The Determinants of Migration into the EU around the turn of the century: a Panel Data Analysis

Ilse Ruyssen[1]

SHERPPA, Ghent University, Tweekerkenstraat 2, 9000 Gent, Belgium

First draft March 2008

Abstract

This paper investigates the determinants of migration into 6 Western European destination countries, i.e. Belgium, Denmark, Finland, Germany, the Netherlands and Spain, using panel migration data from 57 origin countries (non-EU-members because of free movement in the EU) for the period 1996-2005. Our contribution is threefold. First, we use three-dimensional panel data to investigate the determinants of migration. So far, not many three-way panel data studies have been performed, which is mostly due to the lack of data availability. Only recently, countries started to keep track of migration flows in great detail. Second, in dynamic panel models the fixed effects estimator is inconsistent when the time dimension is fixed. For an AR(1) model there is a downward asymptotic bias on the coefficient of the lagged dependent variable. Given this inconsistency, we follow Everaert and Pozzi (2007) who propose a bias correction for the fixed effects estimator using an iterative bootstrap algorithm. Third, due to the reverse impact of migrant flows on stocks, the endogeneity bias comes from two explanatory variables, rather than one. Therefore, we explicitly take into account the dynamic relationship between migrant flows and stocks. When we compare the results of a standard fixed effects estimation, a fixed effects estimation with bootstrap-correction and a general method of moments estimation, we find that the choice of the estimation method has a great impact on the results and that the wrong choice could lead to invalid conclusions about the true factors behind migration.

1. Introduction

In this paper we investigate the determinants of migration into the European Union during 1996-2005. The theoretical framework is based on a model developed by Hatton (1995) to investigate the determinants of migration into the U.K. using a time series analysis. By slightly transforming this

[1] In cooperation with Prof. Dr. G. Rayp, Ghent University and Prof. Dr. G. Everaert, Ghent University.

model, we are able to estimate a dynamic panel model in three dimensions. So far, not many panel data studies on the determinants of migration have been performed, which is mostly due to the lack of data availability. Only recently, countries started to keep track of migration flows in great detail. However, using repeated observations on the same units makes the estimation more realistic than a single time-series or a two-way panel data model.

Usually, this type of models is estimated using a fixed effects estimator. Nevertheless, when the time dimension is small, the estimated coefficient of the lagged dependent variable in an AR(1) model will be downward biased because of its negative correlation with the error term. Moreover, due to the dynamic relationship between migrant flows and stocks, the latter will also be negatively correlated with the error term. This implies that the endogeneity bias is now caused by two explanatory variables, rather than one. Given the inconsistency of the fixed effects estimator for a fixed time period various instrumental variables and generalised method of moments (GMM) estimators have been proposed. However, these estimators are not without problems themselves. In order to avoid these problems analytical corrections for the FE estimator have been proposed (see e.g. Kiviet, 1995; Bun and Carree, 2005b). In this paper however we follow Everaert and Pozzi (2007) who propose a bias correction for the FE estimator using an iterative bootstrap algorithm. Like the analytical correction this bootstrap-correction aims at reducing the bias of the fixed effects estimator while maintaining its higher efficiency compared to GMM estimators. This bootstrap approach does not rely on any theoretical assumptions. Moreover, Monte Carlo results reveal that the bootstrap-corrected fixed effects estimator yields better inference than analytical corrections and GMM estimators in samples with a small to moderate time dimension. It also allows us to explicitly take into account the dynamic relationship between migrant inflows and stocks.

We will perform the standard fixed effects estimation, the bootstrap-corrected fixed effects estimation and the bootstrap-corrected fixed effects estimation which takes into account the relationship between migrant flows and stocks. It is shown that dividing the migrant flows by the population in the origin country is not enough to eliminate all scale effects in a three-way panel model. We will also compare our bootstrap-corrected fixed effects results to various GMM estimates.

The rest of this paper is organized as follows. Section 2 outlines the theoretical model and elaborates how we obtain our empirical specification. Section 3 deals with data choice and collection and presents summary statistics. Section 4 gives a brief overview of the issues regarding the estimation of dynamic panel data and the estimation method we will use. Section 5 examines the estimation results. Section 6 concludes and formulates some directions of future research.

2. Theoretical model and empirical specification

The theoretical framework of this paper is based on a model developed by Hatton (1995) to explain the determinants of migration into the UK. While Hatton (1995) estimates a pure time series model, we prefer to work with panel data which provides us with much more information. Two important features characterize the model. It explicitly takes into account uncertainty in the migration decision and it makes an specific assumption about the formation of expectations regarding future income streams which will be the basis for the migration decision. The model is microeconomic by nature, attempting to explain the probability that an individual chooses to migrate. This probability depends on the difference in expected utility streams (EU) in the origin country o and the destination country d. For a given individual i in a given year t, this difference is written as:

$$d_{it} = EU(y_{dt}) - EU(y_{ot}) + z_{it} \tag{1}$$

with y_t the income and z_{it} the individual's non-pecuniary utility difference between the two locations. According to Hatton (1995), this may also include the cost of migration. The individual's utility is concave $(U'' < 0)$, and specifically given by $U(y_t) = \ln(y_t)$. Hence, the probability of migration can be written as:

$$d_{it} = E\ln(y_{dt}) - E\ln(y_{ot}) + z_{it}. \tag{2}$$

Expanding $E\ln(y_{dt})$ in a Taylor series around Ey_{dt} gives:

$$\begin{aligned} E\ln(y_{dt}) &= \ln(Ey_{dt}) + \frac{1}{Ey_{dt}} E(y_{dt} - Ey_{dt}) - \frac{1}{2(Ey_{dt})^2} E(y_{dt} - Ey_{dt})^2 \\ E\ln(y_{dt}) &= \ln(Ey_{dt}) - \frac{\operatorname{var}(y_{dt})}{2(Ey_{dt})^2}. \end{aligned} \tag{3}$$

Hatton (1995) follows Todaro (1969) in defining $Ey_t = w_t e_t$ where w_t represents the wage and e_t the employment at time t. Utility is determined not only by wages, but also by future employment prospects in the home and destination country. Hence, we can write:

$$\begin{aligned} E\ln(y_{dt}) &= \eta_1 \ln w_{dt} + \eta_2 \ln e_{dt} \\ E\ln(y_{ot}) &= \eta_3 \ln w_{ot} + \eta_4 \ln e_{ot}. \end{aligned} \tag{4}$$

Substituting (4) in equation (3) of the migration probability of individual i at time t gives:

$$d_{it} = \eta_1 \ln w_{dt} + \eta_2 \ln e_{dt} - \eta_3 \ln w_{ot} + \eta_4 \ln e_{ot} + z_{it}. \tag{5}$$

However, the migration decision depends not only on current expected utility but also on future values of the stream of expected utility at home and abroad. Even if the net present value of migrating this year is positive, it might be even higher next year, which makes it interesting for the individual to wait a year. The net present value of the difference in utility streams from $t+1$ on, viewed at time t is denoted d_{it}^{*}. Consequently, the total net present value of migrating today is $d_{it}^{*} + d_{it}$. However, it pays to wait if $d_{it}^{*} > d_{it}^{*} + d_{it}$. Hence, the probability of migrating at time t $\left(m_{it} = 1\right)$ is determined by

$$\Pr(m_{it} = 1) = \Pr(d_{it}^{*} + d_{it} > 0 \cap d_{it} > 0). \tag{6}$$

To capture this, Hatton writes the function for the aggregate migration as

$$M_{odt} = \beta\left(d_t^{*} + \rho d_t\right) = \beta d_t^{*} + \beta\rho d_t \tag{7}$$

where ρ >1 reflects the extra weight given to current conditions, given that potential emigrants could choose to wait if $d_{it} < 0$. Furthermore, it is assumed that the expectations of future utility streams are based on past values of information, namely $d_{t-1}, d_{t-2}, \ldots$ More specifically, expectations are formed by a geometric series of past values of d_t, such that

$$d_t^{*} = \lambda d_t + \lambda^2 d_{t-1} + \lambda^3 d_{t-2} + \lambda^4 d_{t-3} \ldots \tag{8}$$

This would be equivalent to rational expectations if d_t follows an $AR(1)$ process. Following Hatton (1995), we assume that it does. Therefore $d_t = \varsigma d_{t-1} + e_t$, $|\varsigma| < 1$ and $e \sim \mathrm{AR}(0)$. Substituting (8) in (7) and using a Koyck transformation yields

$$M_{odt} = \beta\left(\rho + \lambda\right) d_t - \lambda\beta\rho d_{t-1} + \lambda M_{t-1}. \tag{9}$$

Substituting for d_t gives

$$\begin{aligned} M_{odt} = \beta\left(\rho+\lambda\right)&\left[\eta_1 \ln w_{dt} + \eta_2 \ln e_{dt} - \eta_3 \ln w_{ot} + \eta_4 \ln e_{ot} + z_t\right] \\ &- \lambda\beta\rho\left[\eta_1 \ln w_{dt-1} + \eta_2 \ln e_{dt-1} - \eta_3 \ln w_{ot-1} + \eta_4 \ln e_{ot-1} + z_{t-1}\right] \\ &+ \lambda M_{t-1}. \end{aligned} \tag{10}$$

Hatton (1995) assumes that $\overline{z}$, the mean of z_t over all t is determined by the stock of previous migrants and a time trend. As already mentioned, z_i reflects the individual's non-pecuniary utility difference between the two countries which may also include the cost of migration. The interpretation is as follows. The higher the stock of previous immigrants (MST_{odt}), the lower the costs of migration. This is the idea of network effects: relatives and friends may lower monetary (through useful information or temporary shelter at arrival) and psychological (by easing the feeling of abandoning its roots) costs of migration. The time trend (*t*) serves as a proxy for variations in the costs of migration. They are expected to decrease over time in accordance with the decline of transportation and communication costs. Thus we expect migration flows to be positively influenced by MST_{odt} and *t*. Consequently, $\overline{z}$ can be written as:

$$\overline{z} = \mu_0 + \mu_1 MST_{odt} + \mu_2 t \tag{11}$$

where MST_{odt} is the migrant stock at the beginning of *t* and t a linear trend.

The stock of migration diminishes at a rate δ due to deaths and remigration but on the other hand, more immigrants positively affect the stock again so that

$$MST_{odt} = (1-\delta) MST_{odt-1} + M_{odt-1}. \tag{12}$$

While Hatton (1995) uses this relationship to eliminate MST_{odt-1}, we will use it in a later stage to take into account the relationship between migrant flows and stocks. After rearranging the specification becomes a simple first-order error correction mechanism

$$\begin{aligned}\Delta M_{odt} &= \alpha_0 + s_1 \Delta \ln e_{ot} + s_2 \Delta \ln e_{dt} + s_3 \Delta \ln w_{ot} + s_4 \Delta \ln w_{dt} + c_1 \Delta MST_{odt} \\ &\quad + s_5 \ln e_{ot-1} + s_6 \ln e_{dt-1} + s_7 \ln w_{ot-1} + s_8 \ln w_{dt-1} + c_2 MST_{odt-1} \\ &\quad + c_3 t + (\lambda - 1) M_{odt-1} + \varepsilon_{odt}\end{aligned} \tag{13}$$

where $\alpha_0 = \beta(\rho + \lambda - \lambda\rho)\mu_0 + \lambda\beta\rho\mu_2$, $s_1 = \beta(\rho+\lambda)\eta_1, \ldots, s_4 = \beta(\rho+\lambda)\eta_4$, $s_5 = \beta(\rho+\lambda-\lambda\rho)\eta_1$, …, $s_8 = \beta(\rho+\lambda-\lambda\rho)\eta_4$, $c_1 = \beta(\rho+\lambda)\mu_1$, $c_2 = \beta(\rho+\lambda-\lambda\rho)\mu_1$ and $c_3 = \beta(\rho+\lambda-\lambda\rho)\mu_2$.

Transforming this time series model into a panel data model requires the introduction of possible country-specific and bilateral effects. These country-specific effects allow us to control for characteristics of the sending and receiving countries which do not change over time; think of political instability, political rights, relative freedom and civil liberties. With the inclusion of bilateral effects, we account for instance for distance between home and destination country, common languages, a common border and colonial ties. In their 2003 paper, Egger and Pfaffermayr state that the appropriate specification for a three-way panel model should include a time dummy as well as both country-specific as bilateral fixed effects. This general specification is however identical to the model that includes only a time dummy and bilateral effects.[2] Therefore, we insert only α_{od} into the empirical specification.

Another important issue that comes along with panel data estimation lies in the influence of the size of origin and destination countries. It is likely that bigger origin countries will generate relatively larger migration flows and that these migration flows will be higher in big destination countries compared to smaller ones. To illustrate this, imagine the following. If we would find that an increase in the employment rate in the destination country leads to a higher inflow of immigrants, then we are in fact saying that this increase would be the same in for instance Belgium and Germany. We could however expect that more people would go to Germany than to Belgium because of the larger size and therefore greater possibilities in Germany. It is also more likely to find higher inflows from a country like Russia compared to Iceland. One possibility is to divide the migration flow by the population in the origin country. Then you account for the size of the origin country and the fixed effects could capture the fact that migration is likely to be higher to larger countries (see e.g. Hatton, 1995 and Fertig, 2001). However, this still doesn't entirely solve the problem. Since we are working in a three-dimensional panel, we should also divide migrant flows by the population in the destination country. There is however a better alternative, i.e. taking the natural log of migrant flows and stocks, which immediately removes all problems of scale.[3] Rewriting (7) as

$$\ln M_{odt} = \beta\left(d_t^* + \rho d_t\right) = \beta d_t^* + \beta\rho d_t \tag{14}$$

and the expression for $\overline{z}$ as

$$\overline{z} = \mu_0 + \mu_1 \ln MST_{odt} + \mu_2 t \tag{15}$$

gives a model in double logarithmic form which will be tested below.

[2] For the proof of this statement, Egger and Pfaffermayr (2003) refer to Christensen (1987).

[3] Hatton (1995) emphasize that his original specification is only one in many possible functional forms that could be developed. As a matter of fact, Hatton (1995) founds that using the ln of both the migration flow and the migration stock gives very similar results. We will estimate both the fully logarithmic as the semi-logarithmic model.

Taking these modifications into account, adding $\ln M_{odt-1}$ to both sides of equation (13) and switching from employment (e_t) to unemployment (u_t) as an explanatory variable[4], the final specification becomes

$$\begin{aligned}\ln M_{odt} = {} & \alpha_0 + s_1\Delta\ln u_{ot} + s_2\Delta\ln u_{dt} + s_3\Delta\ln w_{ot} + s_4\Delta\ln w_{dt} + c_1\Delta\ln MST_{odt} \\ & + s_5 \ln u_{ot-1} + s_6 \ln u_{dt-1} + s_7 \ln w_{ot-1} + s_8 \ln w_{dt-1} + c_2 \ln MST_{odt-1} \\ & + c_3 t \; + \lambda \ln M_{odt-1} + \alpha_{od} + \varepsilon_{odt}.\end{aligned} \tag{16}$$

All in all, we have 8 supply determinants (the logs of unemployment and wages in origin and destination country, both in first-differences and in lags), 3 cost determinants of migration (the differenced and lagged log of the migration stock and a time variable) and a determinant reflecting the dynamic process (the log of lagged migration).

This model has two important features. First, both the changes and the levels of the explanatory variables enter the estimation equation separately. This makes it possible to distinguish between short-run and long-run determinants of migration. Second, both the lagged dependent variable and the migrant stock are included in the estimation equation. Unfortunately, the migrant stock can be interpreted in several ways. Although migration stock traditionally enters the estimation equation as a proxy for network effects, it could also be interpreted as reflecting a partial adjustment mechanism (see e.g. Laber, 1972; Dunlevy and Gemery, 1977). In the first sense, we would expect a positive sign, indicating that having friends and relatives already living in the destination country makes migration easier. The latter interpretation however suggests a negative sign of the migrant stock coefficient to prevent migrant inflows to be ever increasing in the future. This implies that getting closer to the equilibrium stock of migrants in the destination country, should reduce the size of migration flows. The problem is that network effects and the adjustment process are not separately identified from the parameter of migrant stocks.

Moreover, as Laber (1972) points out, since the migration stock is the sum of all past migration flows less deaths and return migration, it is itself a function of all those factors which influenced the earlier migration flows. Therefore it will be correlated with all the explanatory variables. However, multicollinearity is no reason to omit the migrant stock variable as this may result in a specification bias as well as in a loss of information regarding the network effect. Although most authors recognize the problem of using migrant stocks as a proxy for network effects, there has not been found an alternative to our knowledge. In this paper we address the problem of the joint inclusion of migration flows and stocks by explicitly taking into account the dynamic relationship between the two. As

[4] The reason for this switch is simply that more data are available for unemployment rates than employment rates.

Hatton (1995) remarked, only by recognising this endogeneity of the migrant stock, the long run steady state could be found for the migration rate and the migrant stock simultaneously.

3. Data and summary statistics

Most research on migration has focused on migration flows into one single country, mostly the United States, the United Kingdom and Germany. As already mentioned, one of our contributions is to estimate the determinants of migration into the European Union. Therefore we limit the set of destination countries to members of the European Union. The next step in the selection of settler countries was a matter of data availability. The subset for which the largest dataset on migrant inflow and stock could be constructed, is formed by Belgium, Denmark, Finland, Germany, the Netherlands and Spain.

With respect to the origin countries, all of them who were or became a member of the European Union during the sample period were excluded. This is because the free movement between member states is a totally different regime. We can expect that the regime under which migration takes place will ultimately have a considerable impact on the factors driving people to migrate. Subsequently, a selection was made based on data availability. We are left with 57 origin countries which are spread all over the world, i.e. in 6 continents.[5] In 2005, migrant inflows into our six destination countries from our sample of origin countries account for 62% of total inflows.

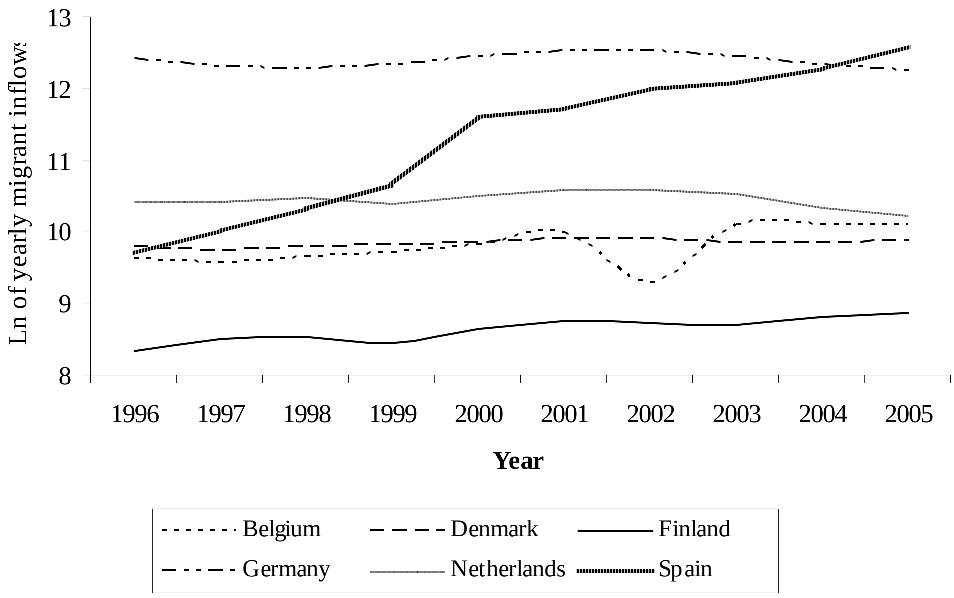
Migration data were kindly provided by the National Institute for Statistics (NIS) and Federal Government Service Economics (FOD) for Belgium, Statistics Denmark, Statistics Finland, the German Federal Statistical Office (SBD), the Central Bureau for Statistics of the Netherlands (CBS), Anuario Estadístico de Immigración 2006 and the Spanish National Statistics Institute (INE). For Denmark, Germany and the Netherlands, the data reveal migration inflows and migrant stocks according to country of origin, while for Belgium it concerns data on inflows and stocks according to nationality. For Finland and Spain, migrant stocks are reported by country of origin while inflow data are broken down by nationality. Caution is necessary in working with these data according to nationality, since naturalizations may give the impression that immigration stock has lowered, while these former foreign people are now considered to be domestic citizens. Appendix table A.1. reveals the average inflows into each destination country, by country of origin (1996-2005). For Belgium and the Netherlands, most migrants come from Morocco, the United States and Turkey. The same holds

[5] The origin countries can be found in Africa (Algeria, Egypt, Mauritius, Morocco, South Africa and Tunisia), Asia (China, Georgia, Hong Kong, Indonesia, Iran, Israel, Japan, Kazakhstan, Kyrgyzstan, Macao, Malaysia, Moldova, Pakistan, Philippines, Singapore, South Korea, Sri Lanka, Thailand and Vietnam), Europe (Bulgaria, Iceland, Macedonia, Norway, Romania, Russia, Switzerland, Turkey and Ukraine), North America (Belize, Canada, Costa Rica, Dominican Republic, El Salvador, Honduras, Jamaica, Mexico, Panama, St. Lucia, Trinidad & Tobago and the United States), South America (Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Peru, Uruguay and Venezuela) and Oceania (Australia and New Zealand). For a list of the origin countries by destination country see appendix table A.1.

for Germany, except that Romania replaces Morocco in the top 3 of German origin countries. In Spain too, most migrants come from Morocco and Romania. Finland mostly attracts Norwegians and Finland especially Russians. The average inflows summed over all origin countries in Germany is the largest, followed by Spain; it is about 7 times the size of migrant inflows into the Netherlands, around 13 times those in Belgium and Denmark and over 42 times the inflows of Finland.

Appendix table A.2. illustrates total yearly migration inflows into each destination country (1996-2005) and table A.3. shows yearly migration inflows into each destination country for our sample of origin countries (1996-2005). The latter is also demonstrated in Figure 1.

Figure 1: Yearly migration inflows into Belgium, Denmark, Finland, Germany, the Netherlands and Spain from our sample of origin countries – Period 1996-2005



Source: Data on migration inflows were kindly provided by the National Institute for Statistics (NIS) and Federal Government Service Economics (FOD) for Belgium, Statistics Denmark, Statistics Finland, the German Federal Statistical Office (SBD), the Central Bureau for Statistics for the Netherlands (CBS) and the Spanish National Statistics Institute (INE).

The dynamics in total yearly inflows and yearly inflows from the origin countries under study are more or less the same for Denmark, Germany and the Netherlands. Although the absolute number of migrant inflows considerably differs between these countries, migration flows are slightly below average until 2000 but then they resume and finally end up at more or less the same level as in 1996. The inflows were the highest in Germany, followed by the Netherlands and the smallest in Denmark. Figure 1 also reveals that in Belgium, Finland and Spain, migration inflows, have increased over the time period considered. This is not the result of a single smooth movement. Migration inflows in Belgium are U-shaped until 2002, when they are almost cut in half. It has to be said that this is not the result of an extraordinary high amount of naturalizations in 2001. After 2002, migration flows increase again and end at an even higher level than in 1996. In Spain, migration flows are initially the same size as in Belgium. They are sharply rising until 2000 and then keep on growing though at a slower rate until they end up at a much higher level than in Belgium. Finland's migration inflows, finally, reveal kind of the same pattern as in Denmark, the only difference being that they end up slightly higher than their starting point. Finland has the lowest migration inflows of all countries in our destination sample.

Due to a lack of real wage data for the set of origin countries, wages are approximated by per capita GDP (see also Fertig, 2001 and Mayda, 2005); in purchasing power parities (ppp) in constant 2000 international dollars to account for differences in living costs between the home and sending countries. Data on GDP per capita comes from the World Development Indicators 2006 from the World Bank. For reasons of data availability, we use unemployment figures instead of employment rates. Figures regarding unemployment were gathered from the World Development Indicators 2007 from the World Bank, National Labor Force Surveys and the Laborsta database from the International Labor Organization.

Table 1 presents the summary statistics for the variables used in the regression analysis.

Table 1: Descriptive statistics

Variable	**Mean**	**St.Dev**	**Min**	**Max**
Log of migrant inflows	5.53	2.55	-6.91	11.45
Lagged log of inflows	5.47	2.59	-6.91	11.22
Migrant inflows divided by thousands of origin population	0.07	0.14	0.00	2.33
Δ log of unemployment rate in origin country	0.00	0.16	-0.96	1.37
Δ log of unemployment rate in destination country	-0.03	0.12	-0.27	0.32
Δ log of per capita GDP ppp in origin country	0.44	0.04	-0.14	0.27
Δ log of per capita GDP ppp in destination country	0.04	0.01	0.01	0.07
Lagged log of unemployment rate in origin country	2.04	0.65	-0.14	3.62
Lagged log of unemployment rate in destination country	1.99	0.44	1.03	3.10
Lagged log of per capita GDP ppp in origin country	8.91	0.84	7.10	10.59
Lagged log of per capita GDP ppp in destination country	10.17	0.12	0.79	10.38
Δ log of migrant stocks	0.05	0.12	-1.10	0.07
Lagged log of migrant stocks	7.53	2.34	0.00	14.56
Migrant stocks/10 000 000	0.00	0.01	0.00	0.21
Time	2001	2.58	1997	2005

Number of observations: 2061. Number of cross-sections: 229. Period: 1997-2005.

4. Estimation method

The model that has been developed above is a dynamic panel model, which allows us to investigate the influence of past behaviour on current behaviour. So far, not many panel data studies on the determinants of migration have been performed, which is mostly due to the lack of data availability. Only recently, countries started to keep track of migration flows in great detail. Hatton (1995) estimated a time series model for the migration determinants into the UK. Fertig (2001) took the model to a higher level by estimating a two-dimensional panel data model for the migration determinants in Germany. Using repeated observations on the same units makes the estimation more realistic than a single time series or a single cross-section. We will go one step further by investigating the factors driving migration in more than one destination country, i.e. a three-dimensional panel data model.

A dynamic panel model is often estimated using the fixed effects (FE) estimator, also referred to as the within or least-squares dummy variable (LSDV) estimator (see e.g. Mitchell and Pain[6], 2003; Clark et al., 2002). The fixed effects estimator calculates the parameter values using only the variation in the explanatory variables through time. Nevertheless, when the time dimension (*T*) is small, the estimated coefficient of the lagged dependent variable $\ln M_{odt-1}$, i.e. λ, in an AR(1) model will be downward biased because of the negative correlation between $\ln M_{odt-1}$ and ε_{odt}. Therefore, the FE estimate of λ is often seen as the lower bound on the true parameter.[7] To obtain the long run coefficients, however, we need to divide the short-run parameters by $1-\lambda$, which is problematic since this would result in a bias for all long-run parameters.[8] Moreover, due to the dynamic relationship between migrant flows and stocks, the latter will also be negatively correlated with the error term. This implies that the endogeneity bias is now caused by two explanatory variables, rather than one.

Given the inconsistency of the fixed effects estimator for a fixed time period various instrumental variables (IV) and generalized method of moments (GMM) estimators have been proposed. The first-difference GMM estimator of Arellano and Bond (1991) and the system GMM estimator of Arellano and Bover (1995) and Blundell and Bond (1998) have been widely used to take into account endogeneity problems (see e.g. Mayda, 2005 who uses GMM to estimate the migration determinants into the OECD during 1980-1995). These estimators are designed for dynamic panels with a large cross-sectional dimension (*N*) and a small *T* (see Arellano and Bond, 1991; Blundell and Bond, 1998). Since in our sample *T=10* which is not that small, the standard uncollapsed instrument matrix is very large. A finite sample may however result in an overfitting bias of the difference and system GMM

[6] Mitchell and Pain (2003) estimate the determinants of international migration into the UK during 1980-2000 building on Hatton's 1995 model, and using panel data. Their starting point is an autoregressive distributed lag (ADRL) model, of which the fixed effects estimator is a restricted form since it imposes equality of all slope coefficients and all error variances. However, this approach is only appropriate when *T* is large. In this case separate time series regressions can be estimated for each group.

[7] Estimating the model using the ordinary least squares (OLS) estimator (see e.g. Hatton, 1995 for his time series analysis of migration determinants into Germany during 1870-1913) would result in an upward bias on the estimated coefficient of $\ln M_{odt-1}$ because of the positive correlation between $\ln M_{odt-1}$ and the country-specific effects α_{od}. Therefore the estimated parameter of the lagged dependent variable is often seen as an upper bound on the true parameter. Performing both the FE and the OLS estimation results in an indication of the interval in which the true parameter should be found.

[8] This paper does not yet contain estimates for the long run relationship. They will however be taken up in the next version of the paper.

estimators (see Zilliak, 1997). To limit the instrument count, we perform both difference and system GMM estimations, once stacking the instrument set and once restricting the lag range to two (the 2nd and 3rd lag). Stacking the instrument set implies taking linear combinations of the moment conditions, i.e. summing over T which considerably reduces the column dimension of the instrument matrix. While this diminishes efficiency, it avoids biases caused by the use of too many moment conditions (see Everaert and Pozzi, 2007).

As linear GMM estimators, the Arellano-Bond and Blundell-Bond estimators have one- and two-step variants. But although two-step is asymptotically more efficient, the two-step standard errors tend to be severely downward biased (see Arellano and Bond 1991; Blundell and Bond 1998). To compensate, we use a finite-sample correction to the two-step covariance matrix derived by Windmeijer (2005). This makes the two-step GMM estimator more efficient than the one-step variant, especially for system GMM. Therefore we only report two-step results.

Wages and unemployment are assumed to be strictly exogenous explanatory variables as it seems very unlikely that migrant flows in year t have an immediate impact on wages and unemployment. Therefore we include them as standard instruments (iv-style). This implies that only lagged flows and the differenced migrant stock need to be instrumented (gmm-style).

We report the Arellano-Bond test for autocorrelation, which is applied to the differenced residuals so as to remove the unobserved and perfectly autocorrelated fixed effects. We can expect AR(1) in first differences since the differenced error term correlated with it's lag due to the common term ε_{odt-1}. So to check for AR(1) in levels, we should look for AR(2) in differences. Autocorrelation indicates that lags of the dependent variable (and any other variables used as instruments that are not strictly exogenous), are in fact endogenous, thus bad instruments (see Roodman, 2006). The Hansen test of over-identifying restrictions tells us if the instrument variables are valid, i.e. if their correlation with the errors is close enough to zero. Finally, the difference-in-Hansen test analyses on the one hand the joint validity of the additional instruments in gmm-style for the levels equation which must be valid for system GMM to be consistent and on the other hand the validity of the instrument variables in iv-style.

However, the IV and GMM estimators are not without problems either. First, Monte Carlo simulations (see e.g. Arellano and Bond, 1991) show that the GMM estimators have a relatively large standard error in comparison to the standard FE estimator. Second, IV and GMM estimators require additional decisions on which instruments to use. Finally, GMM functions perform badly when the initial conditions are not stationary. And although it is hard to find out if this is true from a sample period of 10 years, the evolution of migration flows may very well be an adjustment process towards equilibrium. In this case, the migrant stock would have a negative influence on migration flows indicating that immigration towards our destination countries will not be ever increasing in the future.

In order to avoid these problems analytical corrections for the FE estimator have been proposed. Kiviet (1995) derives a formula for the small sample bias of the FE estimator in a first-order dynamic panel model. Then he uses this approximated error to correct the FE estimator. A simulation study reveals that this technique works noticeably accurate. Unfortunately, this estimator is derived under strict theoretical assumptions. A number of recent papers however succeeded in relaxing some of these theoretical restrictions. Bun and Carree (2005b) for example provide a bias correction for the FE estimator in the presence of cross-sectional and time-series heteroskedasticity. From the simulation results on various designs it is shown that the bias-corrected FE estimators are more efficient than GMM estimators.

In this paper however we follow Everaert and Pozzi (2007) who propose a bias correction for the FE estimator using an iterative bootstrap algorithm. Like the analytical corrections this bootstrap-correction aims at reducing the bias of the FE estimator while maintaining its higher efficiency compared to GMM estimators. Three features of this bootstrap approach should be mentioned. First, it does not rely on any theoretical assumptions. Second, it is easier to implement than the analytical corrections. Third, the long-run effects are estimated directly, unlike in the analytical correction in which the long-run estimate is based on the corrections of the short-run estimates.[9] Moreover, their Monte Carlo results reveal that the bootstrap-corrected FE estimator is comparable to the analytical corrections in terms of estimation but yields better inference in samples with a small to moderate time dimension. Also, it outperforms GMM estimators both in terms of estimation and inference in samples where both T and N are small.[10]

The most important assumption in the correction suggested by Everaert and Pozzi (2007) is that all regressors except for the lagged dependent variable are strongly exogenous. As already mentioned, the theoretical model developed above does not violate this assumption. The only problem is that we jointly include migrant flows and stocks. But we are able to take their relationship into account at the hour of programming the bootstrap procedure.

There are two ways to perform a fixed effects estimation. One way is to include a different dummy for every couple of origin and destination country; i.e. the least squares dummy variable (LSDV) method. However, since in our case N is large (N=2061), it is undesirable to compute all the coefficients for every single cross-section. According to Johnston and Dinardo (1997) however, transforming all the variables by subtracting cross-section specific means and running OLS on these transformed variables brings about exactly the same results. This explains why the FE estimator is also called the within estimator; that is, it uses only the variation within a couple's (of origin and destination country) set of observations. Transforming all variables into matrices of dimension $T*N$,

[9] This paper does not yet contain estimates for the long run relationship. They will however be taken up in the next version of this paper.
[10] In the next version of the paper, the performance of the GMM estimator and the bootstrap-corrected FE estimator will be compared for our setting.

subtracting the mean from each column and then stacking the observations over time and over cross-sections:

$$\begin{aligned}
&\ln M_{od} = \left(\ln M_{od2},...,\ln M_{odT}\right)', \ln M_{od,-1} = \left(\ln M_{od1},...,\ln M_{odT-1}\right)', \\
&x_{od} = \left(x_{od2},...,x_{odT}\right)', \varepsilon_{od} = \left(\varepsilon_{od2},...,\varepsilon_{odT}\right)' \\
&\ln M = \left(\ln M_1',...,\ln M_N'\right)', \ln M_{-1} = \left(\ln M_{1,-1}',...,\ln M_{N,-1}'\right)', \\
&x = \left(x_1',...,x_N'\right)', \varepsilon = \left(\varepsilon_1',...,\varepsilon_N'\right)'
\end{aligned} \tag{17}$$

$$\begin{aligned}
\text{with } x_{odt} = {} & \Delta \ln u_{ot}, \Delta \ln u_{dt}, \Delta \ln w_{ot}, \Delta \ln w_{dt}, \Delta \ln MST_{odt}, \\
& \ln u_{ot-1}, \ln u_{dt-1}, \ln w_{ot-1}, \ln w_{dt-1}, \ln MST_{odt-1}, t
\end{aligned}$$

and subsequently reshaping the variables into $N(T-1)\times 1$ vectors gives:

$$\begin{aligned}
\ln M &= \gamma X + \lambda \ln M_{-1} + \varepsilon \\
&= W\vartheta + \varepsilon
\end{aligned} \tag{18}$$

where $X = \left[\Delta \ln u_{ot} \;\; \Delta \ln u_{dt} \;\; \Delta \ln w_{ot} \;\; \Delta \ln w_{dt} \;\; \Delta \ln MST_{odt} \;\; \ln u_{ot-1} \;\; \ln u_{dt-1} \;\; \ln w_{ot-1} \;\; \ln w_{dt-1} \; \ln MST_{odt-1} \; t\right]$ and $W = \left(X, \ln M_{-1}\right)$ is a $N(T-1)\times(k+1)$ matrix containing the regressors, k being the number of regressors apart from the lagged dependent variable and $\vartheta = \left(\lambda, \gamma\right)$ is a $(k+1)\times 1$ parameter vector.

We know that the estimated parameter of the lagged dependent variable $\hat{\vartheta}$ is biased. This means that $E\left(\hat{\vartheta}\right) \equiv \int_{-\infty}^{+\infty} h f_{\hat{\vartheta}}(h) dh \neq \vartheta$ where $f_{\hat{\vartheta}}(.)$ is the probability distribution of $\hat{\vartheta}$. However, having estimated the coefficients $\tilde{\vartheta}$, the constants $\tilde{\alpha}_{od}$ and the residuals $\tilde{\varepsilon}$, we are able to apply the bootstrap procedure to obtain a bias-corrected FE estimator. We perform 1000 bootstrap samples. In the first bootstrap sample *j*, we resample the residuals within the cross-section identity in the same way for each identity, which we will call $\tilde{\varepsilon}^b$. Subsequently, we use these resampled residuals to calculate $\ln M^b = W^b \tilde{\vartheta} + \tilde{\varepsilon}^b$ with initialization $\ln M_{i1}^b = \ln M_{i1}$, so we initialize $\ln M^b$ with the initial values of $\ln M$. Next, we use the newly constructed variables to estimate the FE estimator $\tilde{\vartheta}^b = \left(\tilde{\lambda}^b, \tilde{\gamma}^b\right)$. Repeating this for 1000 bootstrap samples and calculating the mean of the FE estimator over these 1000 samples eventually allows us to evaluate $\tilde{\vartheta}$ as an estimator of ϑ. For $\tilde{\vartheta}$ to be an unbiased estimator for ϑ, the mean $\tilde{\vartheta}^b$ of the bootstrap distribution of $\tilde{\vartheta}$ should equal the original biased FE estimates $\hat{\vartheta}$, i.e. $\omega = \tilde{\vartheta} - \tilde{\vartheta}^b = 0$. In order to find a parameter vector $\tilde{\vartheta}$ that satisfies this condition we iterate over the bootstrap procedure outlined above and evaluate $\tilde{\vartheta}_{(h)}$, in each iteration *h*, as an estimator for ϑ by calculating $\omega_{(h)} = \tilde{\vartheta}_{(h)} - \tilde{\vartheta}^b_{(h)}$. If $\omega_{(h)} = 0$, $\tilde{\vartheta}_{(h)}$ is taken to be the unbiased estimate for ϑ. If $\omega_{(h)} \neq 0$, $\tilde{\vartheta}_{(h)}$ is updated as $\tilde{\vartheta}_{(h+1)} = \tilde{\vartheta}_{(h)} + \omega_{(h)}$ and we iterate over the bootstrap procedure until this condition is satisfied. As the biased FE estimator $\hat{\vartheta}$ can be thought of as being our first guess for the vector of population parameters ϑ, we initialize the algorithm by setting $\tilde{\vartheta}_{(1)} = \tilde{\vartheta}$.

Now we have drawn one sample out of the population. What we actually want is to draw a much larger amount of samples from the population, say 1000, and then calculate the parameter distribution. To do this, we generate data using the resampled residuals as explained above and perform a bootstrap procedure for each of these iterations. In the end, we can calculate the corrected t-statistics using the corrected coefficients as well as the corrected variance.

5. Estimation results

Table 3 presents the FE estimates of the migration determinants into the EU for the period 1996-2005. The first column of Table 3 presents the standard FE estimation results. All supply parameters show the expected sign except for wages in the origin country. This finding is not unusual and is often explained as the consequence of poverty constraints, indicating that one needs to have a certain income to be able to move abroad (see e.g. Mayda, 2005). Wages in the destination country have highly significant and positive coefficients and so has the lagged unemployment in the origin country. First differenced unemployment in the destination country has a negative coefficient and is significant at the 90% confidence interval. These results indicate that the supply factors play an important role in explaining migration inflows into the European Union around the change of the century. Also the coefficients for the migrant stock, both differenced as lagged, are highly significant and suggest an elasticity of respectively 2.19 and 0.83. Unfortunately, it is impossible to separate the effect of adjustment towards equilibrium (implying a negative coefficient on stocks) and the impact of the network effects (which involves a positive coefficient). As explained above, this means that it is impossible to distinguish between moving quickly towards equilibrium while being tempered by network effects and approaching the equilibrium more slowly with small network effects. Time seems to be working against migration flows and so are passed migration flows. The parameter for the latter though insignificant suggests an elasticity of -0.04. Nevertheless, we know that this estimate is biased and as such considered as a lower bound on the true parameter of lagged migration flows.[11] Based on the interpretation of time as a proxy for decreasing migration costs, this is not as we expected.

[11] Estimating this relationship using OLS results in a parameter value of 0.66, which can be considered as the lower bound on the true parameter. As Roodman (2006) indicates, good estimates of the true parameter should therefore lie in the range between these values – or at least near it, given that these numbers are themselves point estimates with associated confidence intervals. He refers to Bond (2002) who points out that this provides a useful check on results from theoretically superior estimates.

However, this result may reflect the fact that migration policy in these destination countries is becoming more restrictive over time.

An alternative set of estimates using the migration flows divided by the population in the sending country and migrant flows in absolute numbers is given in the appendix table A.4. Contrary to Hatton's (1995) finding of very similar results for his time-series analysis, we obtain very different results. Regarding the supply factors, only unemployment rates are significant with the expected sign. There is no indication for network effects. The coefficient for the lagged migrant flows though is highly significant and positive. These divergent estimates indicate that in a three-way panel model, dividing the migrant flows by the size of the population in the sending country does not entirely eliminate scale effects.

Table 3: FE estimates of migration determinants into the EU 1996-2005 (Dependent variable: $\ln M_{odt}$)

$\ln M_{odt}$	**Fixed Effects**	**Fixed Effects with Bootstrap correction**	**Fixed Effects with Bootstrap correction & link between flows-stocks**
$\Delta \ln w_{dt}$	3.56** (2.01)	3.46* (1.79)	3.51* (1.68)
$\Delta \ln u_{dt}$	-0.39* (-1.89)	-0.40 (-1.24)	-0.41 (-1.22)
$\Delta \ln w_{ot}$	0.07 (0.16)	0.01 (0.02)	-0.10 (-0.17)
$\Delta \ln u_{ot}$	0.13 (1.62)	0.13* (1.77)	0.13 (1.64)
$\ln w_{dt-1}$	5.31*** (3.37)	4.60** (2.29)	4.45** (2.35)
$\ln u_{dt-1}$	-0.21 (-1.39)	-0.20 (-1.15)	-0.19 (-1.06)
$\ln w_{ot-1}$	0.31 (1.05)	0.27 (0.90)	0.19 (0.66)
$\ln u_{ot-1}$	0.18** (2.22)	0.16 (1.35)	0.15 (1.34)
$\Delta \ln MST_{odt}$	2.19*** (4.95)	2.11*** (4.53)	2.29*** (11.07)
$\ln MST_{odt-1}$	0.83*** (4.20)	0.74*** (3.92)	0.86*** (4.00)
Time	-0.20*** (-3.35)	-0.17** (-2.02)	-0.17** (-2.15)
$\ln M_{odt-1}$	-0.04 (-0.33)	0.07 (0.38)	0.06 (0.33)
Constant	339.78*** (3.31)		

Note: t-statistics between brackets. Standard errors are robust to heteroskedasticity and clustered by origin country, both in the standard as in the corrected FE estimation. Number of observations = 2061, number of groups = 229. * Significant at the 90% confidence interval. ** Significant at the 95% confidence interval. *** Significant at the 99% confidence interval.

The second column of Table 3 reports the results of the fixed effects estimator with bias-correction using the bootstrap algorithm. We see that the differenced wage in the destination country becomes slightly less significant while differenced unemployment in the sending country now becomes marginally significant and positive. The coefficients for the migrant stock remain highly significant but fall somewhat to 2.11 for differenced and 0.74 for lagged stocks. And as we expected, the elasticity for the lagged dependent variable though still insignificant now rises to 0.07.

The estimates reported in the final column of Table 3 are fixed effects estimates, corrected with a bootstrap procedure which also takes into account the relationship between migrant flows and stocks. It turns out that only wages in the destination country have positive and significant coefficients, indicating that these are the only important supply factors driving migration. The elasticities for differenced and lagged wages in de receiving country are 3.51 and 4.45 respectively. The coefficient for the differenced migrant stock, i.e. 2.29, becomes even more significant; differenced migrant stocks therefore end up as the most significant explanatory variable. Also the coefficient for lagged migrant stocks is highly significant and suggests an elasticity of 0.86. Time on the other hand negatively affects migrant flows. The coefficient on lagged migration flows is still positive, i.e. 0.06 but remains insignificant.

Although we know that GMM estimators are not without problems, let's still have a look at their estimation results. Appendix table A.5 presents the results of the two-step difference (GMMd) and two-step system GMM (GMMs) estimations, once stacking the instrument set (s) and once restricting the lag range to two (the 2nd and 3rd lag). It becomes immediately clear that the results strongly depend on the estimation method. Estimating the migration determinants with (s)GMMd returns the same parameter signs as GMMd but attaches much more significance to them. The same holds for (s)GMMs and GMMs. While GMMd only assigns significance to wages in the sending country and time, (s)GMMd also stresses the importance of migrant stocks. With respect to the system GMM estimations, the positive significance of lagged migrant flows seems to be the most robust. Migrant stocks are again only significant in the stacked version of the system GMM estimator. Although insignificant, the parameter signs considerably differ between the difference GMM and system GMM estimators. This indicates that the results are not at all robust to the choice of the estimation method.

While the GMMd estimations show no AR(1), in the GMMs estimations there is AR(1) as we expected. From the absence of AR(2) in differenced residuals, we know that there is no AR(1) in their levels. This indicates that the instruments for the lagged dependent variable and the differenced stock are not endogenous and thus good instruments. The Hansen test of over-identifying restrictions however reveals that the instrument variables are not valid as a group, i.e. the correlation between the instrument variables and the errors is not close enough to zero.[12] The only exception is the (s)GMMd

[12] This is also the case for GMM estimations which use the 3^{rd} and 4^{th} lag or the 4^{th} and 5^{th} lag of the gmm-style variables in levels. However, it should be mentioned that the Hansen test of over-identifying restrictions is weak when instruments are many.

estimation. The difference-in-Hansen test reveals that the additional instruments in gmm-style for the levels equation are jointly valid only for the (s)GMMs estimator. This means that the stacked system GMM estimator is consistent but the one using a lag constraint is not. None of the iv-style instruments seem to be valid. As an alternative, we included the differenced and lagged wage in the destination country in gmm-style instead of iv-style, treating them as possibly endogeneous. The estimates are reported in appendix table A.6. For the difference GMM estimations, this renders valid instruments according to both the Hansen test of over-identifying restrictions and the difference-in-Hansen test for the iv-style instruments. For the system GMM estimations however, the instruments remain invalid.

We already mentioned above that one of the disadvantages of working with a GMM estimator is that additional decisions need to be made on which instruments to use. These decisions have shown to have a great impact on the outcome. We come to the conclusion that the GMM-estimates greatly depend on the choice of the instrument set and its lag-structure, and on the choice between the difference and system GMM estimator. Working with a bootstrap-corrected FE estimator makes all of these choices superfluous.

Comparing the (s)GMMd estimates to our bootstrap-corrected FE estimates reported in the second column of table 3, we see that the same variables are found to be significant. However, contrary to the corrected FE estimator, the (s)GMMd shows a positive impact of the differenced unemployment in the destination country (against expectations), and a negative impact of the wage in the origin country, which is more in line with theoretical predictions than what we found with the corrected FE estimator. It also reveals that lagged migrant flows though insignificant negatively affect current inflows. The same holds for the GMMd estimator. The results obtained from the system GMM estimations differ even more from our corrected FE results. Apart from migrant stocks, differenced unemployment in the receiving country, lagged wages in the origin countries and the lagged migrant flows, all variables show the opposite sign. The (s)GMMs estimated coefficient of lagged migrant flows suggests a significant elasticity of 0.28, which is a lot bigger than the 0.06 suggested by the corrected FE. The opposite is true for the (s)GMMs estimated parameter for differenced stocks, which is now only 1.65 compared to 2.11 with the corrected FE estimator. This finding is even more evident from the comparison of GMMs with the corrected FE estimator. With GMMs, the coefficient for lagged migrant flows becomes 0.66[13] while the one for differenced migrant stocks drops to 0.32 and becomes insignificant.

It turns out that the results obtained from the bootstrap-corrected FE estimation and the GMM estimations do not match at all. For one thing, while the difference GMM estimations still puts forward a negative impact of lagged migrant flows, the system GMM estimation suggests a much higher upward correction than the bootstrap-corrected FE estimation.

[13] With this, the GMMs estimated parameter of lagged migrant flows reaches the upper bound of the true parameter, as obtained through OLS estimation.

6. Conclusion

In this paper we investigate the determinants of migration into the European Union during 1996-2005. The theoretical foundations are based on a model developed by Hatton (1995) to examine the factors driving migration in the U.K. using time series data. We slightly transform his model into a dynamic panel data model, which is often estimated by a fixed effects estimator. However, this estimator is inconsistent when the time dimension is small. The coefficient of the lagged dependent variable is downward biased. Moreover, due to the dynamic relationship between migrant flows and stocks, the latter will also be negatively correlated with the error term. This implies that the endogeneity bias is now caused by two explanatory variables, rather than one. We follow Everaert and Pozzi (2007) who propose a bias correction for the FE estimator using an iterative bootstrap algorithm.

This results in a small upward correction of the fixed effects estimator. Accounting for the dynamic relationship between migrant flows and stocks results though insignificant in a point estimate of 0.06 for the coefficient of lagged migrant flows. Migrant stocks are highly significant and accompanied by a positive sign. Unfortunately we are not able to separate the network effects from the adjustment process to equilibrium. Regarding the supply factors, only the wage in the destination country seems to be significant with a positive impact on migrant flows.

One of the disadvantages of working with a GMM estimator is that additional decisions need to be made on which instruments to use. These decisions have shown to have a great impact on the outcome. We come to the conclusion that the GMM-estimates greatly depend on the choice of the instrument set and its lag-structure, and on the choice between the difference and system GMM estimator. Working with a bootstrap-corrected FE estimator makes all of these choices superfluous.

It turns out that the estimates obtained with the bootstrap-corrected FE estimator and those with the GMM estimators do not match at all. For one thing, while the difference GMM estimates suggest a negative impact of lagged flows, the system GMM puts forward a higher upward correction than the bootstrap-corrected FE. So far, the choice between the two needs to be made based on simulation experiments in other work. In the future, we will however try to find out which of these methods is most appropriate in our setting by performing our own simulation experiment.

References

Anuario Estadístico de Inmigración, 2006, Spain, Secretaría de Estado de Inmigración e Emigración.

Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Review of Economic Studies 58, 277-297.

Arellano, M., Bover, O., 1995. Another look at the instrumental-variable estimation of error-components models. Journal of Econometrics 68, 29-52.

Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87, 115-143.

Blundell, R., Bond, S., Windmeijer, F., 2000. Estimation in dynamic panel data models: improving on the performance of the standard GMM estimator. The Institute for Fiscal Studies, Advances in Econometrics: A Research Annual 15, 53-91.

Bun, M. and Windmeijer F., 2007. The weak instrument problem of the systemGMM estimator in dynamic panel data models. Centre for microdata methods and practice, Working Paper CWP08/07.

Bun, M., Carree M., 2005a. Bias-corrected estimation in dynamic panel data models. Journal of Business & Economic Statistics 23 (2), 200-210.

Bun, M., Carree M., 2005b. Bias-corrected estimation in dynamic panel data models with heteroskedasticity. Amsterdam School of Economics, Discussion Paper 2005/03.

Central Bureau for Statistics (CBS) the Netherlands, http://www.cbs.nl, Voorburg/Heerlen.

Central Bureau for Statistics for the Netherlands (CBS), http://www.cbs.nl, Voorburg.

Clark, X., Hatton, T., Williamson, J., 2002. Where do U.S. immigrants come from, and why?. National Bureau of Economic Research, Working Paper 8998.

Dunlevy, J., Gemery, H., 1977. The role of migrant stock and lagged migration in the settlement patterns of nineteenth century immigrants. The Review of Economics and Statistics 59 (2), 137-144.

Egger, P., Pfaffermayr, M., 2003. The proper panel econometric specification of the gravity equation: a three-way model with bilateral interaction effects. Empirical Economics 28, 571-580.

Everaert, G., Pozzi, L., 2007. Bootstrap-based bias correction for dynamic panels. Journal of Economic Dynamics and Control 31, 1160-1184. The Review of Economics and Statistics 51, 189-194.

Federal Public Sector (FPS) Economy Belgium, Directorate-general Statistics and Economic Information, http://economie.fgov.be.

Federal Statistical Office Germany (Statistisches Bundesamt Deutschland), http://www.destatis.de, Wiesbaden.

Fertig, M., 2001. The economic impact of EU-enlargement: assessing the migration potential. Empirical Economics 26, 707-720.

Foders, F., 2003. Long-run determinants of immigration to Germany 1974-1999: a Ricardian framework, Kiel Institute for World Economics, Working Paper 1187.

Greenwood, M., 1969. An analysis of the determinants of geographic labor mobility in the United States.

Hatton, T., 1995. A model of UK emigration, 1870-1913. The Review of Economics and Statistics 77, 407-415.

International Labour Office Bureau of Statistics (laborsta internet), http://laborsta.ilo.org, 1996-2008.

International Monetary Fund, 1997, IMF Annual Report Tables 1997, Washington D.C.

Johnston, J., Dinardo, J., 1997, Econometric Methods, McGraw-Hill/Irwin, fourth edition.

Kiviet, J., 1995. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. Journal of Econometrics 68, 53-78.

Laber, G., 1972. Lagged response in the decision to migrate: a comment. Journal of Regional Science 12 (2), 307-310.

Levy, M., Wadycki, 1973. The influence of family and friends on geographic labor mobility: an international comparison. The Review of Economics and Statistics 55 (2), 198-203.

Mayda, A.M., 2005. International migration: a panel data analysis of economic and non-economic determinants. IZA Discussion Paper 1590.

Mitchell, J., Pain, N., 2003. The determinants of international migration into the UK: a panel based modelling approach. National Institute of Economic and Social Research UK.

National Institute for Statistics (NIS) Belgium, http://www.statbel.fgov.be, Brussels.

National Statistics Institute (INE) Spain, http://www.ine.es, Madrid.

Pirotte, A., 1999. Convergence of the static estimation toward the long run effects of dynamic panel data models. Economic Letters 63, 151-158.

Roodman, D., 2006. How to do xtabond2: an introduction to “difference” and “system” GMM in Stata. Center for global development, Working Paper 103.

Statistics Denmark, http://www.dst.dk, Copenhagen

Statistics Finland, http://www.stat.fi, Helsinki.

Statistisches Bundesamt Deutschland (SBD), http://www.destatis.de, Wiesbaden.

Verbeek M., 2004, A Guide to Modern Econometrics, Erasmus University Rotterdam, 2nd edition, 429 pp.

Windmeijer, F., 2005. A finite sample correction for the variance of linear efficient two-step gmm estimators. Journal of Econometrics 126, 25-51.

Worldbank, 2006, World Development Indicators, 2006, WDI CD-ROM 2006.

Zilliak, J., 1997. Efficient estimation with panel data when instruments are predetermined: an empirical comparison of moment-condition estimators. Journal of Business and Economic Statistics 14 (4), 419-431.

Appendix

Table A.1. Average inflows into each destination country, by country of origin (1996-2005)

Belgium (23)	
Algeria	512
Australia	190
Brazil	417
Bulgaria	372
Canada	515
Chile	118
China	947
Colombia	187
Indonesia	124
Israel	207
Japan	857
Morocco	5636
Norway	244
Pakistan	293
Peru	137
Philippines	438
Romania	861
Switzerland	182
Thailand	400
Tunisia	344
Turkey	2619
United States	2652
Vietnam	186
Total (above)	**18438**
Percentage change	**+59**

Spain (18)	
Algeria	2868
Australia	342
Brazil	6445
Bulgaria	8316
Canada	405
Chile	3159
China	6124
Dominican Republic	5485
Japan	295
Mexico	2613
Morocco	32553
Pakistan	2001
Peru	7652
Philippines	1109
Romania	29167
Russia	3428
Switzerland	3839
United States	3154
Total (above)	**118953**
Percentage change	**+1690**

Finland (43)	
Algeria	19
Australia	14
Argentina	131
Belize	0
Bolivia	11
Brazil	52
Bulgaria	32
Canada	144
Chile	18
China	358
Colombia	40
Egypt	76
Georgia	5
Iceland	47
Iran	160
Israel	64
Jamaica	2
Japan	108
Kyrgyzstan	2
Macedonia	13
Malaysia	28
Mauritius	1
Mexico	25
Moldova	10
Morocco	57
New Zealand	20
Pakistan	85
Peru	17
Philippines	59
Romania	56
Russia	2226
South Africa	34
South Korea	23
Sri Lanka	27
Switzerland	150
Thailand	305
Trinidad & Tobago	1
Tunisia	21
Turkey	353
Ukraine	123
United States	663
Venezuela	9
Vietnam	93
Total (above)	**5680**
Percentage change	**69**

Germany (48)	
Algeria	2562
Argentina	1581
Australia	2840
Bolivia	472
Brazil	5929
Bulgaria	9763
Canada	3697
Chile	1053
China	13140
Colombia	1746
Costa Rica	264
Dominican Republic	1109
Ecuador	853
Egypt	2037
El Salvador	102
Georgia	3598
Honduras	117
Iceland	293
Indonesia	1715
Iran	5845
Israel	1643
Jamaica	178
Japan	5872
Macedonia	3502
Malaysia	688
Mexico	2149
Moldova	2111
Morocco	5072
New Zealand	656
Norway	1367
Pakistan	3558
Panama	81
Peru	1319
Philippines	2179
Romania	21666
South Africa	2224
South Korea	2644
Sri Lanka	2536
Switzerland	8303
Thailand	5929
Trinidad & Tobago	66
Tunisia	2527
Turkey	52248
Ukraine	18486
United States	27170
Uruguay	193
Venezuela	779
Vietnam	5672
Total (above)	**239534**
Percentage change	**-15**

Netherlands (45)	
Algeria	261
Argentina	207
Australia	795
Bolivia	54
Brazil	853
Bulgaria	338
Canada	675
Chile	157
China	2805
Colombia	721
Costa Rica	55
Dominican Republic	562
Ecuador	154
Egypt	681
Honduras	24
Hong Kong	158
Iceland	65
Indonesia	1690
Iran	1431
Israel	437
Jamaica	60
Japan	1280
Macao	8
Malaysia	197
Mexico	258
Morocco	4544
New Zealand	299
Norway	317
Pakistan	614
Peru	205
Philippines	631
Romania	574
Russia	156
Singapore	160
South Africa	1127
South Korea	364
Sri Lanka	491
Switzerland	380
Thailand	802
Tunisia	187
Turkey	5502
United States	3107
Uruguay	34
Venezuela	228
Vietnam	441
Total (above)	**34083**
Percentage change	**-16**

Denmark (52)	
Argentina	82
Belize	1
Bolivia	63
Brazil	276
Bulgaria	112
Canada	508
Chile	81
China	1171
Colombia	135
Costa Rica	28
Dominican Republic	14
Ecuador	45
Egypt	157
El Salvador	11
Georgia	21
Honduras	18
Hong Kong	0
Iceland	1481
Indonesia	106
Iran	415
Israel	224
Jamaica	15
Japan	334
Kazakhstan	12
Macedonia	129
Malaysia	136
Mauritius	7
Mexico	128
Moldova	24
Morocco	194
New Zealand	227
Norway	3112
Pakistan	743
Panama	8
Peru	68
Philippines	321
Romania	248
Russia	547
Singapore	145
South Africa	199
South Korea	113
Sri Lanka	168
Switzerland	555
Thailand	777
Trinidad & Tobago	7
Tunisia	46
Turkey	1096
Ukraine	457
United States	3464
Uruguay	9
Venezuela	75
Vietnam	283
Total (above)	**18593**
Percentage change	**+12**

Total (above) is the sum of average inflows of immigrants into each destination country, by country of origin (1996-2005).

Percentage change is the percentage change of the overall total during the period 1996-2005.

***Source*:** the National Institute for Statistics (NIS) and Federal Government Service Economics (FOD) for Belgium, Statistics Denmark, Statistics Finland, the German Federal Statistical Office (SBD), the Central Bureau for Statistics of the Netherlands (CBS), Anuario Estadístico de Immigración 2006 and the Spanish National Statistics Institute (INE).

Table A.2. Total yearly inflows into each destination country – Period 1996-2005

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	Total inflow
Belgium	61522	58849	61266	68466	68616	77584	38437	81913	85378	24276	626307
Denmark	54445	50105	51372	50236	52915	55984	52778	49754	49860	52458	519907
Finland	13294	13564	14192	14744	16895	18955	18113	17838	20333	21355	188258
Germany	959691	840633	802456	874023	841158	879217	842543	768975	780175	707352	8296233
Netherlands	108749	109860	122407	119151	132850	133404	121250	104514	94019	92297	1138501
Spain	24536	22261	24032	28243	31587	20724	40175	40486	38717	36573	307334

***Source*:** the National Institute for Statistics (NIS) and Federal Government Service Economics (FOD) for Belgium, Statistics Denmark, Statistics Finland, the German Federal Statistical Office (SBD), the Central Bureau for Statistics of the Netherlands (CBS), Anuario Estadístico de Immigración 2006 and the Spanish National Statistics Institute (INE).

Table A.3. Yearly inflows into each destination country for our sample of origin countries - Period 1996-2005

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	Total inflow
Belgium	15183	14145	15322	16378	18176	21916	10796	24172	24111	24178	184377
Denmark	17570	16752	17643	18268	18885	19993	19922	18586	18776	19533	185928
Finland	4216	4948	5000	4679	5726	6370	6159	5932	6628	7144	56802
Germany	244322	218487	214693	228301	253487	275595	275404	250883	227606	206562	2395340
Netherlands	32744	32870	34710	32430	36031	39229	38684	36457	30273	27399	340827
Spain	16343	21990	30061	42636	109380	123636	159627	176656	216648	292548	1189525
Percentage of total inflow	27,03	28,23	29,51	29,67	38,61	41,04	45,86	48,21	49,05	61,80	39,30

Total inflow is the sum of all yearly inflows over the period 1996-2004 for each destination country.

Percentage of total inflow is the share of the yearly inflows from the origin countries covered in total yearly immigration inflows into these destination countries.

Source: Data on migration inflows were kindly provided by the National Institute for Statistics (NIS) and Federal Government Service Economics (FOD) for Belgium, Statistics Denmark, Statistics Finland, the German Federal Statistical Office (SBD), the Central Bureau for Statistics for the Netherlands (CBS) and the Spanish National Statistics Institute (INE).

Table A.4. FE estimates of migration determinants into the EU 1996-2005 (dependent variable: $M_{odt} / \left(Pop_{ot} / 1000\right)$)

$M_{odt} / \left(Pop_{ot} / 1000\right)$	**Fixed Effects**
$\Delta \ln w_{dt} / 100$	13.04 (1.44)
$\Delta \ln u_{dt} / 100$	-2.94 (-1.97)
$\Delta \ln w_{ot} / 100$	1.13 (0.35)
$\Delta \ln u_{ot} / 100$	1.00* (1.89)
$\ln w_{dt-1} / 100$	4.76 (1.41)
$\ln u_{dt-1} / 100$	-1.72** (-2.56)
$\ln w_{ot-1} / 100$	0.16 (0.16)
$\ln u_{ot-1} / 100$	0.85** (1.96)
ΔMST_{odt}	25.74 (0.92)
$MST_{odt-1} / 10\,000\,000$	2.32 (1.27)
$Time$	-0.12 (-0.75)
$M_{odt-1} / \left(Pop_{ot-1} / 1000\right)$	0.88*** (8.59)
Constant	1.82 (0.67)

Note: t-statistics between brackets. Standard errors are robust to heteroskedasticity and clustered by origin country. Number of observations = 2061, number of groups = 229. * Significant at the 90% confidence interval. ** Significant at the 95% confidence interval. *** Significant at the 99% confidence interval.

Table A.5. GMM estimates of migration determinants into the EU 1996-2005 (Dependent variable: $\ln M_{odt}$)

$\ln M_{odt}$	**Difference GMM (twostep, stacked) (s)GMMd**	**System GMM (twostep; stacked) (s)GMMs**	**Difference GMM (twostep, lag(2 3)) GMMd**	**System GMM (twostep, lag(2 3)) GMMs**
$\Delta \ln w_{dt}$	6.70***	-0.11	3.84**	-0.57
	(4.97)	(-0.06)	(2.26)	(-0.17)
$\Delta \ln u_{dt}$	0.34*	-0.14	0.14	-0.30
	(1.86)	(-0.81)	(0.91)	(-0.91)
$\Delta \ln w_{ot}$	-0.50	-0.76	-0.25	-0.49
	(-1.56)	(-1.42)	(-0.69)	(-0.79)
$\Delta \ln u_{ot}$	0.04	-0.11	0.05	-0.06
	(0.41)	(-1.42)	(0.76)	(-0.72)
$\ln w_{dt-1}$	5.12***	-0.65	6.18***	-1.20
	(3.45)	(-0.61)	(3.80)	(-1.53)
$\ln u_{dt-1}$	-0.18	0.14	-0.18	0.01
	(-1.54)	(0.80)	(-1.48)	(0.11)
$\ln w_{ot-1}$	-0.30	0.19**	-0.50	0.09
	(-1.15)	(2.41)	(-1.24)	(1.27)
$\ln u_{ot-1}$	0.12*	-0.13*	0.14	-0.08
	(1.77)	(-1.70)	(1.45)	(-1.39)
$\Delta \ln MST_{odt}$	1.16***	1.65***	0.37	0.32
	(2.91)	(5.74)	(0.48)	(0.24)
$\ln MST_{odt-1}$	0.54**	0.65***	0.50	0.31*
	(2.24)	(5.50)	(1.52)	(1.78)
Time	-0.17***	0.03	-0.20***	0.04*
	(-3.52)	(0.76)	(-3.42)	(1.65)
$\ln M_{odt-1}$	-0.03	0.28**	-0.41	0.66***
	(-0.22)	(2.29)	(-1.21)	(3.87)
Constant		-51.53		-72.24*
		(-0.83)		(-1.71)
Number of observations	1832	2061	1832	2061
Number of instruments	25	29	36	51
Arellano-Bondtest AR(1)	z=-1.85 Pr>z=0.07	z=-2.45 Pr>z=0.01	z=-0.61 Pr>z=0.54	z=-2.31 Pr>z=0.02
Arellano-Bondtest AR(2)	z=-1.07 Pr>z=0.28	z=-0.79 Pr>z=0.43	z=-1.28 Pr>z=0.20	z=-0.62 Pr>z=0.54
Hansen test of over-identifying restrictions	chi2(13) = 21.90 Pr >chi2 = 0.06	chi2(16) = 85.89 Pr >chi2 = 0.00	chi2(24)= 46.28 Pr >chi2 = 0.00	chi2(38) = 99.85 Pr >chi2 = 0.00
Difference-in-Hansen test				
- gmm IV's for levels		chi2(2) = 3.58 Pr >chi2 = 0.17		chi2(14) = 25.41 Pr >chi2 = 0.03
- iv-style	chi2(10) = 20.13 Pr >chi2 = 0.03	chi2(10) = 79.57 Pr >chi2 = 0.00	chi2(10) = 24.89 Pr >chi2 = 0.01	chi2(10) = 51.18 Pr >chi2 = 0.00

(1) Standard errors are robust to heteroskedasticity.

(2) In the first two equations, we stack the gmm instrument variables. In the latter two, we only use the second and third lag of their levels.

(3) Arellano-Bond tests for AR(1) and AR(2): tests for first-order and second-order serial correlation in the residuals, asymptotically distributed as *N*(0,1).

(4) Hansen test: test of over-identifying restrictions for the GMM estimator, asymptotically distributed as χ^2_{df} (null hypothesis: instrument variables are valid as a group). Hansen tests are weak when instruments are many, though.

(5) Difference-in-Hansen test: test of the additional moment conditions used in the system GMM estimators relative to the corresponding first-differenced GMM estimator, asymptotically distributed as χ^2_{df} . Table reports difference-in-Hansen results (null hypothesis: instrument variables are exogenous). Hansen tests are weak when instruments are many, though.

Table A.6. GMM estimates of migration determinants into the EU 1996-2005 (Dependent variable: $\ln M_{odt}$, $\Delta \ln w_{dt}$ & $\ln w_{dt-1}$ possibly not exogenous)

$\ln M_{odt}$	Difference GMM (twostep, stacked) ($\Delta \ln w_{dt}$ & $\ln w_{dt-1}$)	System GMM (twostep; stacked) ($\Delta \ln w_{dt}$ & $\ln w_{dt-1}$)	Difference GMM (twostep, stacked) ($\Delta \ln w_{dt}$)	System GMM (twostep, stacked) ($\Delta \ln w_{dt}$)
$\Delta \ln w_{dt}$	6.32**	-4.49	6.91***	-0.01
	(2.56)	(-1.53)	(3.44)	(-0.01)
$\Delta \ln u_{dt}$	0.37	-0.38	0.24	-0.11
	(1.25)	(-1.33)	(1.02)	(-0.48)
$\Delta \ln w_{ot}$	-2.43	-2.99*	-0.31	-1.17
	(-1.30)	(-1.95)	(-0.22)	(-0.82)
$\Delta \ln u_{ot}$	-0.99	-1.72*	0.28	-0.17
	(-0.87)	(-1.69)	(0.31)	(-0.20)
$\ln w_{dt-1}$	4.70	-1.73	5.03**	0.48
	(1.51)	(-0.73)	(2.11)	(0.46)
$\ln u_{dt-1}$	-0.13	0.03	-0.20	0.34**
	(-0.45)	(0.07)	(-0.93)	(2.15)
$\ln w_{ot-1}$	-2.43	0.09	-0.22	0.17**
	(-1.20)	(1.33)	(-0.14)	(2.16)
$\ln u_{ot-1}$	-0.94	-0.26*	0.36	-0.08
	(-0.82)	(-1.89)	(0.40)	(-0.69)
$\Delta \ln MST_{odt}$	1.31*	1.68***	1.18*	1.85***
	(1.81)	(3.97)	(1.90)	(3.80)
$\ln MST_{odt-1}$	0.63	0.63**	0.63**	0.69***
	(1.57)	(4.38)	(2.15)	(4.66)
Time	-0.06	0.07	-0.18	-0.01
	(-0.41)	(0.85)	(-1.49)	(-0.22)
$\ln M_{odt-1}$	0.02	0.33**	-0.00	0.30**
	(0.13)	(2.49)	(-0.01)	(2.18)
Constant		-122.03		8.22
		(-0.88)		(0.13)
Number of observations	1832	2061	1832	2061
Number of instruments	38	43	30	36
Arellano-Bondtest AR(1)	z=-1.71 Pr>z=0.09	z=-2.72 Pr>z=0.01	z=-1.59 Pr>z=0.11	z=-2.42 Pr>z=0.02
Arellano-Bondtest AR(2)	z=-1.01 Pr>z=0.31	z=-0.85 Pr>z=0.40	z=-1.03 Pr>z=0.30	z=-0.76 Pr>z=0.45
Hansen test of over-identifying restrictions	chi2(26) = 32.20 Pr >chi2 = 0.18	chi2(30) = 92.57 Pr >chi2 = 0.00	chi2(18)= 22.87 Pr >chi2 = 0.20	chi2(23) = 99.21 Pr >chi2 = 0.00
Difference-in-Hansen test				
- gmm IV's for levels		chi2(4) = 18.50 Pr >chi2 = 0.00		chi2(3) = 5.37 Pr >chi2 = 0.15
- iv-style	chi2(7) = 6.75 Pr >chi2 = 0.46	chi2(10) = 48.56 Pr >chi2 = 0.00	chi2(7) = 4.27 Pr >chi2 = 0.75	chi2(8) = 66.73 Pr >chi2 = 0.00

(1) Standard errors are robust to heteroskedasticity. All gmm-instrument variables are stacked.

(3) In the first 2 columns, we include $\Delta \ln w_{dt}$ and $\ln w_{dt-1}$ in gmm-style, while in the last two columns only $\Delta \ln w_{dt}$.

(4) Arellano-Bond tests for AR(1) and AR(2): tests for first-order and second-order serial correlation in the residuals, asymptotically distributed as $N(0,1)$.

(5) Hansen test: test of over-identifying restrictions for the GMM estimator, asymptotically distributed as χ^2_{df} (null hypothesis: instrument variables are valid as a group). Hansen tests are weak when instruments are many, though.

(6) Difference-in-Hansen test: test of the additional moment conditions used in the system GMM estimators relative to the corresponding first-differenced GMM estimator, asymptotically distributed as χ^2_{df}. Table reports difference-in-Hansen results (null hypothesis: instrument variables are exogenous). Hansen tests are weak when instruments are many, though.