

# Labour Market Integration of Immigrants and their Children in Germany: Does Personality Matter?

Anna-Elisabeth Thum  
European University Institute

October 11, 2010

## Abstract

Educational attainment, length of stay, differences in national background and language skills play an acknowledged important role for the integration of immigrants. But integration is also a social process, which suggests that psychological factors are relevant. This paper explores whether and to what extent immigrants and their children need to believe in their ability to control their own success, in other words their sense of control. To quantify this personal trait I use a measure of an individual's sense of control over outcomes in life - such as finding a job. This measure is known in psychology as "the locus of control". I first estimate an exogenous measure. Then I address the problem that this measure is actually endogeneous in a labor market outcome equation by employing a model in which the sense of control is an endogenized latent factor in a simultaneous equation model. The determinants of this sense of control as well as its effect on the probability of being employed are examined. The model is estimated using an implemented Bayesian Markov Chain Monte Carlo algorithm. Results with endogenized personality indicate that, on average, immigrants believe less than natives in being able to control outcomes in life, but children of immigrants have already a stronger sense of control than their parents. The paper also finds that sense of control over life's outcomes positively contributes to the probability of being employed. This means that immigrants and their children face a double disadvantage on the labor market: they are disadvantaged because of their status as an immigrant and they have a lower sense of being able to control their situation, which is a personality trait that matters on the labour market.

## 1 Introduction

Does personality matter for integration of immigrants into the labor market? Intuition and common sense suggest a positive answer, but scientifically it is not as

straightforward to tell, especially for the economist. This paper does not invent a new immigrant "homo oeconomicus" with a mathematically modelled personality, feelings, culture and other human features. Instead, it suggests a starting point to include an aspect of personality, a noncognitive skill, into an econometric model of labor market integration for immigrants. Measures of noncognitive skills and their inclusion into economic models have recently been studied in the economic literature - theoretically as well as empirically. This body of literature is still small but since it follows the trend of derationalizing the "homo oeconomicus", it is promising. Studies of the role of noncognitive skills have been undertaken for natives and there is a small literature on noncognitive skills for immigrants. The aim of this paper is to enrich this small body of literature by using an elaborated statistical tool to address problems of measurement error and endogeneity, which are very evident when working with measures of psychological concepts, such as noncognitive skills. This tool allows to endogenize measurement of noncognitive skills by taking into account its determinants. It additionally allows construction of a measure out of an informative set of measures of the noncognitive skill. This is an advantage since psychological concepts are more complex to measure than a naturally quantitative variable such as age or years of schooling.

The personality aspect, or noncognitive skill, we consider is called in psychological terms "locus of control", developed by Rotter (1966). We follow Heckman, Stixrud and Urzua (2006) in this approach. The locus of control is a measure for the degree to which an individual believes he (or she - it matters for both genders) has control over the happenings in his life<sup>1</sup>. It is represented as a scale reaching from "external" to "internal". A high *external* locus of control indicates that the individual believes that his life is controlled by forces outside of his own influence and he or she does not have a high feeling of controlling his life. A high *internal* locus of control however indicates that the individual believes strongly in his ability to control his life. One hypothesis of this paper is that an immigrant, who believes in controlling his outcomes has a higher incentive to provide the effort to integrate. The locus of control is strongly linked to the concept of "motivation": if individuals feel they have control over their lives, they believe in a causality between their actions and outcomes and this will motivate them to take actions.

---

<sup>1</sup>Osborne-Groves (2006) has already used this measure. See Heckman, Stixrud and Urzua (2006).

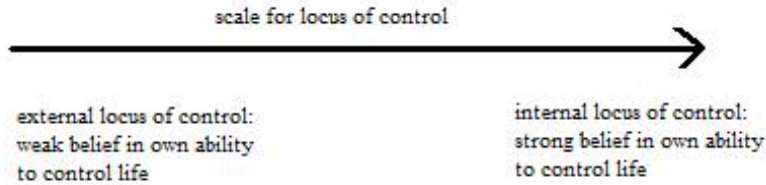


Figure 1: Measuring Personality - the Locus of Control

The locus of control of an individual develops over time, with foundations being laid in childhood and adolescence. Education, family background, maybe religion and also their personal immigration history - such as the arrival date in the host country - can play a role to determine an individual's locus of control. This is why we use a statistical tool, which allows to construct a measure of the locus of control, that depends on its determinants. In this way we address endogeneity problems of some variables, which could be included as controls in the employment equation and be correlated to the measure of the locus of control. This tool is based on work by Fahrmeir and Raach (2006).

This paper finds that a more *internal* locus of control has a positive effect on the probability of being employed. We find that being an immigrant has a significantly negative effect on having a more *internal* locus of control. Immigrants have on average a more *external* index of control. The same is true for the second generation, but the effect is not significant.

The paper is structured as follows. Section 2 gives an overview over the existing literature on labor market integration of immigrants and their children, and in particular in Germany, and over the use of psychological concepts in economics and in the literature on labor market integration. In section 3 I introduce the data I use and describe the main variable definitions of the model. Section 4 presents the model. In section 5 I analyze the empirical results and section 6 contains conclusions.

## 2 Labor Market Integration of Immigrants and Their Children

The theoretical and empirical study of economic integration of immigrants was initiated by a seminal paper by Chiswick (1978), in which he shows that a catch-

up process of “assimilation” of immigrants’ earnings to those of the indigenous population takes place. This assimilation depends positively and crucially on the time spent in the host country. Among the first, Borjas (1985) shows that assimilation depends not only on the duration of stay but differs across cohorts. Borjas (1987) shows, that this cohort effect can be due to a change in the mix of countries-of-origin among the immigrant population.

A debate on how to study integration of immigrants economically commenced and a search for the main determinants of labour market integration began among mainly labour economists. A body of literature is concerned about different measures of integration, such as LaLonde and Topel (1992), Baker and Dwayne (1994) and Eckstein and Weiss (2002). Skills are considered as main factor of labour market integration, such as language acquisition in assimilation. A prominent example is work by Chiswick and Miller (1999). Human capital investment and transferability of skills as an important factor is, for example, studied by Duleep and Regrets (1999). Cultural Factors are studied by a smaller body of economic literature, but ethnic differences in immigrants’ performance are acknowledged, for example by Chiswick (1988).

This literature mainly covers the United States and Australia. There is a recently developing body considering the German case, acknowledging the difference in labour market structure and immigration history with respect to the US. In particular, Dustmann and Schmitt (2000) study for example wage performance of immigrant women in Germany. Dustmann (1993) shows, that the status as a temporary vs permanent migrant has an effect on earnings adjustment. Constant, Zimmermann and Zimmermann (2006) examine the role ethnic self-identification. A comparative study of integration in Germany and Denmark by Hinte and Zimmermann (2005) provides a valuable overview over findings in the field.

Main findings on determinants of labour market integration in Germany are that formal education is particularly important since on the German labour market a high importance is attached to formal educational qualifications. Fertig and Schmidt (2001) argue that in the European context, discrimination might play a greater role for immigrants’ earnings than in the US case.

In this research, in line with a body of research on the role played by personality as a determinant for earnings, I propose to investigate specifically on the importance of personality in form of non-cognitive traits for the labour market performance of immigrants.

## 2.1 Integration in Germany

In July 2006 the German Integration summit was held and was supposed to be the starting point for a national plan of an integration process. Preceding the summit was the ratification of a new integration policy, the Immigration Act of 2005. This is a nationwide policy program to integrate permanent migrants. These events provide evidence for the fact that German policy makers acknowledge the status of Germany as a country of positive net immigration.

Main goal of the national integration plan is to address language and educational deficits because of the high importance attached to formal degrees on the German labor market<sup>2</sup>. This is partly in line with policy recommendations of an OECD report<sup>3</sup> on the issue. However, the OECD recommendations stress the importance of vocational training, access to self-employment, improved organization of temporary work agencies and early language training. The OECD report also argues that anti-discrimination policies and initiatives should be introduced, especially to address the discrimination in terms of non-recognition of foreign degrees. According to Liebig (2007) there is no clear-cut evidence on discrimination in terms of wages. This picture could however be distorted by discrimination in terms of qualifications.

Germany has accepted immigrants since 1954. The rapid economic recovery after the war demanded a work force and contracts were made with those countries that provided workers – mainly Mediterranean. Even after the oil crisis in 1973, a large proportion of immigrants and their families, some of whom came under family re-unification policies, stayed on and became permanent residents. This phenomenon is a main origin for a second-generation of migrants, born in Germany of foreign parents. Between 1988 and 1993 another wave of immigrants, asylum seekers and ethnic Germans, came to Germany. The two waves of different types immigration add complexity in the assessment of integration if immigrants in Germany.

Up to beginning of the 1990s the labour market situation of the foreign born population was similar to that of the native born<sup>4</sup>. However from the beginning of the 1990s Germany experienced a period of economic stagnation - with exception of the years 1998 and 2000 - and at the same time a particularly strong inflow of immigrants. The immigrants were hit strongest by the economic downturn<sup>5</sup>. Some steps to integrate immigrants were realized on a political level as early as 1974 when language classes for foreign workers were introduced. But the government nevertheless did not recognize Germany as an immigration country nor the necessity of the integration of immigrants up to the ratification of the Immigration Act mentioned above.

The developments would prove an integration policy useful. Elements of this policy would obviously include language promotion, equality of chances and social integration as measures of integration. These measures are supported if the immigrants have the will to integrate and to cooperate, in other words if they are motivated and believe in their success in integrating.

---

<sup>2</sup>See for example Gang and Zimmermann (1999).

<sup>3</sup>The report "The Labour Market Integration of Immigrants in Germany" (2007) by Thomas Liebig was published in the *OECD Social Employment and Migration Working Papers*.

<sup>4</sup>According to Liebig (2007) an exception was the employment rate of immigrant women and in particular of Turkish immigrant women.

<sup>5</sup>According to Liebig (2007) immigrants experienced a decrease in employment rates by ten percentage points, whereas the natives' employment rates decreased only by three percentage points.

## 2.2 Labor Market Outcomes and Psychological Factors in Economics

The so-called bell-curve argument started by Herrnstein and Murray (1994) and their book "The Bell Curve: Intelligence and Class" in which they state, that IQ - or cognitive abilities - matters for socioeconomic success. They also argue that IQ is genetically inherited and is distributed across the population in form of a bell-curve and that some groups are marginalized in society due to their position within this curve. A response<sup>6</sup> to this discussion was given by Bowles, Gintis and Osborne (2001a, 2001b), henceforth BGO. In their work BGO claim that not only the cognitive skills matter but also personal traits and subsequently they provide a theoretical foundation and empirical evidence for their statement.

BGO believe that there are four puzzles in the wage determination literature which can be solved by taking into account personality traits as determinants of earnings. Firstly, earnings differ among apparently similar individuals. According to BGO in the United States only up to a third of the variation of log earnings are explained by the classical earnings determinants age, gender, education, parental education and occupation. Secondly, the advantages that successful parents transmit to their children go beyond the possibility to offer higher educational quality and the genetic transmission of cognitive skills. According to BGO studies seeking to explain the covariance of parental economic status and the respondents' income find considerable unexplained variance even after taking into account measures of IQ and the quality of schooling. Thirdly, there are supposedly irrelevant personality traits (such as home cleanliness) that matter for wage determination. Apparently, variables such as height, beauty or obesity are often robust determinants in earnings equations. Fourthly, findings on the effectiveness of school resources are controversial. Some authors find a positive impact on later earnings but not on educational performance. BGO provide evidence that non-cognitive traits might explain these puzzles. For example, managers rank such traits high on a list of desirable attributes a worker should have. They also cite a series of papers by Heckman and coauthors<sup>7</sup> which study the labor market success of General Educational Development (GED) diplomas. GED diplomas are tests are usually taken by high-school drop outs. The authors find that the cognitive skills of these drop-out are higher than those of high school graduates, but that GED graduates lack non-cognitive skills and usually have behavioral difficulties.

There are many personality traits and it is a complex task to determine those most relevant for labor market returns. BGO summarize this literature to some extent and find, that the traits that matter, are motivation as opposed to fatalism - measured in this paper by the locus of control, communication skills, attitudes, self-esteem and conscientiousness. A preference for challenge

---

<sup>6</sup> Another prominent response was "Lessons from the Bell Curve" by Heckman in 1995.

<sup>7</sup> The authors name Heckman, Hsee and Rubinstein (1999) "The GED as a Mixed Signal", *University of Chicago* and Cameron and Heckman (1993) "The Non-equivalence of the High School Equivalents", *Journal of Labour Economics*.

and fear of failure have an effect and can be seen as the risk attitude of an individual. Borghans, Duckworth, Heckman and terWeel (2008) have written an exhaustive account of the connection between personality psychology and economics and examine the explanatory power and interpretation of personality traits in economic models. They extensively review the relevant literature. The Five-Factor model on the five major dimensions of personality can also be an indicator of which noncognitive skills or personality traits to use. This literature claims that the five traits of extraversion, agreeableness, conscientiousness, emotional stability and openness constitute a personality. Farkas (2003) argues that additionally to the five factors, leadership, sociability and social sensitivity have an effect on socioeconomic success of individuals. Mueller and Plug (2006) estimate the effect of all five elements of the five-factor personality theory<sup>8</sup> on earnings of Wisconsin high school graduates. They find that all five elements have a significant effect on earnings.

Motivation, measured in this paper by the locus of control, is a trait studied widely among researchers on the relationship between labor market success and non-cognitive traits. Duncan and Dunifon (1998) for instance have written a study of the long-run effects of motivation on labor market success.

The literature seems to agree that social background in turn plays a role in determining the personal traits affecting socioeconomic success and that noncognitive skills develop over time Cunha, Lochner and Masterov (2006) and Cunha and Heckman (2007) and Cunha, Heckman and Schennach (2010) develop a framework to model the change in noncognitive skill over time. Cunha, Heckman and Schennach (2010) model the development of both cognitive and non-cognitive skills as production functions at different time periods. Inputs into this production process are the parental environment, investments made at each period and initial personal endowments. The production process is modelled by a dynamic factor model with endogenous inputs. An advantage of their methodology is that they do not rely on test scores, which have no natural scale. They identify the scale by estimating the effect of the latent factors on adult outcomes. The authors find that interventions are more successful at the early stages of cognitive development, whereas noncognitive skills can also be successfully formed at later stages.

In the literature on immigration and integration of immigrants, and especially in the psychological literature, acculturation and identity strategies are seen as key factors for integration (see Berry (2001)). Sociability is also studied as a key factor for integration by dePalo, Faini and Venturini (2006). Fertig (2004) analyses the differences in leisure-time activities and attitudes of foreign immigrants, ethnic Germans and different generations and finds that both generations have differences in attitudes compared to Germans. Second generation immigrants seem to be the most fatalist and pessimist.

This paper enriches the literature on psychological determinants in economic

---

<sup>8</sup>The five factors are extroversion, agreeableness, conscientiousness, neuroticism, and openness to experience.

models for immigrants by studying the effect of the locus of control on immigrants' labor market outcomes.

### 3 Data and Variable Definitions

We use data from the 2007 wave of the German Socioeconomic Panel. Despite some inaccuracies the GSOEP is a valuable dataset, especially rich through questions going beyond purely observable characteristics. It is particularly of interest for this study since it includes personality questions as well as migrant-related and detailed data on labor market returns and educational history of individuals. The sample consists of immigrants and natives aged 17-30 in 1999 (so 25-38 in 2007), not in education in 2007, with provided information on the dependent variables. Employment  $D$ , age and education are measured in 2007. Psychometric measures, importance of religion are measured in 1999. The sample size is 1812. There are 111 immigrants (6.1% of the sample) and 243 children of immigrants (13.4% of the sample). The German statistical office reports a percentage of 8.8% of "foreign population" (inhabitants of Germany with foreign nationality) in Germany in 2008 <sup>9</sup>.

An immigrant is defined as "*foreign born with no German nationality at birth*", immigrants' children ("second generation") are defined as "*born in Germany with no German nationality at birth*".

To measure education levels I constructed three categories according to the ISCED<sup>10</sup> classification. ISCED 0-2 includes education up to the level of general elementary schooling and indicates a low education level, ISCED 3-4 includes "middle vocational schooling" and "vocational plus Abitur" and indicates a medium education level and ISCED 5-6 includes "higher vocational schooling" and "higher education" and indicates a high education level. Each category is controlled for by a dummy variable.

To take into account the different nationalities present in the sample I constructed three geopolitical nationality groups "EU15", "Central Europe and former Soviet Union" and "Turkey". Turkish immigrants are a large group among German non-nationals. A foreign language indicator takes the value one, if the only language spoken at home is the foreign language.

#### 3.1 Measuring Personality: The Locus of Control

Personality measures have been developed by personality psychologists using self reported questionnaires with so-called psychometric questions. They found five factors - openness to experience, conscientiousness, extraversion, agreeableness, neuroticism. Motivation can be allocated to conscientiousness, which includes the facet "striving to achievement"<sup>11</sup>. Personality psychologists have found that

<sup>9</sup>see [http://www.statistik-portal.de/Statistik-Portal/de\\_jb01\\_jahrstab2.asp](http://www.statistik-portal.de/Statistik-Portal/de_jb01_jahrstab2.asp)

<sup>10</sup>See UNESCO (2006): ISCED 1997 - International Standard Classification of Education, [www.uis.unesco.org](http://www.uis.unesco.org).

<sup>11</sup>See Borghans, Duckworth, Heckman and terWeel (2008).

personality is partly inherited and partly determined by the environment.

The psychometric questions to measure the locus of control are based on measures formulated by Rotter (1966). They are chosen from the following set of questions, present in the 1999 sample of the German Socioeconomic Panel:

1. How my life goes depends on me
2. Compared to other people, I have not achieved what I deserve
3. What a person achieves in life is above all a question of fate or luck
4. If a person is socially or politically active, he/she can have an effect on social conditions
5. I frequently have the experience that other people have a controlling influence over my life
6. One has to work hard in order to succeed
7. If I run up against difficulties in life, I often doubt my own abilities
8. The opportunities that I have in life are determined by the social conditions
9. Inborn abilities are more important than any efforts one can make
10. I have little control over the things that happen in my life

The answers can be "totally disagree", "slightly disagree", "slightly agree", "totally agree". I merge the first two categories to improve identification of the model, since the first two categories are characterized by low frequencies. Agreement with questions 1,4,6,9 is seen as an internal locus of control whereas agreement with questions 2,3,5,7,8,10 is seen as an external locus of control. The questions are chosen using the correlation matrix of the ten items. Five items display bivariate correlations with each other above 0.3, which is considered sufficiently large in this context. These five items are

2. Compared to other people, I have not achieved what I deserve
3. What a person achieves in life is above all a question of fate or luck
5. I frequently have the experience that other people have a controlling influence over my life
7. If I run up against difficulties in life, I often doubt my own abilities
10. I have little control over the things that happen in my life

### **3.1.1 The Relation between Economic Preferences and Personality Measures**

Previous literature in economics has attempted to link psychometric questions to economic parameters. As mentioned above, Borghans, Heckman, Duckwoth and terWeel (2008) survey this emerging literature. They come to the conclusion that personality traits introduced both on an empirical and a theoretical level into economics can be fruitful for economic theory. According to the authors classical economic theory should incorporate the fact that economic preferences might be consequences of constraints imposed by cognitive skills and personality traits. They name the example that an agent with a high rate of time preference might be due to the fact that the agent cannot imagine the future. This ability to

imagine would be interpreted as a personal trait<sup>12</sup>. The authors see latent factor theory as a crucial connecting tool between psychology and economics. But the authors see as a major problem with this approach, especially in economics, the problem of reverse causality between personality traits and outcomes. A more self-confident person is prone to have a higher income, but a higher income also increases self-confidence. Borghans, Heckman et al (2008) name the work of Carneiro, Hansen and Heckman (2003) as a successful example for incorporating psychometric questions in an economic outcome model in a way that addresses the problem of endogeneity.

Borghans, Golsteyn Heckman and Meijers (2009) also contribute to this literature and study the psychological determinants of the risk aversion parameter, often found in economic theory. They study in particular the gender difference in risk aversion and find that people who are less agreeable, less neurotic and who have more ambition are less risk averse. They also find evidence for the hypothesis that the differences in risk aversion across gender are due to the fact that women differ in terms of their non-cognitive skills from men.

The theoretical economic literature also gives some economic insight applicable to the psychometric questions in this paper. For the questions 2 and 7 a working paper by Eeckhout and Weng (2009) provides some economic contents. In their labor learning model, neither workers nor firms know the worker's type, but they learn it by the wages a worker receives over time. Workers and firms can observe all wages received of all workers across time. This outcome process depends on a random error term. Such a model can explain why an individual would assume his abilities are of a low type if he receives only low wages. In question 7, "running up against difficulties in life", can be understood as receiving only lower wages. The fact that every worker can see the outcome of all other workers can explain why workers might think that others have achieved more than they, given their beliefs about their own abilities.

Another attempt to economically model the content of a set of psychometric questions - in this case the locus of control - was made by Bowles, Gintis and Osborne (2001a). They develop a theoretical model to explain the advantage for an employer to employ a motivated worker by setting up a signalling model. They interpret the locus of control as an employee's preference, which reduces the employer's cost to induce the employee's effort. This simply means that a worker, who believes more in his own success is more motivated to induce effort. This could also be interpreted that a worker with intrinsic motivation will not need an external source to make him provide an effort. This is a desirable trait of a worker for an employer and the authors call such a trait an "incentive enhancing preference". Bowles, Gintis and Osborne (2001a) use a principal-agent model to interpret the relationship between incentive-enhancing preferences and earnings in an economic way. In the model the employer (the principal) and the employee (the agent) cannot contract on effort. The employer has an imperfect measure of effort. The employee has an effort-dependent belief

---

<sup>12</sup>As addressed in our econometric strategy below, this trait could also be due to the family background and previous experiences.

on the termination of the contract. An incentive-enhancing preference is in this model an employee's attribute, which makes him work harder at every wage level. As an example of an incentive-enhancing preference in this model the authors name the locus of control. An individual with an external locus of control is a fatalist person who does not believe that effort can change any outcomes in his life. In Bowles' et al (2001a) model fatalism would enter the model via lowering the worker's belief on the effectiveness of effort on lowering the probability of termination of the contract. In the model fatalism lowers the marginal subjective benefit to exerting effort in the agent's first order condition. By entering the first order condition greater fatalism lowers the agent's optimal choice of effort.

The reviewed literature in the paragraph shows that there is a growing recognition in economics and translation into economic concepts of psychological concepts and psychometric measures.

## 4 Econometric Strategy

As a first step we estimate the model in a traditional way: we treat the Rotter index as exogenous and estimate a psychometric model separately from an economic outcome model. As a second step we endogenize the Rotter index, which allows us to examine the effects of the factors determining the locus of control. This approach allows a comparison between the traditional method and an integrated methodology which allows to address the problem of endogeneity and to assess the effect of determinants on the locus of control. In the next sections I will first outline the two methodologies - the traditional one and the integrated one and then proceed to the interpretation of the results.

### 4.1 Exogenous Personality : Two step estimation procedure

#### 4.1.1 The Personality Model

As a first step, we treat the Rotter index as exogenous and we estimate two models separately. The first model is a classic factor model. Factor models have been developed in psychology to measure intelligence<sup>13</sup>. Later factor models were also used to measure other personality traits, in political science for measuring concepts and in financial economics to measure latent concepts which influence financial markets.

The main idea of factor models is to use a set of measures for the concept "intelligence", "discipline", "peace" or "beliefs on the stock market" and to divide the joint variation among these measures into a common part  $\theta$  and a random part  $\varepsilon$  and to estimate the common part  $\theta$  and its effect on the measures, indicated by  $\alpha$ .  $\theta$  indicates in this paper the locus of control, which is

---

<sup>13</sup>See Spearman (1904).

measured using a set of questions related to the locus of control<sup>14</sup>. The model is a simultaneous equation model of the five psychometric questions above. Each psychometric question is modelled as an ordered probit model. All five questions are assumed to depend on a latent factor  $\theta$ , the locus of control, and an independent random error term  $\varepsilon^M$ . The psychometric questions all depend differently on the latent factor - each question has a different factor loading  $\alpha^M$ , which can be interpreted as a coefficient of the latent factor in the regression of  $M$  on  $\theta$ . The model is estimated with a maximum likelihood methodology, implemented in STATA. The model takes the form:

$$\begin{aligned}
 M_1 &= \{1, 2, 3\} \\
 M_1^* &= \alpha^{M_1}\theta + \varepsilon^{M_1} \\
 M_2 &= \{1, 2, 3\} \\
 M_2^* &= \alpha^{M_2}\theta + \varepsilon^{M_2} \\
 M_3 &= \{1, 2, 3\} \\
 M_3^* &= \alpha^{M_3}\theta + \varepsilon^{M_3} \\
 M_4 &= \{1, 2, 3\} \\
 M_4^* &= \alpha^{M_4}\theta + \varepsilon^{M_4} \\
 M_5 &= \{1, 2, 3\} \\
 M_5^* &= \alpha^{M_5}\theta + \varepsilon^{M_5}
 \end{aligned}$$

**Parametric Identification of Factor Models** Here I give a brief outline of the identification of factor models. Factor models take the form of the measurement equation above:

$$M^* = \alpha^M\theta + \varepsilon^M$$

Consider  $M^*$  to be computable.

The identification of factor models is based on the covariance matrix of the items:

$$cov(M^*) = \Lambda\Sigma_f\Lambda' + \Omega_e$$

where

$$\begin{aligned}
 \theta &\perp \varepsilon^M \\
 \varepsilon^M &\sim N(0, 1)
 \end{aligned}$$

$\Lambda$  - matrix of factor loadings  $\alpha^M$

$\Sigma_f$  - variance-covariance matrix of the factors

$\Omega_e$  - diagonal matrix of "uniqueness"-variances of  $\varepsilon^M$

$K$  - number of factors  $\theta$

$L$  - number of items  $M$

---

<sup>14</sup>These questions have been developed by Rotter (1966).

The goal is to identify  $K \times L$  factor loadings  $\Lambda$  and  $K$  variances of factors  $\Sigma_f$ . The elements of  $cov(M)$  are observable and the elements of  $\Omega_e$  are determined by our distributional assumption on  $\varepsilon^M$ . So we can identify the unobservable elements  $\Lambda$  and  $\Sigma$  with the  $(L(L - 1)/2)$  observable off-diagonal elements of  $cov(M)$ . So, we need that

$$L(L - 1)/2 \geq (L \times K) + K.$$

The number of unique terms in  $cov(M)$  needs to be equal to or larger than the number of factor variances and factors. In our case  $K = 1$  and  $L = 5$ . So we have

$$\begin{aligned} 5 * 4/2 &\geq 5 + 1 \\ 10 &\geq 6 \end{aligned}$$

#### 4.1.2 The Employment Model

The second model is an employment model. The latent factor  $\theta$ , estimated through the model above, is treated as an additional explanatory variable in the employment equation. The model takes the form:

$$\begin{aligned} D &= \{0, 1\} \\ D^* &= \beta_0^D + \alpha^D \theta + \beta^D X + \varepsilon^D \end{aligned}$$

## 4.2 Endogenizing Personality: Simultaneous Equation Model

The models above treat the personality measure, the locus of control, as exogenous. As a next step, we endogenize the locus of control - firstly to address its endogeneity in an employment equation and secondly to find out, how the locus of control is determined. Especially, we are interested in whether immigrants and their children have different positions on the locus of control scale. Additionally, the methodology, we use for this, allows to estimate all parameters and the locus of control at the same time. This avoids treating the measure of the locus of control as an observed quantity as we did above.

The model is a linear parametric simultaneous equation model with an embedded factor model structure, as described above. The simultaneous equation model contains the equations for the economic outcome  $D$  and for the measures  $M$ . In this paper the latent concept "locus of control" is endogenized and so I add another equation in the simultaneous equation model to determine  $\theta$

The model then takes the following form:

$$\begin{aligned}
D &= \{0, 1\} \\
D^* &= \beta_0^D + \alpha^D \theta + \beta^D X + \varepsilon^D \\
M &= \{1, 2, 3\} \\
M^* &= \alpha^M \theta + \varepsilon^M \\
\theta &= \gamma W + \varepsilon^\theta
\end{aligned}$$

where  $D$  is an employment indicator,  $M$  signifies psychometric measures for the locus of control. Since  $M$  and  $D$  are categorical variables we need to impose a probit structure on the variables, so  $D^*$  and  $M^*$  indicate the latent underlying variables for the probit models for  $M$  and  $D$ .  $X$  comprises the observable variables (called *direct effects*).  $W$  comprises the observable variables (called *indirect effects*) "age", "gender", "immigrant", "religion important", "father upper secondary education", "mother upper secondary education", "father education missing", "mother education missing", "employment in 99", "still in education in 99".

#### 4.2.1 Identification Assumptions

The identification strategy is parametric and we need to make assumptions on the distributions of the error terms and of the latent concept.

$$\begin{aligned}
\varepsilon_D &\sim N(0, 1) \\
\varepsilon_M &\sim N(0, 1)
\end{aligned}$$

We need to impose normalization conditions on  $D_i^*$  and  $M_i^*$ :  $V(D_i^*)$  is normalized to 1,  $V(M_i^*)$  is normalized to 1 and we impose normality on  $\theta$  conditional on  $W$ .

Finally we need to impose conditional independence conditions:

$$\begin{aligned}
\theta &\perp \varepsilon_M | W \\
\theta &\perp \varepsilon_D | W, X \\
X &\perp \theta | W \\
D &\perp M | \theta, X \\
M_j &\perp M_{j-1} | \theta \quad \forall j
\end{aligned}$$

For all tricategorical items the cutpoint between the first and the second category is  $c_1 = 1$ .

#### 4.2.2 Estimation: The Gibbs Sampler

The likelihood function of the model under the assumption of independently and identically distributed observations is given by

$$\begin{aligned}
& \prod_{i=1}^N f(M_i, D_i, M_i^*, D_i^*, \theta_i | X_i, W_i, \alpha, \beta, \gamma, c) \\
= & \prod_{i=1}^N f(M_i^*, D_i^*, \theta_i | X_i, W_i, \alpha, \beta, \gamma, c) \prod_{i=1}^N f(M_i, D_i | \theta_i, M_i^*, D_i^*, X_i, W_i, \alpha, \beta, \gamma, c) \\
= & \prod_{i=1}^N f(M_i^*, D_i^*, \theta_i | X_i, W_i, \alpha, \beta, \gamma, c) \prod_{i=1}^N f(M_i, D_i | c)
\end{aligned}$$

where the factor loadings are written as  $\alpha = (\alpha^M, \alpha^D)$  and the coefficients as  $\beta = \beta^D$ . The first simplification follows from exploitation of the product rule. The second step follows from the fact that ordinal responses are solely determined by the underlying variables  $D_i^*$  and  $M_i^*$  and by the cutpoints  $c$ . We can factor out the likelihood function  $f(M_i^*, D_i^*, \theta_i | X_i, W_i, \alpha, \beta, \gamma, c)$  into  $f(M_i^*, \theta_i | \cdot) f(D_i^*, \theta_i | \cdot)$  due to the conditional independence assumptions above. The likelihood functions of  $D_i^*$  and  $M_i^*$  written separately are

$$\begin{aligned}
& \prod_{i=1}^N [f(M_i^*, \theta_i | \alpha, \gamma, c, M_i, W_i) \{ \sum_{k_M=1}^{K_M} 1(M_i = k_M) 1(c_{k_M-1} < M_i^* < c_{k_M}) \}] \\
& \prod_{i=1}^N [f(D_i^*, \theta_i | \alpha, \beta, \gamma, D_i, X_i, W_i) \{ \sum_{k_D=1}^{K_D} 1(D_i = k_D) 1(c_{k_D-1} < D_i^* < c_{k_D}) \}]
\end{aligned}$$

Each of the factors  $f(M_i^*, \theta_i | \cdot)$  and  $f(D_i^*, \theta_i | \cdot)$  needs to be multiplied by two indicators - an indicator which equals one if the observation  $M_i$  ( $D_i$ ) falls in category  $k_M$  ( $k_D$ ) and an operator indicating that  $M_i^*$  ( $D_i^*$ ) must fall between the two cutpoints  $c_{k_M-1}$  ( $c_{k_D-1}$ ) and  $c_{k_M}$  ( $c_{k_D}$ ) according to its category.

$\theta_i$  is unobservable and will be estimated. To make the mechanism by which  $\theta_i$  influences  $M_i^*$  and of  $D_i^*$  perspicuous we integrate out  $\theta_i$  and obtain the distributions of  $M_i^*$  and  $D_i^*$  conditional on the parameters of the model and on the data.

$$\begin{aligned}
f(M_i^* | \alpha, c, \gamma, M_i, W_i) &= \int_{\theta} f(M_i^* | \alpha, c, \theta_i, M_i) f(\theta_i | \gamma, W_i) d(\theta_i) \\
f(D_i^* | \alpha, \beta, \gamma, D_i, X_i, W_i) &= \int_{\theta} f(D_i^* | \alpha, \beta, c, \theta_i, D_i, X_i) f(\theta_i | \gamma, W_i) d(\theta_i)
\end{aligned}$$

It becomes obvious that the likelihood function of the model is a high-dimensional integral, which cannot be solved analytically and needs to be solved by numerical methods. Markov Chain Monte Carlo<sup>15</sup> methods provide a way

<sup>15</sup>The Gibbs sampler and Bayesian statistics are assessed in chapter one.

to estimate the parameters of interest by sampling from the integral. The main advantage of the Gibbs sampler is its relative computational ease.

The Gibbs sampler is a Bayesian method. The Bayesian paradigm specifies statistical models as a posterior joint distribution, composed of the two elements prior distribution and likelihood function. The prior distribution contains the beliefs of the researcher about the parameters before taking into account the information in the data. The prior is combined with the likelihood function, which contains the information of the data. The posterior joint distribution is obtained by simply multiplying the priors with the likelihood and it can be written as

$$f(\beta, \alpha, \gamma, \theta_i, M^*, D^*, c | M, D, X, W) \\ \propto f(\beta)f(\alpha)f(\gamma)f(c) \prod_{i=1}^N f(M_i, D_i, M_i^*, D_i^*, \theta_i | X_i, W_i, \alpha, \beta, \gamma, c)$$

where  $f(\beta)f(\alpha)f(\gamma)f(c)$  are the priors for the coefficients of  $X$ , the factor loadings, the coefficients of  $W$  and the cutpoints.

The Gibbs sampler is an algorithm which samples from this joint posterior distribution in a sequential way. The idea of the Gibbs sampler is to sample one of the elements among  $M_i^*, D_i^*, \beta, \alpha, \gamma, c$  and  $\theta$  at a time, conditioning on the last sampled values for the remaining elements and on the data. This procedure is equivalent to sampling from a set of conditional distributions sequentially. Each conditional distribution is a conditional posterior distribution of a parameter value given the last sampled values of the other parameters and the data. These conditionals - each of them constitutes one step of the Gibbs sampling algorithm - are called "full conditionals". The closed form of the full conditionals follows from the properties of the model. After a sufficient amount of iterations, the algorithm converges under a set of regularity conditions<sup>16</sup> and the sampled values are samples from the true posterior<sup>17</sup>. The algorithm for the model in this paper ran for 100 000 iterations and convergence statistics do not indicate that the algorithm has not converged. In the following I derive the full conditionals of the model.

First a value is sampled from the posterior conditional distribution (or full conditional) of the latent underlying variables, then from the posterior conditional distribution of the factor loadings and so forth. For the second iteration the same procedure is repeated, conditioning on the sampled values from the first iteration. The very first iteration starts with a set of specified initial values. The algorithm is not sensitive to the choice of the starting values.

<sup>16</sup>These conditions are assessed in chapter one.

<sup>17</sup>For the theory MCMC algorithms and on the Gibbs sampler, see Robert and Casella (2004).

### 4.2.3 The Posterior Conditional Distribution of the Latent Underlying Variables

Albert and Chib (1993) propose a data augmentation procedure to sample latent underlying variables in a threshold model. It follows from their work that the full conditional for the latent underlying variable of the binary response is

$$f(D^*|\alpha^D, \beta^D, \theta, D, X) \propto \prod_{i=1}^N f(D_i^*|\beta^D X_i^D + \alpha^D \theta_i, 1) \left\{ \sum_{k_D=1}^{K_D} 1(D_i = k_D) 1(c_{k_D-1} < D_i^* < c_{k_D}) \right\}$$

where  $\alpha^D, \beta^D, \theta$  signify the last sampled values from the previous iteration of the algorithm. It follows from the normality assumptions on  $\theta$  and  $\varepsilon$  that  $f(D_i^*|\theta_i, \alpha, \beta, D_i, X_i)$  is normally distributed - with mean  $\beta^D X_i^D + \alpha^D \theta_i$  and  $V(D_i^*)$  normalized to one as indicated above.

The latent underlying variable is distributed as the following truncated normal distributions:

$$\begin{aligned} D_i^*|\alpha, \beta, \theta, D, X &\sim TN_{(-\infty, 0)}(\beta^D X_i^D + \alpha^D \theta_i, 1) \text{ if } D_i = 0 \\ D_i^*|\alpha, \beta, \theta, D, X &\sim TN_{(0, \infty)}(\beta^D X_i^D + \alpha^D \theta_i, 1) \text{ if } D_i = 1 \end{aligned}$$

Similarly, the full conditionals for the polytomous variables are

$$f(M^*|\alpha, \theta, c, M, X) \propto \prod_{i=1}^N f(M_i^*|\alpha^M \theta_i, 1) \left\{ \sum_{k_M=1}^{K_M} 1(M_i = k_M) 1(c_{k_M-1} < M_i^* < c_{k_M}) \right\}$$

The latent underlying variables of the polytomous items is distributed as the following truncated normal distribution:

$$M_i^*|\alpha, \theta, c, M, X \sim TN_{(c_{k_M-1}, c_{k_M})}(\alpha^M \theta_i, 1)$$

### 4.2.4 The Posterior Conditional Distribution of the Factor Loadings

The full conditional for the factor loadings for  $D$  and  $M$  can be written as<sup>18</sup>

$$\begin{aligned} f(\alpha^D|\beta, \theta, D, X, D^*) &\propto f(\alpha^D) \prod_{i=1}^N f(D_i^*|\beta^D X_i^D + \alpha^D \theta_i, 1) \\ f(\alpha^M|\theta, M, X, M^*) &\propto f(\alpha^M) \prod_{i=1}^N f(M_i^*|\alpha^M \theta_i, 1) \end{aligned}$$

<sup>18</sup> As above  $\beta$  and  $\theta$  denote the last sampled values.

where we choose normal priors  $f(\alpha^D) = N(0, 1)$  and  $f(\alpha^M) = N(0, 1)$ . If we rewrite the equation for  $D^*$  and  $M^*$  as

$$\begin{aligned} D_i^* - \beta^D X_i^D &= \alpha^D \theta_i + \varepsilon_i^D \\ M_i^* &= \alpha^M \theta_i + \varepsilon_i^M \end{aligned}$$

we can treat it as a normal regression model and derive for  $M$  and  $D$

$$\begin{aligned} \alpha^M | \theta_i, M_i, M_i^* &\sim N [(\theta_i' \theta_i + 1)^{-1} \theta_i' (M_i^*), (\theta_i' \theta_i + 1)^{-1}] \\ \alpha^D | \beta, \theta_i, D_i, X_i, D_i^* &\sim N [(\theta_i' \theta_i + 1)^{-1} \theta_i' (D_i^* - \beta^D X_i^D), (\theta_i' \theta_i + 1)^{-1}] \end{aligned}$$

#### 4.2.5 The Posterior Conditional Distribution of the Direct Coefficients

Similarly to the procedure for the factor loadings, we can write the model as

$$D_i^* - \alpha^D \theta_i = \beta^D X_i^D + \varepsilon_i^D$$

For the coefficients, we choose to set diffuse priors as well. The full conditionals for the intercepts are, according to Albert and Chib (1993,p.671)

$$\beta^D | \alpha, \theta_i, D_i, X_i, D_i^* \sim N \left[ (X_i' X_i)^{-1} X_i' (D_i^* - \alpha^D \theta_i), (X_i' X_i)^{-1} \right]$$

#### 4.2.6 The Posterior Conditional Distribution of the Cutpoints

We assume a uniform prior for the cutpoints and can write for the full conditionals for the polytomous responses

$$c^M | \alpha, \theta, M, M^* \sim \text{unif} \left[ \begin{array}{l} \max\{\max\{M_i^* : M_i = k_M\}, c_{M-1}\}, \\ \min\{\min\{M_i^* : M_i = k_{M+1}\}, c_{M+1}\} \end{array} \right]$$

#### 4.2.7 The Posterior Conditional Distribution of the Latent Factors

Similarly as for the procedure for coefficients and factor loadings, we can rewrite the model as

$$\begin{aligned} D_i^* - \beta^D X_i^D &= \alpha^D \theta_i + \varepsilon_i^D \\ M_i^* &= \alpha^M \theta_i + \varepsilon_i^M \end{aligned}$$

and treat it as a normal regression model, where  $\theta_i$  is the parameter to be estimated. Carneiro, Hansen and Heckman (2003) specify a mixture of normals as the prior for the latent factors. We treat the latent factors as endogenous

depending on  $\gamma W_i$ . We treat  $\theta_i$  in the same way as  $M_i^*$  and  $D_i^*$  for which the priors are implicitly determined by the prior distributions of the other parameters and by the assumptions on the distribution of  $\varepsilon_i^M$  and  $\varepsilon_i^D$ . The prior of  $\theta_i$  is therefore implicitly determined by the priors of the other parameters of the model and by the assumptions on the distributions of  $\varepsilon_i^M, \varepsilon_i^D$  and  $\varepsilon_i^\theta$ .

We can derive the full conditional for the latent factor as:

$$\begin{aligned} & f(\theta|\beta, \alpha, c, \gamma, X, W, D^*, M^*) \\ & \propto \prod_{i=1}^N f(M_i^*|\alpha^M \theta_i, 1) f(D_i^*|\beta^D X_i^D + \alpha^D \theta_i, 1) \end{aligned}$$

We do not need to condition on  $D$  and  $M$  since they are implicitly known through  $D^*$  and  $M^*$  and  $c$ . Our dependent variables are ordinal and for identification reasons their variances and error variances have been set to one.

The posterior conditional distribution of  $\theta_i$  is given by:

$$\begin{aligned} & \theta_i|\beta, \alpha, \gamma, c, X_i, W_i, D_i^*, M_i^* \\ & \sim N \left[ \begin{array}{c} \gamma W_i + (\alpha^{D'} \alpha^D + \alpha^{M'} \alpha^M + 1)^{-1} \\ (\alpha^{M'} (M_i^* - \alpha^{M'} \gamma W_i) + \alpha^D (D_i^* - \beta^D X_i^D - \alpha^D \gamma W_i)), \\ I - \alpha^{D'} (\alpha^{D'} \alpha^D + \alpha^{M'} \alpha^M + 1)^{-1} \alpha^D \\ - \alpha^{M'} (\alpha^{D'} \alpha^D + \alpha^{M'} \alpha^M + 1)^{-1} \alpha^M \end{array} \right] \end{aligned}$$

#### 4.2.8 The Posterior Conditional Distribution of the Indirect Coefficients

The posterior we sample from can be written as

$$\begin{aligned} & f(\gamma|\theta, W) \\ & \propto f(\gamma) f(\theta|\gamma, W) \end{aligned}$$

The model for the latent variable is

$$\theta = \gamma W + \varepsilon^\theta$$

We assume a diffuse prior for the coefficient  $\gamma$ . Similar to the procedures above we get:

$$f(\gamma|\theta, W) \sim N((W'W)^{-1}W'\theta), (W'W)^{-1})$$

## 5 Results

Table 1 shows the results of a simple Mincer equation for employment including two dummy variables for immigrants and for the second generation. The base

category are native-born who have the German nationality at birth. All coefficients display the expected signs. Being an immigrant lowers the employment probability significantly. This effect is attenuated for the second generation whereas the coefficient for the second generation is not significant. Being married does not significantly increase the probability of being employed whereas having children does<sup>19</sup>. The table shows results from two different estimation methods - firstly, from the Markov Chain Monte Carlo methodology, which is used later for the incorporation of the latent factor and secondly, the traditional Maximum Likelihood estimator for probit models. The results show that both methods give quite similar results. This fact is taken as evidence for the correct implementation of the Markov Chain Monte Carlo method since the assumptions used for both methods are the same. It can also be interpreted as evidence for the fact that the priors chosen are not very informative and that the MCMC estimator takes the information given in the data more strongly into account - in other words the data is informative.

In table 2 the results for a model taking into account the ethnic background and the language spoken at home. I split the sample into four ethnic groups - four dummy variables control for Turkish, central European (this group includes the former Soviet Union), EU15 (including Switzerland and the US) and German nationalities. The base category are all German-born with German nationality at birth<sup>20</sup>. The results for the control variables are similar to the table 1. The ethnic variables show that only the Turkish first and second generation have a significant disadvantage compared to native Germans. It must be noted however, that the sample sizes for the different ethnic groups are small and this might affect the significance of the variables.

---

<sup>19</sup>It would be of interest to split the sample into men and women since it could certainly be the case that the coefficient for marital status differs largely between men and women and therefore renders the coefficient insignificant once taking the whole sample. But here the interest lies in the immigrant population and I consider the immigrant and the second generation samples as too small to be able to split the sample.

<sup>20</sup>There is no group for immigrants with German nationality since there are none in the sample.

$\beta^D$	MCMC	ML
Intercept	-1.00 (-1.99,-0.04)	-0.481 (0.340)
Age	0.03 (0.01,0.05)	0.036 (0.009)
Gender	-0.56 (-0.70,-0.421)	-0.573 (0.072)
Immigrant	-0.40 (-0.67,-0.13)	-0.406 (0.139)
Second generation	-0.16 (-0.34,0.036)	-0.115 (0.098)
Low education	-0.40 (-0.58,-0.21)	-0.411 (0.096)
High education	0.31 (0.14,0.47)	0.293 (0.085)
Marital status	0.08 (-0.07,0.24)	0.078 (0.081)
Children under 16	0.23 (0.08,0.37)	0.260 (0.077)

Table 1: Estimates of the Employment Equation:  $D^* = \beta X + \varepsilon_D$ , estimated by MCMC and ML

$\beta^D$	ML	MCMC
Intercept	-0.57 (0.342)	-1.34 (-2.31,-0.31)
Age	0.04 (0.000)	0.04 (0.02,0.06)
Gender	-0.58 (0.072)	-0.59 (-0.72,-0.44)
Turkish immigrant	-0.65 (0.210)	-0.65 (-1.06,-0.23)
Central European immigrant	-0.07 (0.279)	-0.06 (-0.60,0.50)
EU15 immigrant	0.47 (0.352)	0.50 (-0.18,1.23)
Turkish second generation	-0.27 (0.177)	-0.45 (-0.87,-0.02)
Central European second generation	0.31 (0.256)	0.32 (-0.28,0.95)
EU15 second generation	0.18 (0.234)	0.16 (-0.30,0.64)
German second generation	-0.08 (0.135)	-0.13 (-0.38,0.12)
Immigrant foreign language spoken at home	-0.52 (0.332)	-0.53 (-1.19,0.12)
Second generation foreign language spoken at home	-0.26 (0.330)	-0.25 (-0.89,0.41)
Low education	-0.36 (0.097)	-0.37 (-0.56,-0.17)
High education	0.28 (0.081)	0.28 (0.11,0.45)
Marital Status	0.10 (0.08)	0.10 (-0.06,0.25)
Children under 16	0.26 (0.773)	0.26 (0.10,0.41)

Table 2: Estimates of the Employment Equation:  $D^* = \beta X + \varepsilon_D$ , estimated by MCMC and ML

## 5.1 Adding Personality

Table 3 to 5 show the results for a simple model of employment - for natives, immigrants and their children, adding a measure of the locus of control. The results in table 3 are quite similar to those of table 1. The coefficient for immigrants is slightly less negative than without controlling for the locus of control and the coefficient for the second generation is much less negative but still insignificant. Again, being an immigrant is a disadvantage on the labour market and the disadvantage is attenuated for the second generation. The last row in table 3 show the results for the locus of control. They show that having a more internal locus of control has a positive and significant effect on the employment probability. We can see that a  $2.5 \sigma$  locus of control units can compensate for being an immigrant and 1 for being a second generation immigrants. About four  $\sigma$  locus of control units can compensate for having a low educational attainment level as opposed to a medium one.

The three columns in table 3 refer to three different ways of estimating the model - one Maximum Likelihood method and two MCMC methods. The Maximum Likelihood method is a two-step method whereas the MCMC methodology allows to estimate all parameters simultaneously. The former way of estimating a model with latent variables is easy to implement since most software includes a routine to estimate a conventional factor model and a routine to estimate a probit model. Both models - the probit model for the outcome equation and the factor model - can be estimated using maximum likelihood procedures<sup>21</sup>. A two-step methodology has the disadvantage, that the latent factor is treated as an observable variable in the second stage. This is a less efficient method than a method estimating all parameters simultaneously and taking into account that the latent variable is an estimated and not an observable entity. On the other hand, if the latent factor is estimated in a wrong way, any mistake is carried on to the estimation of the remaining parameters. My results show that both estimation methods render highly similar results - even for the coefficient of the latent factor.

As outlined above, the locus of control is treated here as a variable that is determined by socioeconomic conditions and possibly partly also by genetical heritage. The second and third column show results for a model in which the locus of control is estimated without assuming that it is determined by other variables and for a model in which the locus of control is estimated assuming that it is determined by socioeconomic variables, respectively. The table shows that the results do not differ greatly.

Table 4 shows the estimations of the factor loadings. They are positive for all items and using all three ways of estimating the model. They differ in size depending on the estimation methodology used and on whether the latent factor

---

<sup>21</sup>A maximum likelihood method for estimating a factor model is Rao's canonical-factor method, which is based on maximizing the determinant of the correlation matrix of the items by seeking the highest canonical correlation with the items. This methodology is based on the assumption that the factors are continuous. Certainly this is a less accurate methodology but for the mere purpose of comparison to the methodology used throughout my dissertation I assume the level of accuracy sufficient.

is treated as exogenous or as endogenous. In table 5 I show the coefficients of the determinants of the locus of control. The results show that immigrants have a much lower level of beliefs that they can influence the outcomes of their lives; they seem to be more fatalistic. The same is true for the second generation but the effect is attenuated. Age and the duration of stay in Germany have a small, significantly positive effect on the locus of control. Mother's education - interestingly, as opposed to father's education - has a positive and significant effect on the locus of control. Education does not have a significant effect on the locus of control, but the fact of still being in education does.

The outlined results suggest that motivation matters for everyone and that immigrants can overcome their gap by motivation. But immigrants have a *double disadvantage* since they have less motivation than natives and more motivation would actually help them find a job.

$\beta^D, \alpha^D$	ML	MCMC exog	MCMC endog
Intercept	-0.46 (0.341)	-1.38 (-2.34,-0.29)	-0.75 (-1.71, 0.23)
Age	0.04 (0.009)	0.04 (0.02,0.05)	0.03 (0.01,0.05)
Gender	-0.57 (0.072)	-0.58 (-0.71,-0.43)	-0.57 (-0.70,-0.42)
Immigrant	-0.37 (0.141)	-0.37 (-0.65,-0.09)	-0.34 (-0.62,-0.05)
Second generation	-0.09 (0.098)	-0.09 (-0.29,0.10)	-0.11 (-0.31,0.08)
Low education	-0.39 (0.095)	-0.40 (-0.59,-0.20)	-0.40 (-0.59,-0.21)
High education	0.27 (0.085)	0.27 (0.10,0.45)	0.27 (0.10,0.44)
Marital status	0.07 (0.081)	0.07 (-0.09,0.23)	0.02 (-0.13,0.18)
Children under 16	0.26 (0.077)	0.26 (0.11,0.42)	0.15 (0.001,0.29)
Locus of control	0.12 (0.041)	0.17 (0.04,0.24)	0.13 (0.05,0.21)

Table 3: Estimates of the Employment Equation:  $D^* = \alpha\theta + \beta X + \varepsilon_D$ , estimated by MCMC and ML

$\alpha^M$	ML	MCMC exog	MCMC endog
Not achieved what I deserve	0.49	0.99 (0.88,1.10)	0.62 (0.56,0.69)
Achievements are question of luck	0.44	0.72 (0.64,0.80)	0.44 (0.39,0.49)
Other people influence my life	0.55	1.17 (1.06,1.30)	0.72 (0.65,0.79)
Doubt my abilities	0.52	0.98 (0.88,1.08)	0.58 (0.53,0.65)
Little control over my life	0.67	1.84 (1.60,2.09)	1.37 (1.19,1.56)

Table 4: Estimates of the Psychometric Question Equations:  $M^* = \alpha\theta + \varepsilon_M$

	$\gamma$	Bayesian confidence intervals
Age	0.04	0.03, 0.04
Gender	-0.03	-0.15, 0.08
Immigrant	-0.80	-1.38, -0.18
Second generation	-0.23	-0.44, -0.00
Religion important	-0.03	-0.17, 0.10
Low education in 1999	-0.03	-0.17, 0.10
High education in 1999	0.16	-0.02, 0.35
Father highly educated	0.00	-0.19, 0.19
Mother highly educated	0.35	0.11, 0.59
In education in 1999	0.22	0.05, 0.37
Time stayed in Germany	0.04	0.00, 0.07

Table 5: Estimates of Determinants of the Locus of Control:  $\theta = \gamma W + \varepsilon_\theta$ , estimated by MCMC

In table 6 to 8 I add nationalities and an indicator of whether German is spoken at home. The control variables do not change much in size or sign. The results show that only Turkish immigrants have a significant disadvantage on the German labour market once controlling for the locus of control. It should be noted that, as mentioned above, the sample sizes for the separate ethnic groups are small. A Turkish immigrant can compensate his disadvantage on the labour market by six  $\sigma$  units of belief in being able to determine his success. As in table 2, speaking only the foreign language at home does not cause a significant disadvantage for immigrants or for the second generation. The locus of control still has a positive and significant effect. Table 6 displays again three columns for the three ways of estimating the model - by Maximum Likelihood, by MCMC assuming a latent factor not determined by other variables and assuming a latent factor determined by socioeconomic controls. Again, the results are largely similar across the methods.

Table 7 show the results for the factor loadings which are all positive and significant. Table 8 shows the results for the determinants of the locus of control, when adding nationalities and an indicator of whether German is spoken at home. Age has a small and significantly positive effect on the locus of control and mother's education seems to be an important determinant. Turkish immigrants as well as the Turkish second generation have a significantly lower belief in being able to determine their own success.

Again, the results show that - especially the Turkish - immigrants and their children have a double disadvantage on the labour market: they are disadvantaged in terms of employment and additionally they lack in belief to be able to do something to be successful in life - a skill which matters on the labour market.

$\beta^D, \alpha^D$	ML	MCMC exog	MCMC endog
Intercept	-0.55 (0.343)	-1.43 (-2.41,-0.37)	-1.41 (-2.41,-0.39)
Age	0.04 (0.009)	0.04 (0.02,0.06)	0.04 (0.02,0.06)
Gender	-0.58 (0.072)	-0.59 (-0.72, -0.44)	-0.58 (-0.72,-0.44)
Low education	-0.36 (0.098)	-0.36 (-0.55,-0.16)	-0.36 (-0.55,-0.17)
High education	0.26 (0.086)	0.26 (0.09,0.43)	0.26 (0.09,0.42)
Turkish immigrant	-0.61 (0.211)	-0.63 (-1.05,-0.21)	-0.60 (-1.02,-0.18)
Central European Immigrant	-0.06 (0.282)	-0.05 (-0.61,0.52)	-0.06 (-0.60,0.51)
EU15 immigrant	0.51 (0.355)	0.55 (-0.17,1.26)	0.56 (-0.15,1.26)
Turkish second generation	-0.25 (0.177)	-0.44 (-0.87,0.00)	-0.41 (-0.85,0.01)
Central European second generation	0.30 (0.255)	0.32 (-0.30,0.94)	0.32 (-0.29,0.92)
EU15 second generation	0.19 (0.235)	0.19 (-0.29,0.66)	0.18 (-0.28,0.66)
German second generation	-0.07 (0.135)	-0.11 (-0.37,0.14)	-0.11 (-0.36,0.15)
Immigrant foreign language spoken at home	-0.47 (0.334)	-0.48 (-1.14, 0.19)	-0.47 (-1.13,0.19)
Second generaion foreign language spoken at home	-0.18 (0.333)	-0.14 (-0.81,0.51)	-0.16 (-0.82,0.49)
Marital status	0.09 (0.082)	0.09 (-0.07,0.25)	0.09 (-0.08,0.24)
Children under 16	0.26 (0.077)	0.26 (0.11,0.41)	0.26 (0.11,0.41)
Locus of control	0.10 (0.042)	0.10 (0.02,0.23)	0.10 (0.02,0.18)

Table 6: Estimates of the Employment Equation:  $D^* = \alpha\theta + \beta X + \varepsilon_D$ , estimated by MCMC and ML

$\alpha^M$	ML	MCMC exog	MCMC endog
Not achieved what I deserve	0.49	0.99 (0.89,1.09)	0.62 (0.57,0.69)
Achievements are question of luck	0.43	0.72 (0.64,0.79)	0.44 (0.3904596 0.4925686)
Other people influence my life	0.55	1.16 (1.05,1.26)	0.72 (0.65,0.79)
Doubt my abilities	0.55	0.98 (0.88,1.08)	0.58 (0.53,0.64)
Little control over my life	0.67	1.88 (1.60,2.15)	1.33 (1.17,1.51)

Table 7: Estimates of the Psychometric Question Equations:  $M^* = \alpha\theta + \varepsilon_M$

	$\gamma$	Bayesian confidence intervals
Age	0.04	0.03,0.04
Gender	-0.04	-0.16,0.07
Religion important	-0.04	-0.17,0.10
Low education in 1999	-0.02	-0.16,0.12
High education in 1999	0.17	-0.20,0.35
Father highly educated	-0.01	-0.20,0.18
Mother highly educated	0.34	0.10,0.58
In education in 1999	0.15	0.00,0.30
Turkish immigrant	-1.06	-1.72,-0.39
Central European immigrant	-0.19	-0.76,0.38
EU15 immigrant	-0.60	-1.39,0.16
Turkish second generation	-0.55	-0.86,-0.22
Central European second generation	0.02	-0.37,0.42
EU15 second generation	-0.22	-0.61,0.15
German second generation	-0.12	-0.39,0.15
Immigrant foreign language spoken at home	-0.40	-0.97,0.18
Second generation foreign language spoken at home	-0.10	-0.74,0.53
Time stayed in Germany	0.03	-0.00,0.07

Table 8: Estimates of Determinants of the Locus of Control:  $\theta = \gamma W + \varepsilon_\theta$ , estimated by MCMC

In order to see whether the locus of control matters more for immigrants than for natives, the graphs in appendix A show the effect of the locus of control on employment for immigrants and natives and for children of immigrants and natives. The graphs show that a more internal locus of control has a positive effect for everyone - for immigrants, natives, for children of immigrants and for men and women. They also indicate an apparent gap between immigrants and natives and a smaller gap between children of immigrants and natives.

## 6 Conclusion

This paper set out to examine whether personality traits matter for the labour market performance of immigrants to Germany, using the example of the locus of control - the belief of an individual in their own ability to control their lifecourse. We find that a strong belief in control over one's life has a positive effect on the probability of being employed. Immigrants have a more external locus of control than natives, which means that they believe that their lives are more controlled by external circumstances than by themselves. The second generation also has a more external locus of control than natives, but it is already more internal than that of immigrants. This is evidence for a generational convergence of migrants' locus of control towards that of natives. Immigrants have a double disadvantage on the labour market : they are disadvantaged by lower employment chances because of their status and they are additionally disadvantaged due to having a lower sense of being able to control their life and this sense of control positively matters for the probability of being employed.

There seems to be a barrier in the German labour market towards immigrants, which can be overcome by self-confidence, belief in success, personal dedication and commitment. To a certain degree, an effort to adjust to the new country can be expected of an immigrant. But the German labour market should be of such a structure, that this effort should not be necessary to overcome discrimination, but only to start an adjustment process. German integration policies should include some measures to enhance a stronger belief in immigrants, that they will be able to manage their situation. The paper provides evidence for the success and appropriateness of policies or behavior towards migrants, which encourage their belief in their success.

## 7 Appendix: Figures

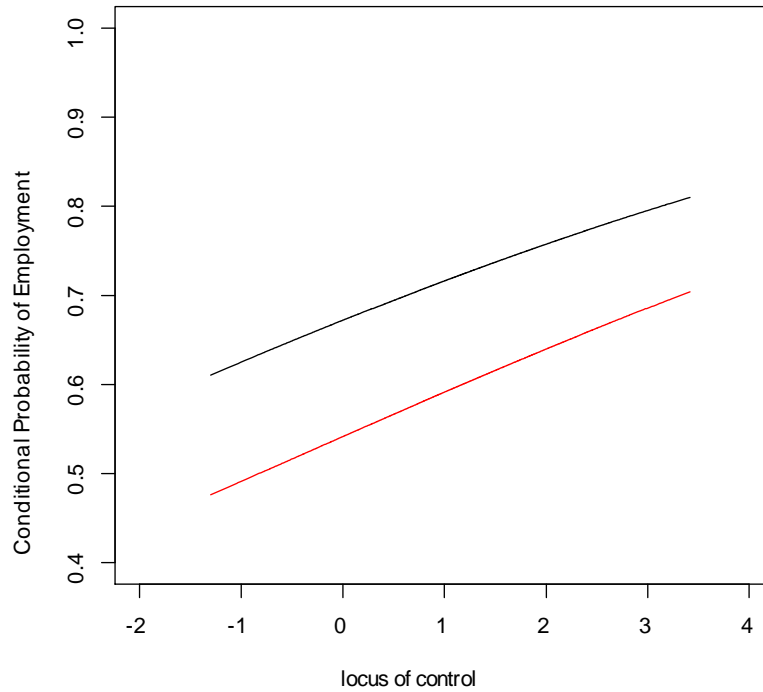


Figure 1: Figure 1: The Conditional Probability of Being Employed for Natives vs Immigrants (32 year old, medium educated, married men with children)

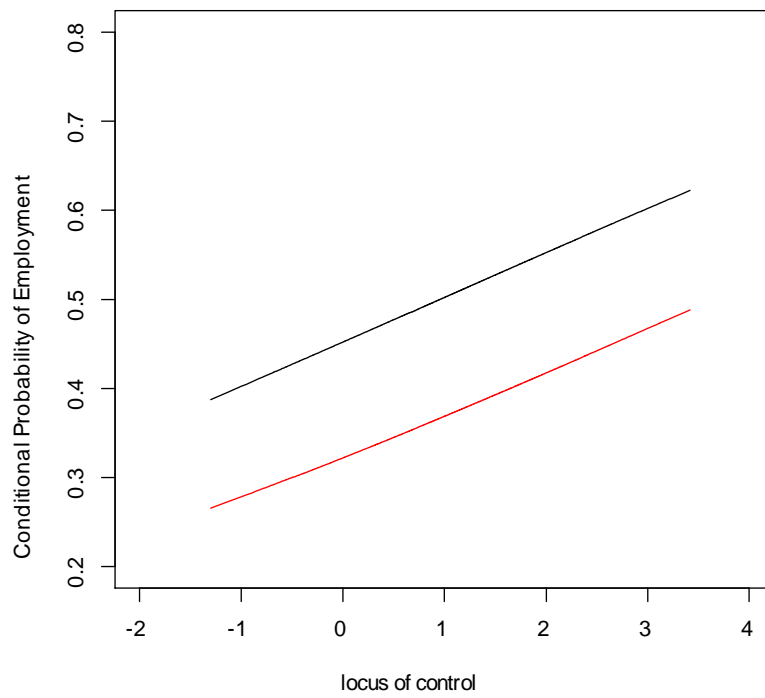


Figure 2: The Conditional Probability of Being Employed for Natives vs Immigrants (32 year old, medium educated, married women with children)

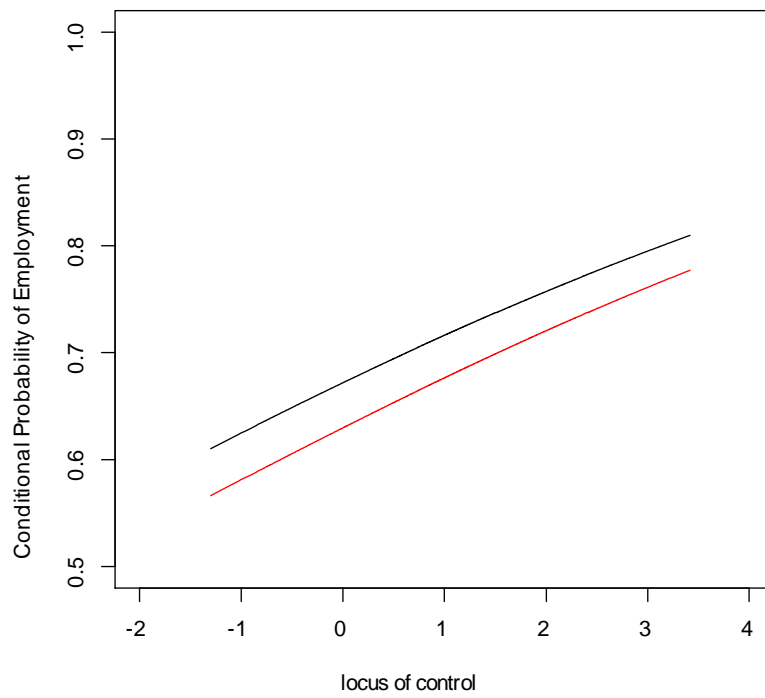


Figure 3: The Conditional Probability of Being Employed for Natives vs Second Generation (32 year old, medium educated, married men with children)

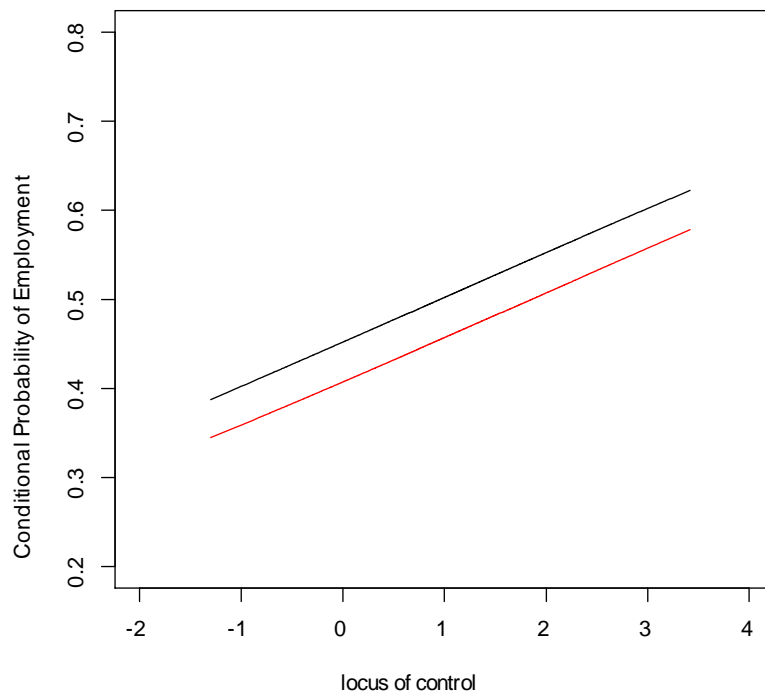


Figure 4: The Conditional Probability of Being Employed for Natives vs Second Generation (32 year old, medium educated, married women with children)

## References

- [1] Albert, J.H. & Chib, S. (1993) : Bayesian Analysis of Binary and Polychotomous Response Data, *Journal of the American Statistical Association*, 88 (422), 669-679.
- [2] Baker, M. & Dwayne B. (1994) : The Performance of Immigrants in the Canadian Labor Market, *Journal of Labor Economics* 12.
- [3] Berry, J.W. (2001) : A Psychology of Immigration, *Journal of Social Issues*, 29, 541-62.
- [4] Borghans, L.; Duckworth, A.L.; Heckman, J. & terWeel, B. (2008) : The Economics of Psychology and Personality Traits, *Journal of Human Resources*, 43, 972-1059.
- [5] Borghans, L.; Golsteyn, B.; Heckman, J. & Meijers, H. (2009) : Gender Differences in Risk Aversion and Ambiguity Aversion, *NBER Working Papers 14713*, National Bureau of Economic Research, Inc.
- [6] Borjas, G. (1985) : Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants, *Journal of Labor Economics*.
- [7] Borjas, G. (1987) : Self-Selection and the Earnings of Immigrants *American Economic Review* 77.
- [8] Bowles, S.; Gintis, H. & Osborne M. (2001a): Incentive-Enhancing Preferences: Personality, Behavior and Earnings, *American Economic Review* 91 (2), 155-158.
- [9] Bowles, S.; Gintis, H. & Osborne, M. (2001b): The Determinants of Earnings: A Behavioral Approach, *Journal of Economic Literature* 39(4), 1137-1176.
- [10] Caner, P.; Hansen, K. & Heckman J. (2003), Estimating Distributions of Treatment Effects with an Application to the returns of Schooling and Measurement of the Effects of Uncertainty of College Choice, *International Economic Review* 44(2), 361-442.
- [11] Chiswick, B.R. (1978) : The Effect of Americanization on Earnings of Foreign-born Young Men, *Journal of Political Economy* 86(5), 897-921.
- [12] Chiswick, B.R. (1988) : Differences in Education and Earnings across Racial and Ethnic Groups: Tastes, Discrimination and Investments in Child Quality, *Quarterly Journal of Economics* 103 (3).
- [13] Chiswick, B.R. & Miller, P.W. (1999): Language skills and earnings among legalized aliens, *Journal of Population Economics* 12(1):63-89.

- [14] Constant, A. & Zimmermann L. & Zimmermann, K.F. (2006): Ethnic Self-Identification of First Generation Migrants, *IZA Discussion Papers, Institute for the Study of Labor (IZA)*.
- [15] Cunha, F. & Heckman, J. (2007) : The Technology of Skill Formation, *American Economic Review* 97(2), 31-47.
- [16] Cunha, F.; Heckman, J.; Lochner, L. & Masterov,D. (2006): Interpreting the Evidence on Life Cycle Skill Formation. In Eric A. Hanushek and Frank Welch, eds., *Handbook of the Economics of Education*, chapter 12. Amsterdam, North-Holland, 697–812.
- [17] Cunha, F.; Heckman, J. & Schennach, S. (2010) : Estimating the Technology of Cognitive and Noncognitive Skill Formation, *NBER Working Paper No. 15664*.
- [18] de Palo, D.; Faini, R. & Venturini, A. (2006) : The Social Assimilation of Immigrants, *IZA Discussion Papers, Institute for the Study of Labor (IZA)*.
- [19] Duleep, H. & Regrets, M. (1999) : Immigrants and Human Capital Investment, *American Economic Review*.
- [20] Duncan, G.J. & Dunifon, R. (1998) : Long-Run Effects of Motivation on Labor Market Success, *Social Psychology Quarterly*, pp 33-48 : A Study of Asian Immigrants and their family ties. Kalamazoo MI: Upjohn Institute of Economic Research.
- [21] Dustmann, C. (1993) : Earnings Adjustments of Temporary Migrants, *Journal of Population Economics* 6.
- [22] Dustmann, C. & Schmidt, C. (2000) : The Wage Performance of Immigrant Women: Full-Time Jobs, Part-Time Jobs, and the Role of Selection, *IZA Discussion Papers* 233, *Institute for the Study of Labor (IZA)*.
- [23] Eeckhout, J. & Weng, X. (2009) : Assortative Learning, *working paper*
- [24] Eckstein, Z. & Weiss, Y. (2002) : The Integration of Immigrants from the Former Soviet Union in the Israeli Labor Market", in Ben-Bassat Avi, (ed.), *The Israeli Economy, 1985-1998: From Government Intervention to Market Economics*, Essays in Memory of Prof. Michael Bruno, MIT Press. 2002, pp. 349-378.
- [25] Fahrmeir, L. & Raach, A. (2006) : A Bayesian semiparametric latent variable model for mixed responses, *Psychometrika*.
- [26] Farkas,G. (2003): Cognitive Skills and Noncognitive Traits and Behaviors in Stratification Processes, *Annual Review of Sociology*, 29, 541–62.
- [27] Fertig, M. & Schmidt, C. (2001) : First- and Second-Generation Migrants in Germany, What do we Know and What do People Think, *IZA Discussion Papers, Institute for the Study of Labor (IZA)*.

- [28] Fertig, M. (2004): The Societal Integration of Immigrants in Germany, *IZA Discussion Papers, Institute for the Study of Labor (IZA)*.
- [29] Gang, I. & Zimmermann, K.F. (1999) : Is child like parent? Educational Attainment and Ethnic Origin, *Journal of Human Resources*.
- [30] Heckman, J. (1995) : Lessons from the Bell Curve, *Journal of Political Economy* 103 (5), 1091.
- [31] Heckman, J.; Stixrud, J. & Urzua, S. (2006) : The Effects of Cognitive and Non-cognitive Abilities on Labor Market Outcomes and Social Behavior, web supplement available at [jenni.uchicago.edu/noncog](http://jenni.uchicago.edu/noncog), *Journal of Labor Economics*.
- [32] Herrnstein, R.J. & Murray, C.A. (1994) : The Bell Curve: Intelligence and Class, *Structure in American Life*, New York, Free Press.
- [33] Hinte, H. & Zimmermann, K.F. (2005) : Arbeitsmarkt & Zuwanderung, Springer, Berlin, Heidelberg.
- [34] LaLonde, R.H. & Topel, R.H. (1992) : Immigrants in the American Labor Market : Quality, Assimilation and Distributional Effects, *American Economic Review* 8.
- [35] Liebig, T. (2007) : The Labor Market Integration of Immigrants, *OECD Social Employment and Migration Working Papers*.
- [36] Mueller, G. & Plug, E. (2006) : Estimating the Effect of Personality on Male and Female Earnings, *Industrial and Labor Relations Review* 60 (1), 3-22.
- [37] Osborne-Groves, M. (2006) : How important is your personality? Labor market returns to personality for women in the U.S. and U.K., *Journal of Economic Psychology*.
- [38] Robert, C.P. & Casella, G. (2004) : Monte Carlo statistical methods (2nd edition). Springer, New York.
- [39] Rotter, J.B. (1966) : Generalized Expectancies for Internal versus External Control of Reinforcement, *American Psychological Association*.
- [40] Spearman, C. (1904) : General Intelligence, Objectively Determined and Measured, *American Journal of Psychology* 15, 201-293.