionally proved difficult because there is only one (sic!) individual reporting to be self-employed and the trade union variable contains almost no valid observations. Therefore, I chose not to further investigate the second test for this dataset.

I refer to the work of Locher [29] for some tests we discussed briefly in section 3, but did not implement. Locher also looks at the effect of changes in the subjective estimated remaining time T of the option to migrate (we assumed infinitely living options). The results of Locher offer support for the ROTM: the number of migrants seems to rise if people believe the option will expire sooner.

[22] The figure shows estimations with a bandwidth of 200, but results were comparable for a large class of bandwidths.

Conclusion

In this thesis, I have tried to extend the real option model of migration of Burda [6], by adding a model for the anticipated income gain for the potential migrants. The suggestion of Burda to develop the theory in a dynamic programming framework was followed. Possible tests for the theory were investigated formally. Two test were found promising, as they base on predictions of the real options theory that are not made by other migration models. The first was the upward shift in the trigger level of migration (intentions), caused by increases in the subjective estimation of the variance of the stochastic process driving the wage differentials between the regions. The second was the specific nonlinearity in the dependence of the latent variable thought to drive the migration (intentions) on the anticipated income difference.

For the empirical implementation of the first test, I closely followed Burda, Härdle, Müller and Werwatz [7]. However, I argument that the form of the nonlinearity of the dependence on the anticipated wage gain that would offer support for the real options approach to migration is different from the one proposed by these authors. It must also be noted that in the alternative specification, the dependence of the migration intentions on the income gain would be harder to discern from the linear case.

Two different datasets were used for the empirical tests. The first dataset was used to model German East-West migration intentions. This dataset contains data which allows to estimate a model for the anticipated wage gain for each individual. This way we were able to evaluate assumptions on the dependency of the anticipated wage gain and the reported income. The second dataset contains data on ethnic Germans in Russia. For this dataset, some assumptions on the dependence of the anticipated wage gain on measured income for this dataset were necessary, as it lacks information on income of migrants.

Following Burda et al. [7], I first estimated the dependence of the latent variable driving migration (intentions) on (among other variables) the measured income, using different model specifications and allowing the data to freely determine the exact form of the nonlinear dependence through a semi-parametric approach. The estimated form was found to be quite robust over different model specifications and the different waves of the dataset. However, it offered no support for the real options approach to migration, as it contradicts the predictions, in both the specification of Burda et al. and the proposed alternative, under the assumption the wage gain depends positively on the measured income.

The results of the separate model for the anticipated wage gain for the first dataset suggests a highly nonlinear relation between the anticipated wage gain and the measured wage gain. Here too, the specific form proved robust over different model specifications and the different waves. Some theoretical considerations were made, that offer support for the specific nonlinear form that was estimated; the most important being the fact that individuals with a low income are often involuntarily unemployed and therefore have a higher anticipated or expected wage gain of migration. Nevertheless, it should be emphasised that the very limited number of observations make the results are probably not very reliable. Also the model specification could probably be greatly improved. Especially the use of wage information on West-Germans could contribute to this aim.

Combining the two estimations revealed the estimated dependence of the latent variable driving migration intentions on the reported income can be explained rather well by the estimated relation between the anticipated wage gain and the measured wage. Assuming the latter estimation reflects a true underlying structure would make the classical 'Marshallian'(human capital) model to migration produce a predicted dependence of the latent variable on the measured income similar to the structure that is seen in the data.

As a good model is small and explains a lot of phenomena, the 'Marshallian' migration model seems to be the model of choice. However, it must be noted that the classic model is more simple mathematically, but theoretically less appealing. The real options theory is probably valid, at least to some extent, in the case of migration.

A possible cause of the fact we do not find proof of it in the GSOEP data, is that the real options approach is a bad description for migration decisions for German East-West migration, as the sunk costs are relatively low. This should be less of a problem with the second dataset.

Estimation with the data on ethnic Germans, revealed the estimated dependence of the latent variable on the *measured* income to be almost exactly as predicted by the real options theory, in the specification of Burda et al. As we argued the relation between the anticipated an measured income to be nonlinear , this can not be readily be interpreted as a proof of the real options approach of migration. Further research could reveal if the observed nonlinearity in the dependence of the migration propensity on the measured income is caused by the 'low-income'-effect of the anticipated income gain from migration we measured for the first dataset. As the real options theory should describe Russian-German migration well, given the large sunk costs, the nonlinearity in the influence of the anticipated income gain on the latent variable driving migration intentions should be pronounced.

The predicted shift in the trigger level was given relatively little attention, because of the theoretical difficulties related to measuring this variable from the data. Estimation was only done for the GSOEP dataset and gave no definitive conclusion on whether the predicted shift is present in the data or not.

The main conclusion of this thesis was already mentioned in the introduction: with the developed methods and datasets, we find only little support for the real options approach to migration. Perhaps a clearer (easier?) test could be developed in theoretical frameworks other than the one we used. Interesting approaches in this direction are Locher [29] and Parikh and Van Leuvensteijn [35], which consider the use of aggregated data on migration flows.

Appendix

Summary Statistics for the GSOEP dataset

Name		1991	1992	1993	1994	1995	1996	1997	1998	1999
	n	2618	2243	2017	1916	1915	1806	1704	1598	1572
	Min	1	1	1	1	1	1	1	1	1
Y	Max	4	4	4	4	4	4	4	4	4
	Avg	2.74	2.679	2.664	2.647	2.646	2.597	2.663	2.588	2.643
	Min	0	0	0	0	0	0	0	0	0
BIN	Max	1	1	1	1	1	1	1	1	1
	Avg	0.4107	0.3928	0.3626	0.3666	0.3123	0.3377	0.3419	0.3071	0.3594
	Min	0	0	0	0	0	0	0	0	0
MIG	Max	1	1	1	1	1	1	1	1	1
	Avg									
	no	26	51	79	100	105	112	111	106	126
	Min	0	0	0	0	0	0	0	0	0
SEX	Max	1	1	1	1	1	1	1	1	1
	Avg	0.4645	0.4575	0.4636	0.4571	0.4615	0.4687	0.4722	0.4736	0.4676
	Min	0	0	0	0	0	0	0	0	0
PARTNER	Max	1	1	1	1	1	1	1	1	1
	Avg	0.8782	0.889	0.8945	0.8964	0.8993	0.9049	0.899	0.8992	0.8944
	Min	0	0	0	0	0	0	0	0	0
OWNER	Max	1	1	1	1	1	1	1	1	1
	Avg	0.3584	0.3717	0.3792	0.4050	0.4292	0.4607	0.4813	0.4914	0.5248
	Min	0	0	0	0	0	0	0	0	0
FF	Max	1	1	1	1	1	1	1	1	1
	Avg	0.8595	0.8679	0.9307	0.859	0.948	0.9459	0.9516	0.9477	0.9513
	Min	0	0	0	0	0	0	0	0	0
ENV	Max	10	10	10	10	10	10	10	10	10
	Avg	6.143	6.251	6.4	6.455	6.576	6.574	6.498	6.646	6.701
	Min	0	0	0	0	0	0	0	0	0
UNI	Max	1	1	1	1	1	1	1	1	1
	Avg	0.2748	0.2872	0.2978	0.2993	0.3037	0.3081	0.3168	0.3149	0.3327
	Min	18	18	18	18	18	18	18	18	18
AGE	Max	65	65	65	65	65	65	65	65	65
	Avg	38.19	38.43	38.47	39.27	39.88	40.4	41.03	41.47	42.16
	Min	0	0	0	0	0	0	0	0	0
ATT	Max	1	1	1	1	1	1	1	1	1
	Avg	0.1656	0.1838	0.1713	0.1988	0.1933	0.1737	0.1694	0.1568	0.1724
	Min	0	0	0	0	0	0	0	0	0
KID	Max	1	1	1	1	1	1	1	1	1
	Avg	0.7699	0.7738	0.7652	0.7427	0.7249	0.7085	0.6731	0.6728	0.6506
	Min	0	0	0	0	0	0	0	0	0
REQTR	Max	1	1	1	1	1	1	1	1	1
	Avg	0.6771	0.6688	0.6896	0.6849	0.6694	0.6737	0.6936	0.7066	0.6947
	Min	0	0	0	0	0	0	0	0	0
SAME	Max	1	1	1	1	1	1	1	1	1
	Avg	0.08779	0.1038	0.1140	0.1184	0.1217	0.1253	0.1382	0.1470	0.1355

continued on next page...

...summary statistics continued

Name		1991	1992	1993	1994	1995	1996	1997	1998	1999
	Min	18	18	18	18	18	18	18	18	18
TREIMAN	Max	78	78	78	78	78	78	78	78	78
	Avg	40.02	40.37	40.37	38.52	40.17	38.58	40.7	41.73	39.26
	Min	0	0	0	0	0	0	0	0	0
PART	Max	1	1	1	1	1	1	1	1	1
	Avg	0.2233	0.1038	0.09895	0.1120	0.1169	0.1196	0.1131	0.1236	0.1317
	Min	0	0	0	0	0	0	0	0	0
BLUE	Max	1	1	1	1	1	1	1	1	1
	Avg	0.3393	0.3173	0.3029	0.3041	0.2833	0.3058	0.2783	0.2713	0.2774
	Min	1	1	1	1	1	1	1	1	1
QUALW	Max	10	10	10	10	10	10	10	10	10
	Avg	7.993	7.485	7.368	7.104	7.119	7.13	7.157	7.17	7.118
	Min	0	0	0	0	0	0	0	0	0
JEOP	Max	1	1	1	1	1	1	1	1	1
	Avg	0.4454	0.3633	0.2140	0.2062	0.164	0.1663	0.1212	0.1601	0.1330
	Min	0	0	0	0	0	0	0	0	0
ECO	Max	1	1	1	1	1	1	1	1	1
	Avg	0.7267	0.5816	0.6396	0.6289	0.4733	0.6864	0.7786	0.7704	0.7708
	Min	1	1	1	1	1	1	1	1	1
LOSS	Max	4	4	4	4	4	4	4	4	4
	Avg	2.453	2.228	2.843	2.827	2.352	2.929	2.58	2.904	2.8271
	Min	200	200	200	200	200	200	200	200	200
AINC	Max	7000	7000	7000	7000	7000	7000	7000	7000	7000
	Avg	1550	2085	2533	2826	2965	3124	3235	3265	3462
	Min	0	0	0	0	0	0	0	0	0
INC	Max	7000	7000	7000	7000	7000	7000	7000	7000	7000
	Avg	2446	2988	3524	3774	3906	4040	4096	4131	4292
	Min	10	10	10	10	10	10	10	10	10
HOURS	Max	990	990	990	990	990	990	990	990	990
	Avg	409.6	432.2	435.6	433.4	432.3	429.9	437.1	436.8	424.5
	Min	0	0	0	0	0	0	0	0	0
SELF	Max	1	1	1	1	1	1	1	1	1
	Avg	0.040	0.046	0.0517	0.0520	0.0587	0.0643	0.0672	0.0696	0.0782
	Min	1	1	1	1	1	1	1	1	1
WHH	Max	6	6	6	6	6	6	6	6	6
	Avg	1.800	1.876	1.863	1.937	1.866	1.863	1.961	1.922	1.945
	Min	0	0	0	0	0	0	0	0	0
BIL	Max	1	1	1	1	1	1	1	1	1
	Avg	0.7153	0.738	0.7644	0.7675	0.776	0.7859	0.8006	0.8055	0.8136
	Min	1	1	1	1	1	1	1	1	1
SATJ	Max	10	10	10	10	10	10	10	10	10
	Avg	6.692	6.896	6.933	6.846	6.77	6.8	6.799	6.814	6.815
	Min	1	1	1	1	1	1	1	1	1
SATD	Max	10	10	10	10	10	10	10	10	10
	Avg	6.92	6.394	6.543	6.681	6.796	6.933	6.96	7.175	7.363
	Min	1	1	1	1	1	1	1	1	1
SATI	Max	10	10	10	10	10	10	10	10	10
	Avg	5.128	5.268	5.768	5.784	5.743	5.857	5.684	5.686	5.859

Table 5: Summary statistics for the GSOEP dataset

Variable name	Description
Y	Migraton intentions
BIN	High migration intention
MIG	Migrated
SEX	Female
PARTNER	Married or living together
OWNER	Owner of the house
FF	Has family or friends in West-Germany
ENV	Environmental satisfaction
UNI	University degree
AGE	Respondent age
ATT	Does not feel strongly connected to local community
KID	Has children
REQTR	Current job requires specific training
SAME	Has had the current job for more then 15 years
TREIMAN	Treiman job prestige index
PART	Part time job
BLUE	Blue collar worker
QUALW	Estimation of quality of life in West-Germany
JEOP	Fear of job loss in foreseeable future or unemployed
ECO	Envirionmental satisfaction
LOSS	Fear of job loss in foreseeable future
AINC	(personal) Labour income
INC	Household income
HOURS	Average hours worked per week
SELF	Self-employed
WHH	White-collar worker, managerial
BIL	Completed secondary school
SATJ	Satisfaction with current job
SATD	Satisfaction with current house
SATI	Satisfaction with current income

Table 6: Variable description for the GSOEP dataset

The reported summary statistics are with respect to a censored subsample $200 < AINC < 7000$, containing only nonmigrants. For the MIG variable, the statistics are with respect to a sample containing migrants and non-migrants.

All ordered variables are ordered from low to high.

Density estimation of the reported personal income

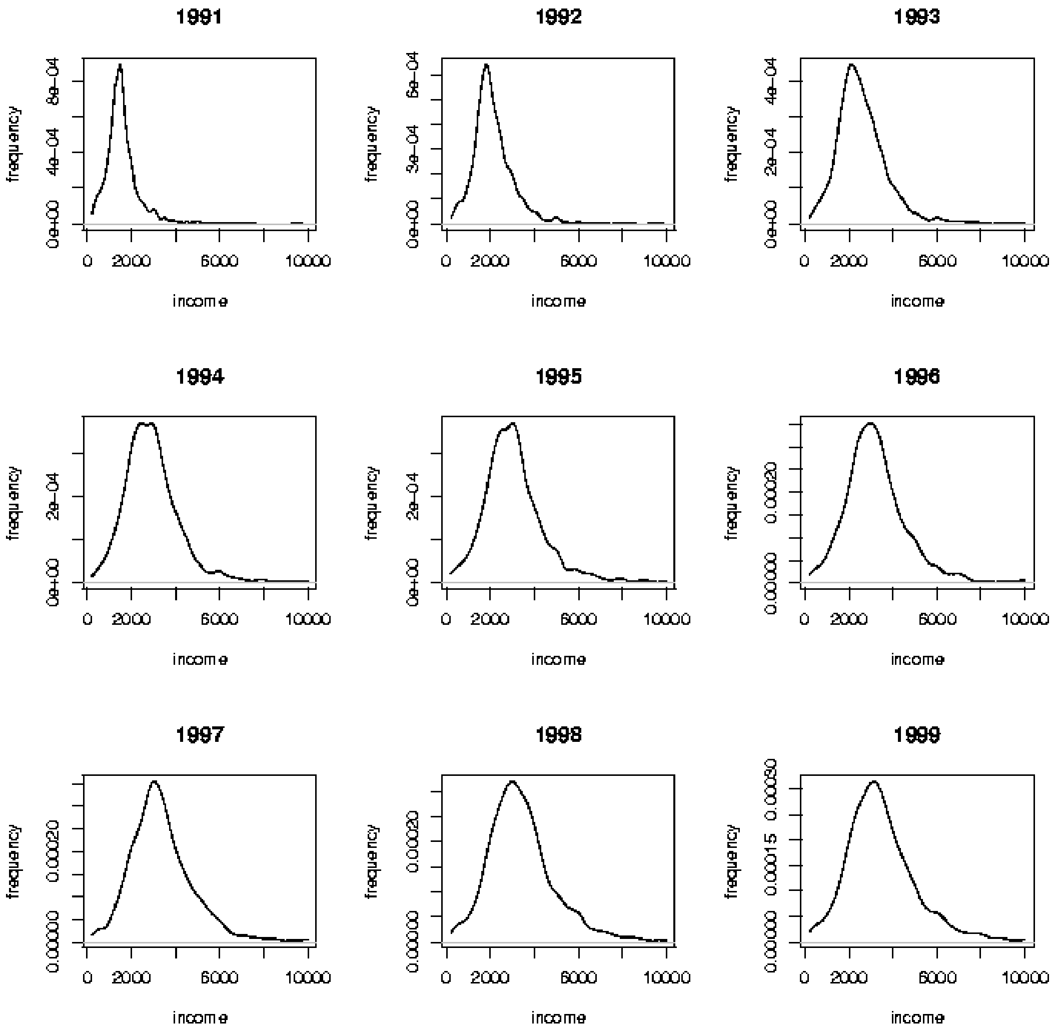


Figure 17: Density estimations of the income distribution

The marginal influence of income: binary response

GAM: $E[BIN = 1] = G(m(AINC))$

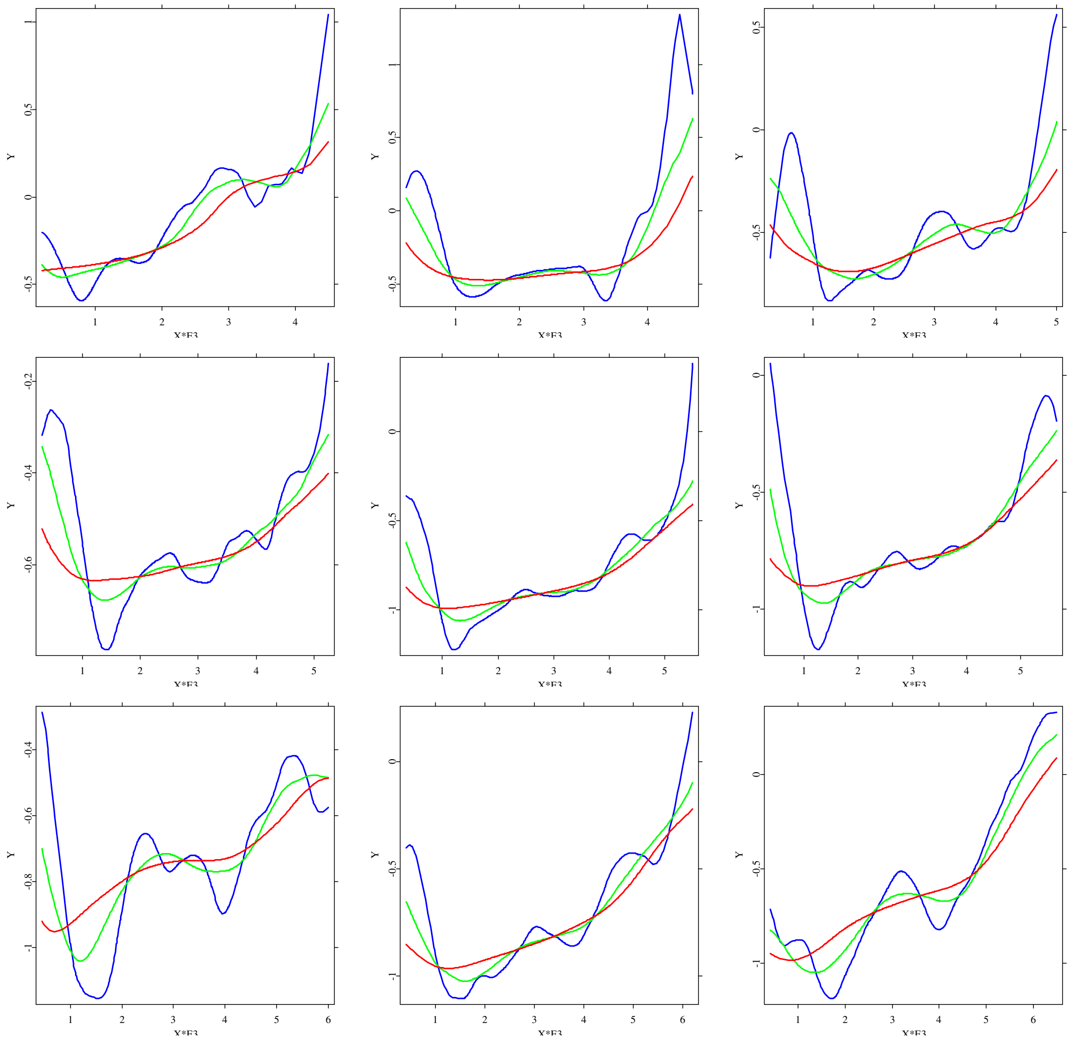


Figure 18: GAM estimated influence of income (AINC) on U^*

A small GPLM model: $E[BIN = 1] = G(SEX + PARTNER + OWNER + FF + JEOP + ENV + UNI + AGE + m(AINC))$

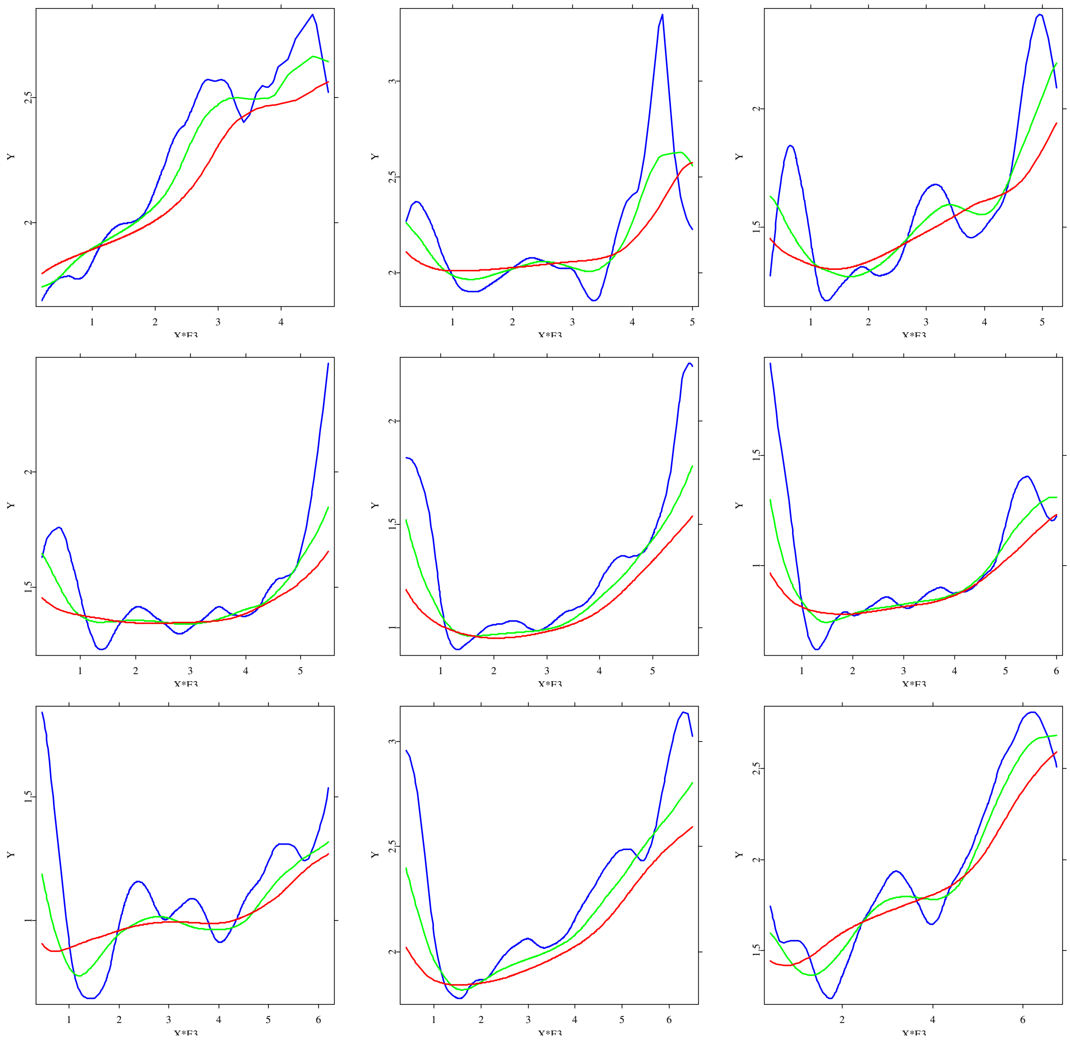


Figure 19: Small GPLM estimates

A large GPLM model: $E[BIN=1]=G(SEX+PARTNER+OWNER+FF+JEOP+ENV+UNI+AGE+ATT+QUALW+KID+SAME+WHH+SATI+SATD+SATJ+REQTR+m(AINC))$

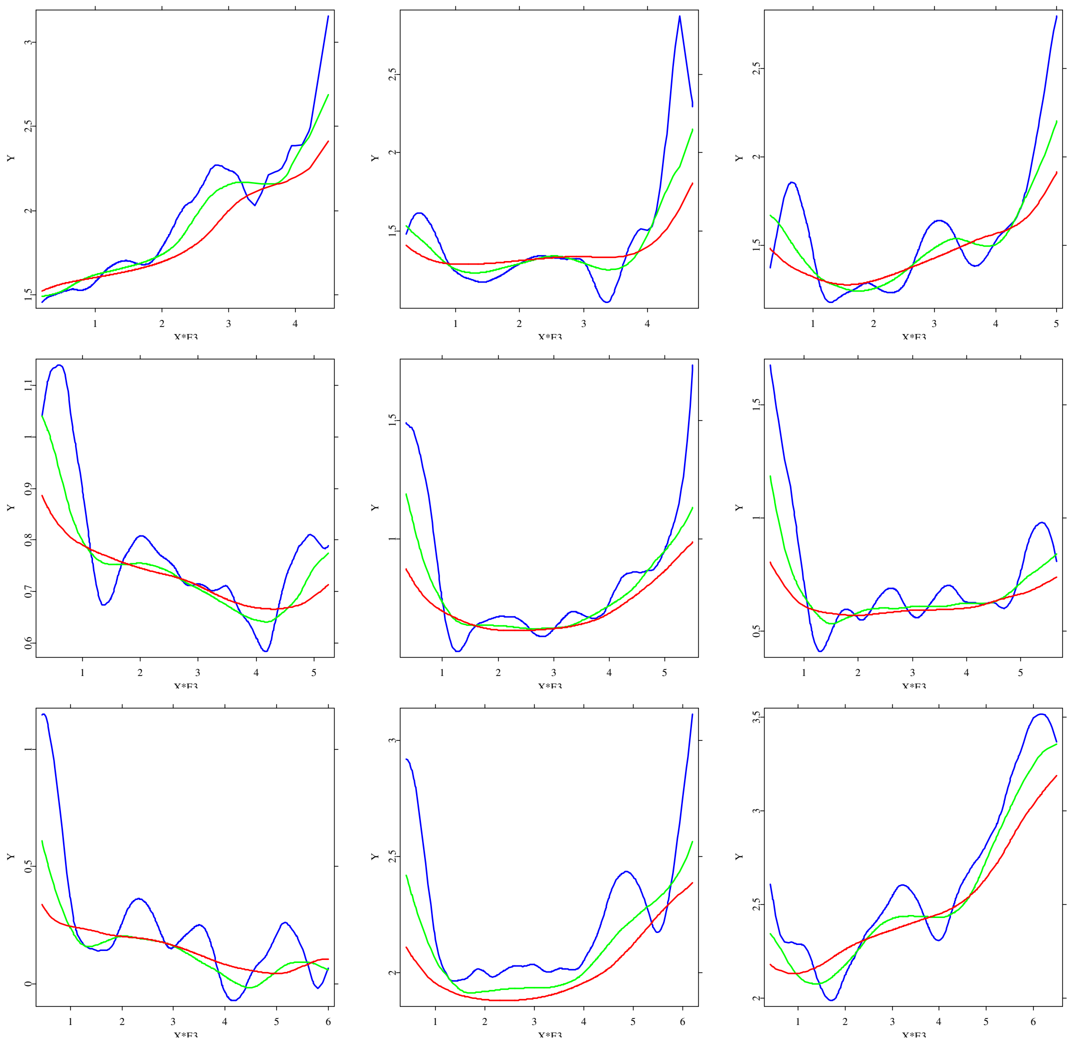


Figure 20: Large GPLM estimates

Ordered Response

Kernel regression: $E[Y] = m(AINC)$

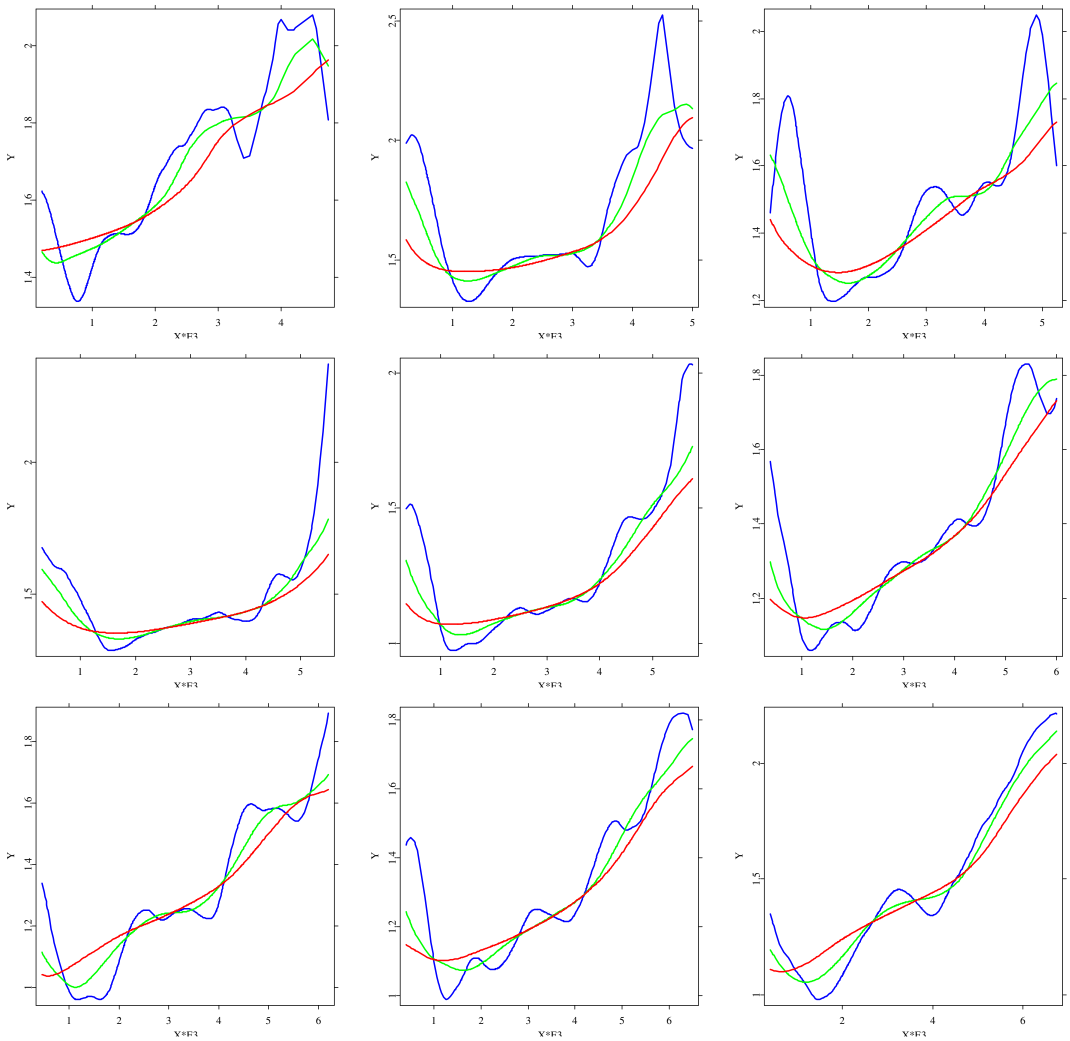


Figure 21: Kernel regression estimates

A small PLM: $E[Y] = SEX + PARTNER + OWNER + FF + JEOP + ENV + UNI + AGE + m(AINC)$

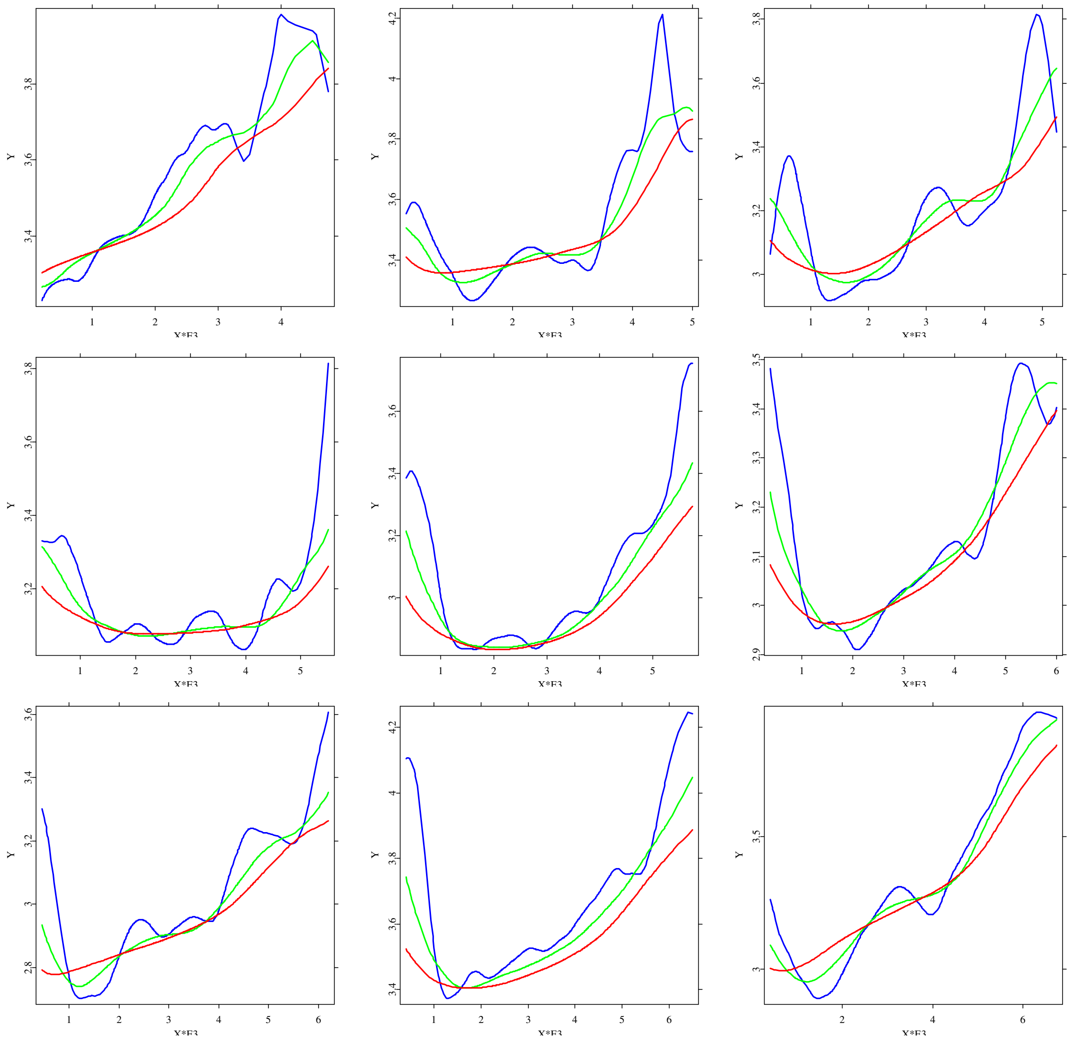


Figure 22: PLM estimates

A large PLM: $E[Y] = SEX + PARTNER + OWNER + FF + JEOP + ENV + UNI + AGE + ATT + QUALW + KID + SAME + WHH + SATI + SATD + SATJ + REQTR + m(AINC)$

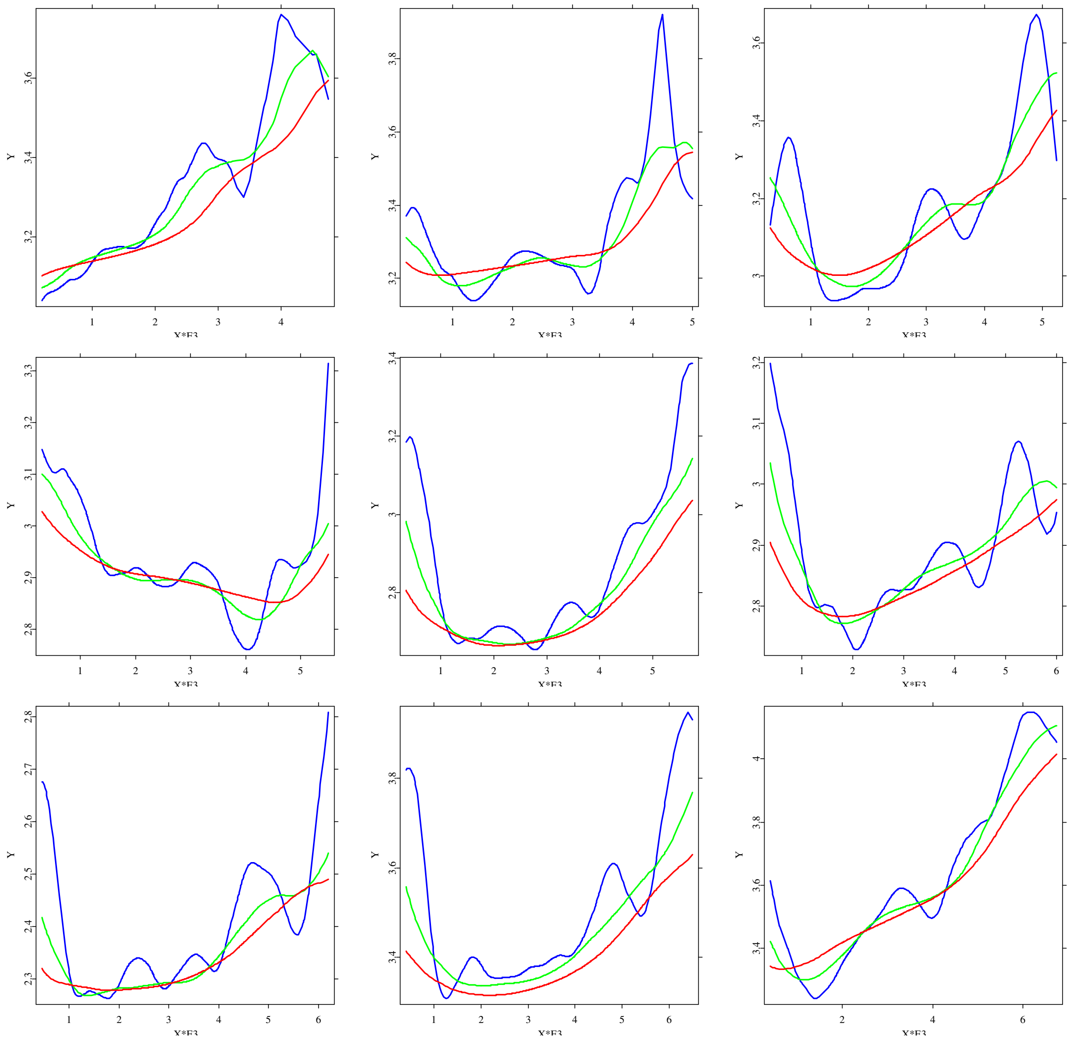


Figure 23: Large PLM estimates

Contour Plots of the estimated $W^W(W^E)$

Heckman self-selection-corrected W^W estimation, plotted against the observed W^E

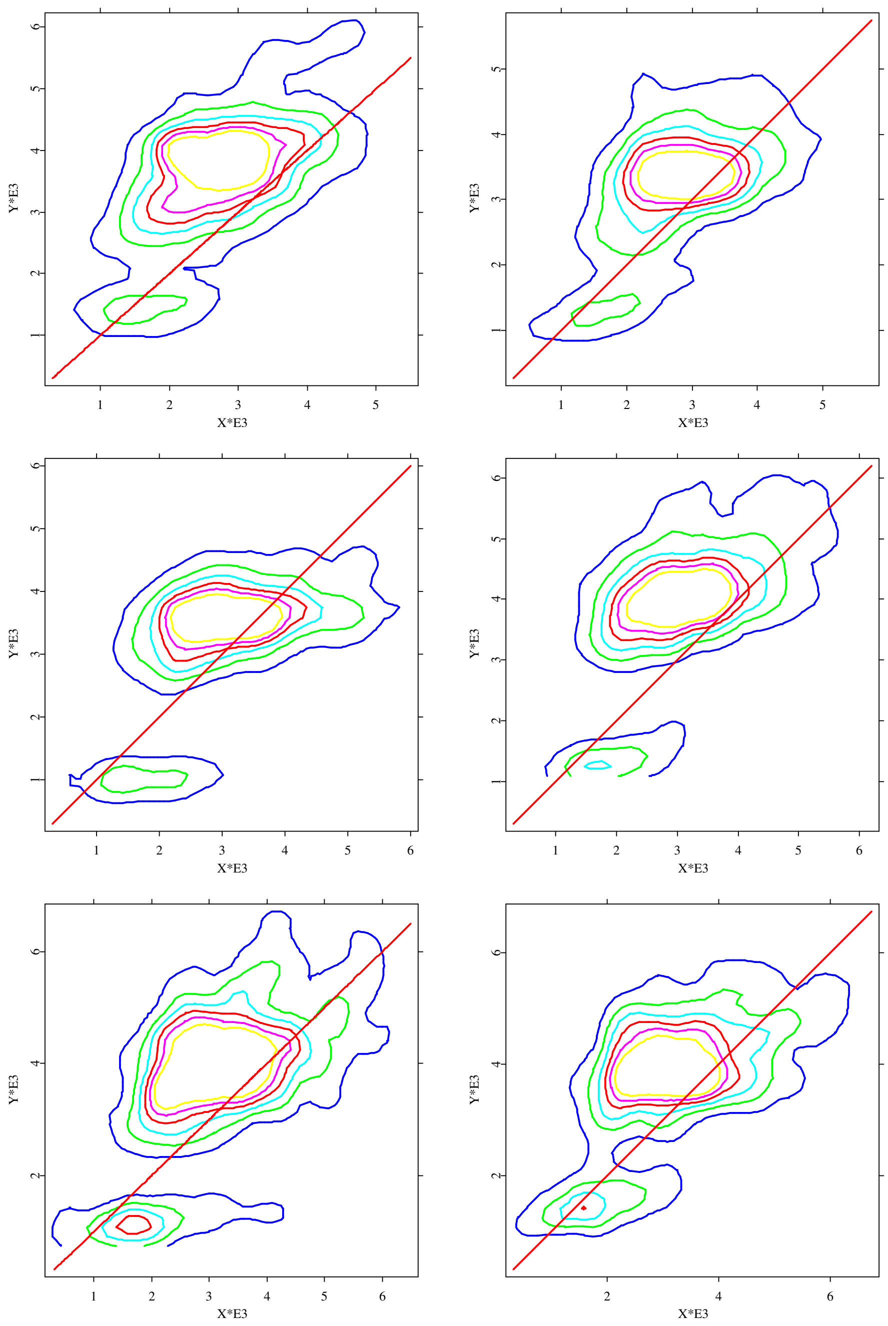


Figure 24: Expected west-wages, plotted against observed wages

Heckman corrected W^W estimation, plotted against observed W^E, for $AINC \leq 2700$

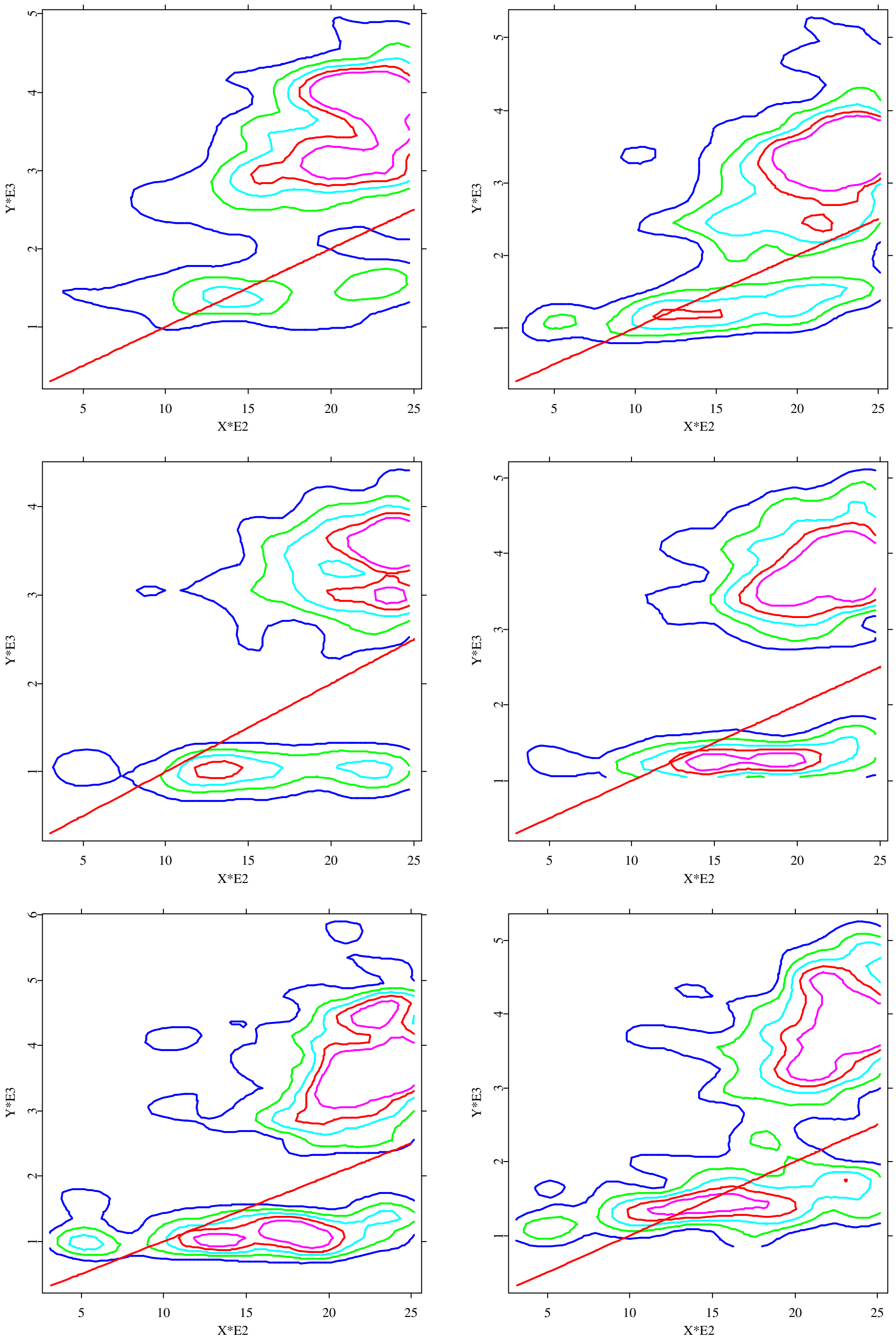


Figure 25: Expected west-wages, plotted against observed wages, closeup

Heckman corrected W^W estimation, plotted against estimated W^E

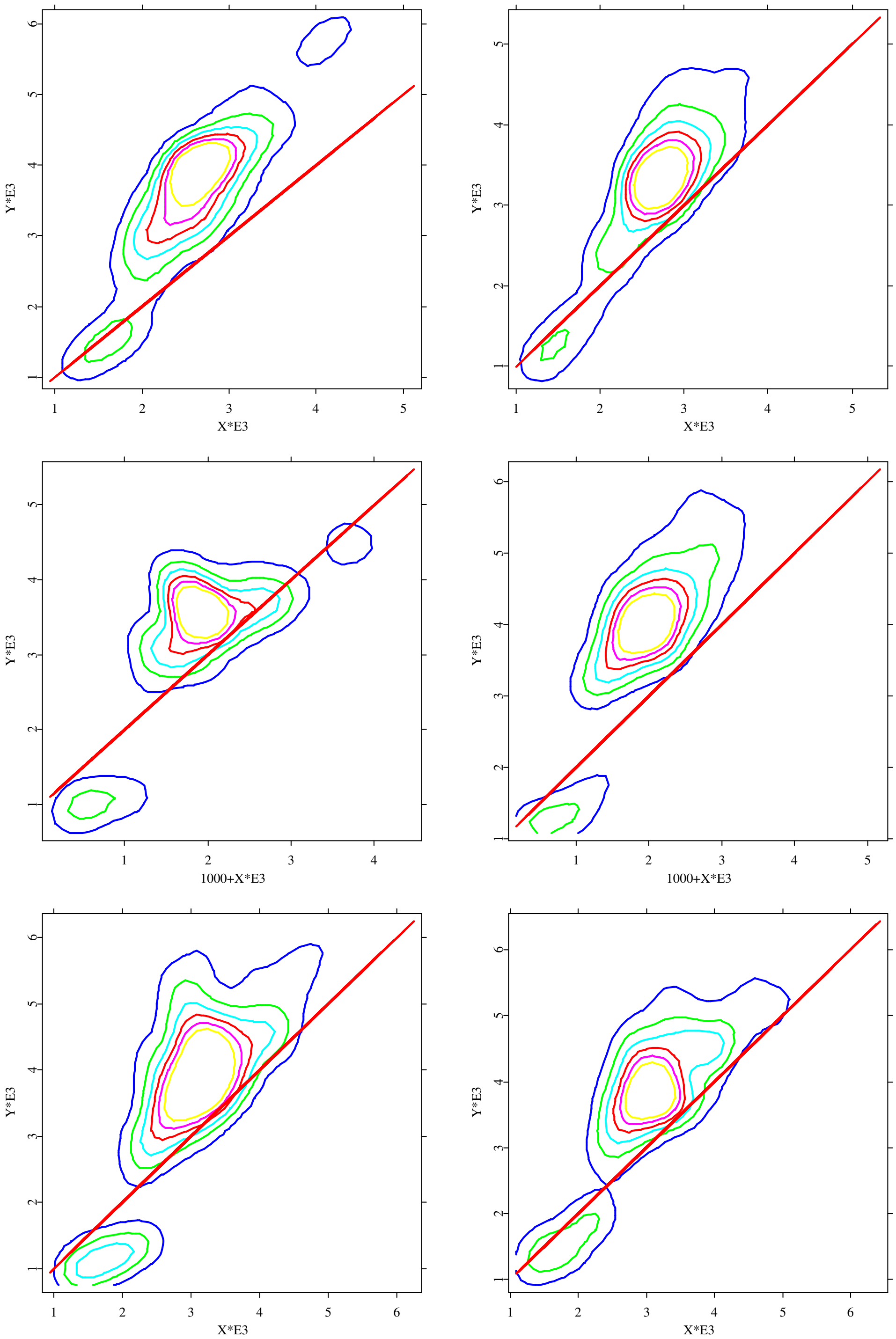


Figure 26: Expected west-wages, plotted against estimated wages

Heckman corrected W^W estimation, plotted against estimated W^E, for $AINC \leq 2700$

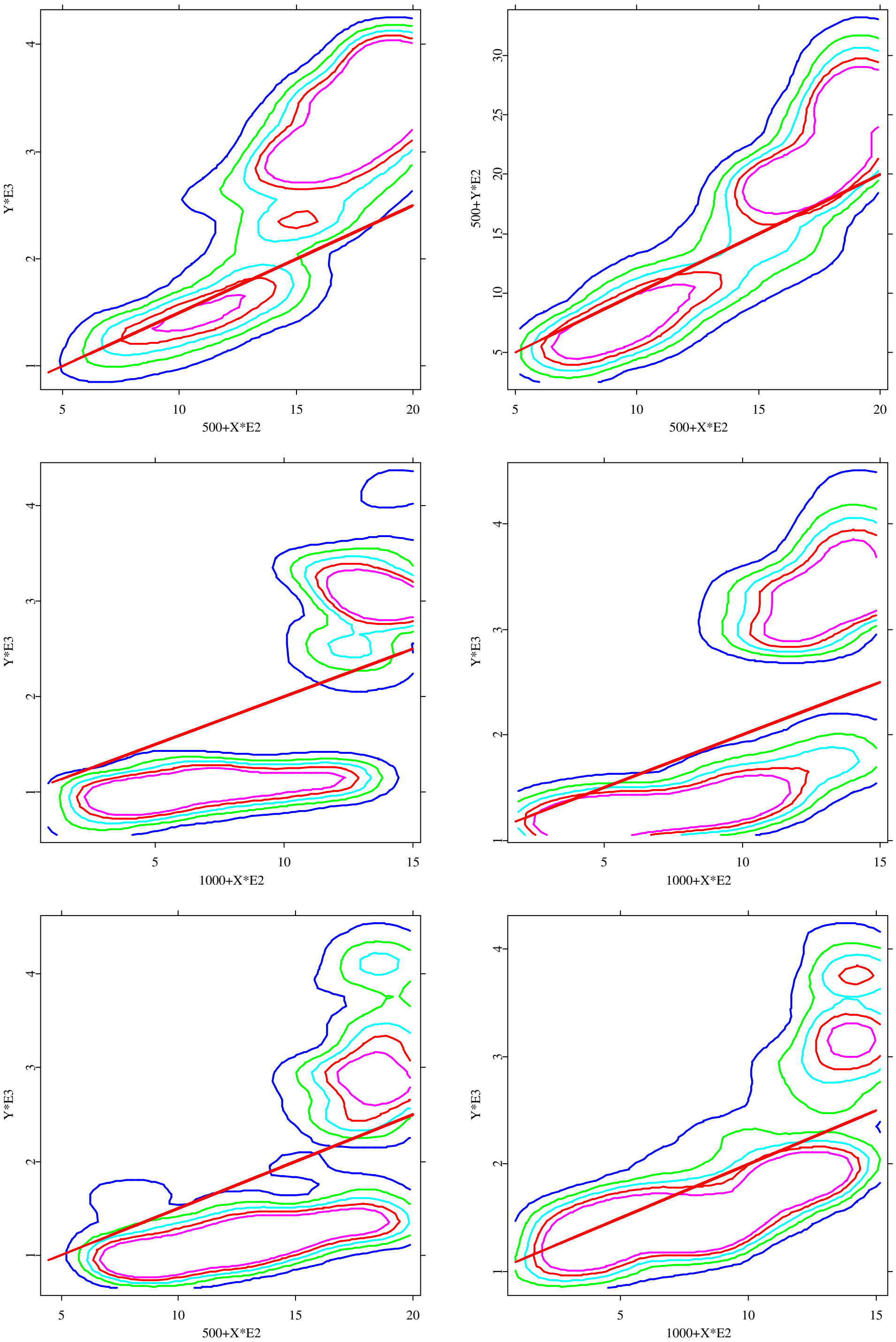


Figure 27: Expected west-wages, plotted against estimated wages, closeup

Not self-selection-corrected estimated W^W, plotted against observed W^E

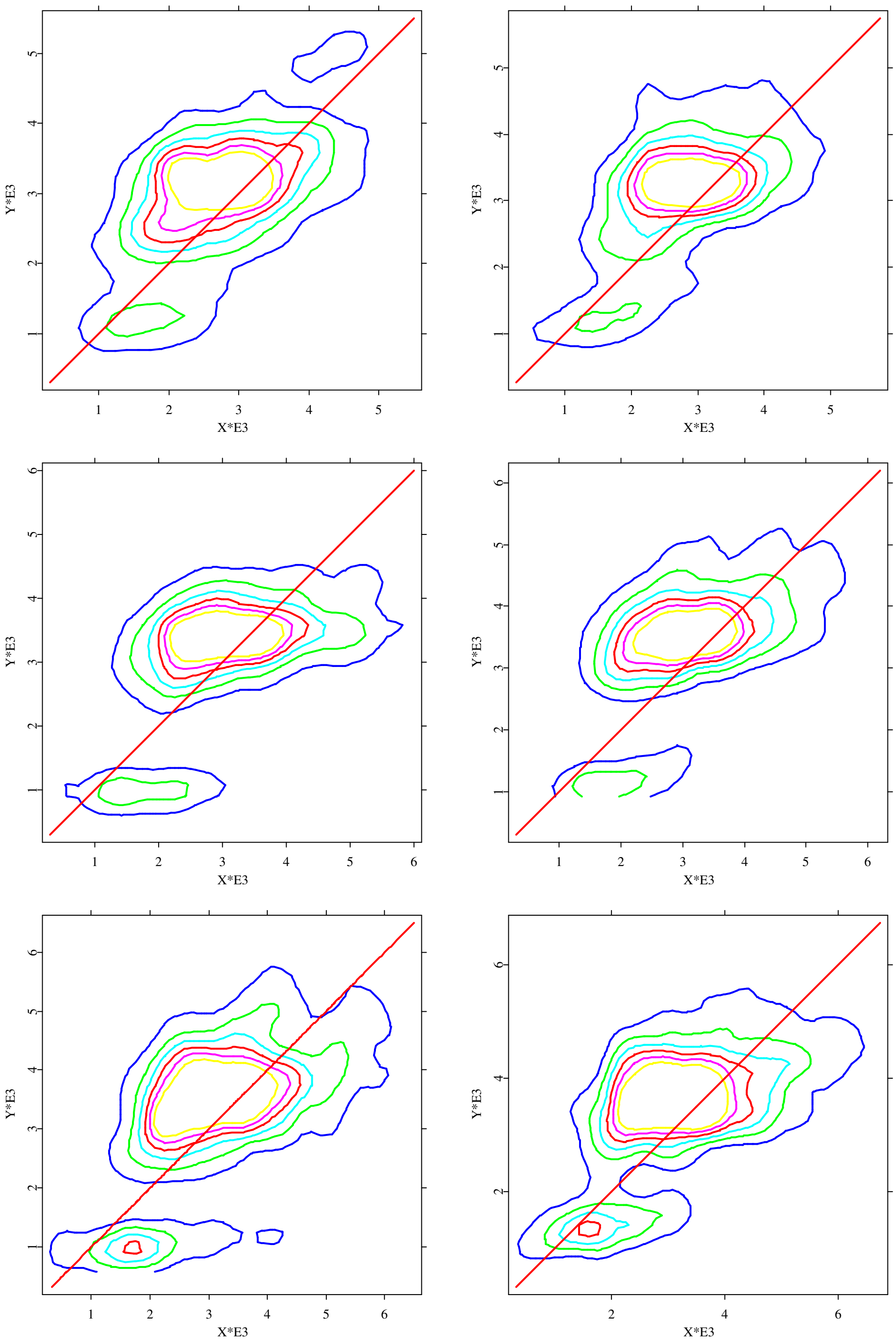


Figure 28: Anticipated west-wages, plotted against observed wages

Not self-selection-corrected W^W estimation, plotted against observed W^E, for $AINC \leq 2700$

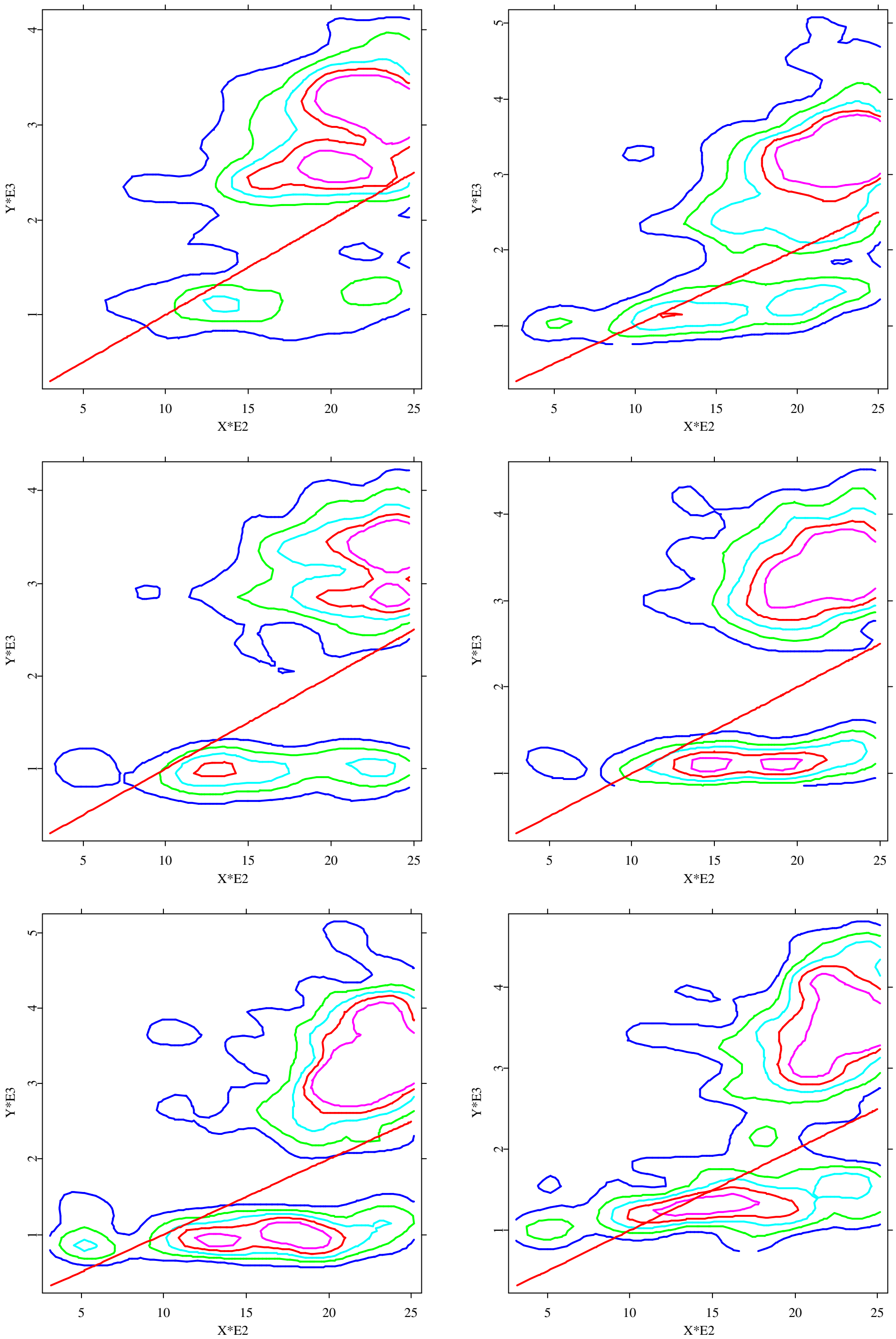


Figure 29: Anticipated west-wages, plotted against observed wages, closeup

Not self-selection-corrected estimated W^W, plotted against estimated W^E

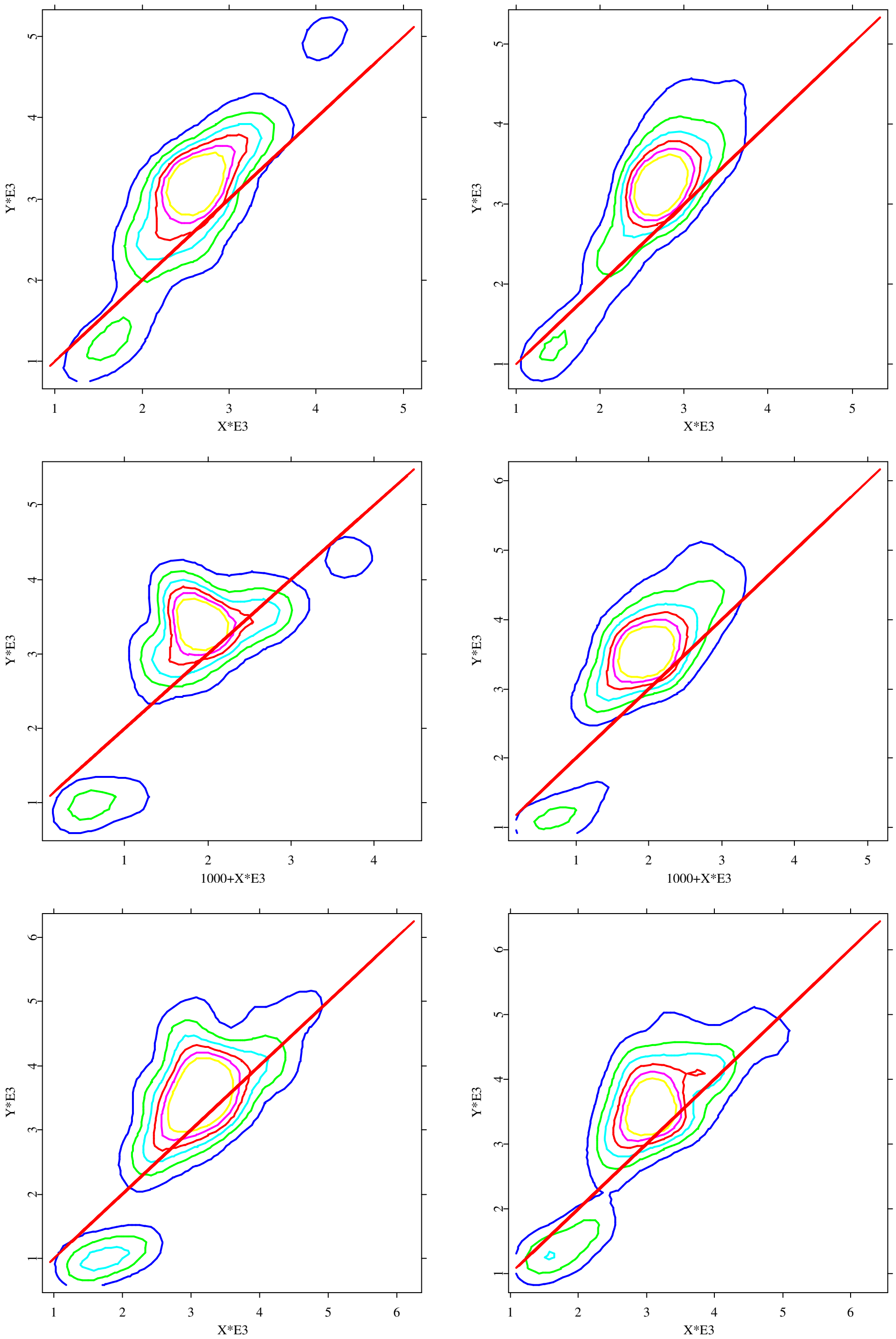


Figure 30: Anticipated west-wages, plotted against estimated wages

Not self-selection-corrected W^W estimation, plotted against estimated W^E, for $AINC \leq 2700$

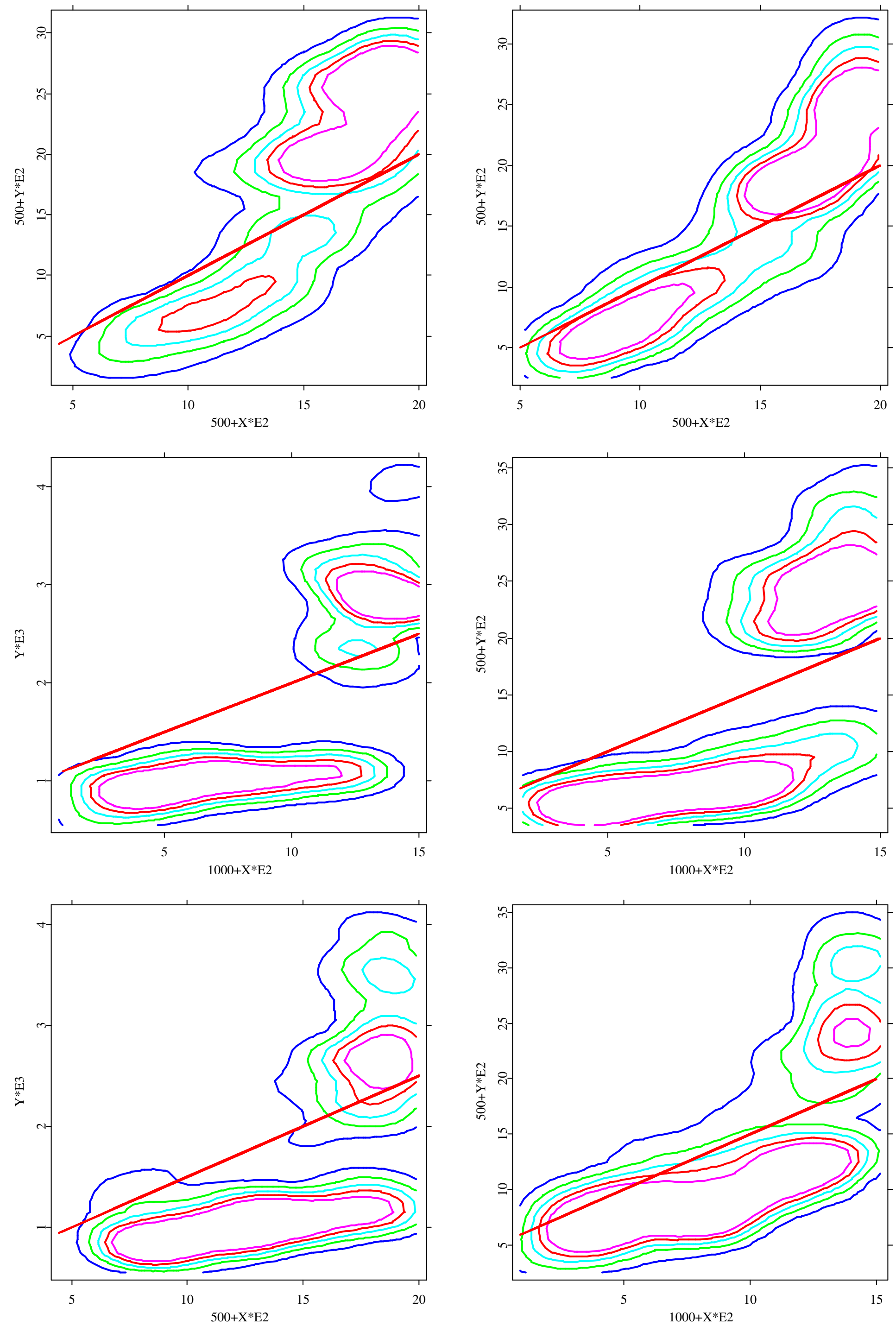


Figure 31: Anticipated west-wages, plotted against estimated wages, closeup

References

[1] Takeshi Amemiya. *Advanced Econometrics*. Harvard University Press, Cambridge, Massachusetts, 1985.

[2] Donald W.K. Andrews and Marcia A. Schafgans. Semiparametric estimation of a sample selection model. *Cowles Foundation Discussion Papers, Cowles Foundation, Yale University*, (1119), April 1996.

[3] Joshua Angrist. Estimating the labor market impact of voluntary military service using social security data on military applicants. *Econometrica*, March 1998.

[4] George J. Borjas. Self-selection and the earnings of immigrants. *The American Economic Review*, 77(4):531–553, September 1987.

[5] Michael C. Burda. The determinants of east-west migration: Some first results. *Center for Economic Policy Research Discussion Paper*, 764, January 1993.

[6] Michael C. Burda. Migration and the option value of waiting. *Center of Economic Policy Research Working Papers*, (1229), 1995.

[7] Michael C. Burda, Wolfgang Härdle, Marlene Müller, and Axel Werwatz. Semiparametric Analysis o fGerman East-West Migration Intentions: facts and Theory. *Journal of Applied Econometrics*, 13(5), 1998.

[8] Michael C. Burda and Christoph M. Schmidt. Getting behind the east-west wage differential: Theory and evidence. *Working paper of the SFB373. Humboldt Universität Berlin*, (77), 1997.

[9] Julie Davanzo. Microeconomic approaches to studying migration decisions. In Gordon F. De Jong and Robert W. Gardner, editors, *Migration Decision Making*, pages 90–129. Pergamon Policy Studies, 1981.

[10] Julie Davanzo and James R. Hosek. Does migration increase wage rates? an analysis of alternative techniques for measuring wage gains to migration. *A Rand Note*, N-1582-NICHD, July 1981.

[11] A. Dixit and R. Pindyck. *Investment under Uncertainty*. Princeton: Princeton University Press, 1994.

[12] Avinash Dixit. Investment and Hysteresis. *The Journal of Economic Perspectives*, 6(1), 1992.

[13] Avinash Dixit. Entry and exit decisions under uncertainty. *Journal of Political Economy*, 3:620–638, 1997.

[14] Thomas A. Dunn, Michaela Kreyenfeld, and Mary E. Lovely. Communist human capital in a capitalist labor market : The experience of east german and ethnic german immigrants to west germany. *Vierteljahrshefte zur Wirtschaftsforschung*, 66(1), 1997.

[15] Peter A. Fisher. *On the Economics of Immobility*, chapter 4. Paul Haupt Verlag, Bern, 1999. Dissertation der Wirtschafts- und Sozialwissenschaften der Universitt der Bundeswehr Hamburg.

[16] The R Foundation for Statistical Computing. The r project for statistical computing. Available from the www at http://www.r-project.org/.

[17] Bildungsverein fr Volkskunde in Deutschland DIE LINDE e. V. Geschichte der russlanddeutschen. available from http://www.russlanddeutschegeschichte.de/.

[18] J Franke, Härdle W., and C. Haffner. *Einführung in die Statistik der Finanzmärkte*. Springer Verlag, 2001.

[19] John P. Haisken-DeNew and Joachim R. Frick. Dektop companion to the german socio-economic panel study(gsoep). Available from http://www.diw.de/deutsch/sop/service/dtc/index.html.

[20] W. Härdle, E. Mammen, and M. Müller. Testing parametric versus semiparametric modelling in generalized linear models. *Journal of the American Statistical Association*, 93:1461–1474, 1998.

[21] W. Härdle, M. Müller, S. Sperlich, and A. Werwatz. Non - and Semiparametric Modelling. http://www.xplore-stat.de/ebooks/ebooks.html, 2001.

[22] John R. Harris and Michael P. Todaro. Migration, unemployment and development: A two-sector analysis. *The American Economic Review*, 60: 126–142, 1970.

[23] J. J. Heckman. Sample selection as a spcification error. *Econometrica*, 47: 153–161, 1979.

[24] Joel L. Horowitz and Wolfgang Härdle. Testing a parametric model against a semiparametric alternative. *Econometric Theory*, 10:821–848, 1994.

[25] George R Judge, R. Carter Hill, William E. Griffith, Helmut Lütkepohl, and Tsoung-Chao Lee. *Introduction to the Theory and Practice of Econometrics*. John Wiley and Sons, New York, 1988.

[26] Yasmeen Khwaja. Should i stay or should i go? migration under uncertainty: A new approach. *Working Paper Series of the School of Oriental and African Studies, Department of Economics, University of London.*, (113), 2000.

[27] L. F. Lee. Some approaches to the correction of selectivity bias. *Review of Economic Studies*, 49:355–372, 1982.

[28] Kung-Yee Liang and Scott L. Zeger. Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, 73(1):13–21, 1986.

[29] Lilo Locher. Testing for the option value of waiting. *Discussion Paper of the Institute for the Study of Labour(IZA), Bonn*, (405), 2001.

[30] Ilya Malyakin. The 'volga german autonomour republic': its absence as a detail of the political landscape. available from http://russia.jamestown.org/pubs/view/pri_004_010_006.htm.

[31] McDonald, R. and D. R. Siegel. The Value of Waiting to Invest. *Quarterly Journal of Economics*, 101(4):707–727, 1986.

[32] Marlene Müller. Computer-assisted generalized partial linear models. In D. W. Scott, editor, *Proceedings of the 29th Symposium on the Interface, Houston Texas, May 14-17*, volume 29, pages 221–230. 1997.

[33] Robert A. Nakosteen and Michael Zimmer. Migration and incme: The question of self-selection. *Southern Economic Journal*, 46:840–851, 1980.

[34] Paul O'Connell. Migration under uncertainty: 'try your luck' or 'wait and see'. *Journal of Regional Science*, 47, 1997.

[35] Ashok Parikh and Michiel Van Leuvensteijn. Interal migration in regions of germany: A panel data analysis. *University of East Anglia, Mimeo*, 2001.

[36] R. S. Pindyck. Irreversible Investment, Capacity Choice, and the Value of the Firm. *American Economic Review*, 77(5):969–985, 1988.

[37] R. S. Pindyck. Irreversibility, Uncertainy and Investment. *Journal of Economic Literature*, 29:1110–1148, 1991.

[38] J.L. Powell. Semiparametric estimation of bivariate latent variable models. *University of Wisconsin-Madison, Working Paper*, 8704.

[39] E. Ravenstein. The laws of migration. *Journal of the Statistical Society*, 59: 214–301, 1889.

[40] Brian Ripley. The gee package. Available from the www at http://cran.r-project.org/src/contrib/PACKAGES.html#gee.

[41] A. D. Roy. Some thoughts on the distribution of earnings. *Oxford Economic Papers*, pages 135–146, 3.

[42] Rubin. Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics*, 2:1–26, 1977.

[43] Johannes Schwarze. Beinflusst das Lohngefälle zwischen Ost- und Westdeutschland das Migrationsverhalten der Ostdeutschen? -Eine Mikroökonometrische Untersuchung mit Daten des Sozio-Oekonomischen Panels 1991 bis 1994. *Allgemeines Statistisches Archiv*, 80(1), 1996.

[44] Thomas A. Severini and J.G. Staniswalis. Quasi-likelihood estimation in semiparametric models. *Journal of the American Statistical Association*, 89:501–511, 1994.

[45] Thomas A. Severini and W.H. Wong. Profile likelihood and conditionally parametric models. *Annals of Statistics*, 20:1768–1802, 1992.

[46] Insan Tunali. Rationality of migration. *International Economic Review*, 41 (4):893–920, 2000.

[47] XploRe. XploRe - the interactive statistical computing environment. http://www.xplore-stat.de, 2002.