DIW Berlin

Deutsches Institut
für Wirtschaftsforschung

Potential Migration from Central and Eastern Europe into the EU-15 – An Update

Report for the European Commission, DG Employment and Social Affairs

Patricia Alvarez-Plata
Herbert Brücker (Coordinator)
Boriss Siliverstovs
DIW Berlin

Final Report

Berlin, 2003

Contents

Executive Summary

The purpose of this study is to update the estimates of potential for migration from the ten accession candidate countries of Central and Eastern Europe to the present EU Member States. A study carried out by European Integration Consortium on behalf of the DG Employment and Social Affairs of the European Commission estimated the annual growth of the migrant population from the ten Central and Eastern European countries (CEEC-10) in Germany at around 220,000 persons p.a. immediately after the hypothetical introduction of free movement in 2002. The long-run stock of the migrant population was estimated at around 2.5 million persons, which corresponds to a share of 2.5 per cent of the population in the CEEC-10. An extrapolation of these estimates to the present EU Member States resulted in an annual increase of the foreign population from the CEEC-10 of some 335,000 persons and a long-run migrant stock of 3.9 million residents (3.9 per cent of the population of the CEEC-10). Although a large number of studies confirm these results, there have also been estimates of either much higher or much lower migration potential.

In order to further reduce uncertainty as to the magnitude of migration potential, this study pursues three main objectives: (i) to assess whether the recent changes in the economic determinants of migration will affect the magnitude of migration potential, (ii) to evaluate the forecasting performance of different estimation procedures in order to improve the reliability of our estimates and to analyse the reasons for the variety the estimates of the migration potential in the empirical literature, and (iii) to analyse the implications of effective migration restrictions during the transitional periods for the short-run and long-run migration potential.

Change in the migration conditions

The migration conditions between the EU and the CEECs have changed only slightly since the study carried out by the European Integration Consortium: the income difference in purchasing power parities is estimated by Eurostat to be two percentage points below the income difference used in the old scenario, and the conditions for economic growth have tended to stabilise in many of the accession candidate countries. Moreover, in the main host country for immigration from the CEECs, Germany, the economic conditions have deteriorated significantly in 2001 and 2002. Migration stocks in the present EU have increased only slightly since 1998: we estimate the number of citizens from the CEECs residing in the present EU-15 at 840,000 persons in 1998, and at around 1.0 million persons in 2001/2002. The number of employees from the CEECs in the EU-15 can be estimated at around 400,000.

Since the migration scenario is based on long-term assumptions on growth and convergence of the relevant variables, these minor changes affect the size of the migrant potential only marginally: In the model of the European Integration Consortium, it reduces the short-run migration potential by no more than one or two thousand persons p.a. and has almost no impact on the long-run migration potential.

Forecasting performance of different econometric estimators

Most empirical studies estimate the long-run migration potential from the CEECs in the EU at between 2 per cent and 4 per cent of the population. Some studies find, however, that the long-run migration potential is much greater: a study carried out by the Ifo-institute estimates the migration potential from the CEECs in Germany at between 8 and 10 per cent of the CEECs' population in the long-run equilibrium, which would correspond to a migration potential in the EU-15 of between 12 and 15 per cent – given the present distribution of migrants from the CEECs across the EU-15 countries.

In order to improve our projections of migration potential and shed some light on the reasons for divergent findings, we compared the forecasting performance of different estimation procedures. The variety of migration projections cannot be traced back to differences in the explanatory variables: almost all models discussed in the empirical literature explain migration by income and employment opportunities in the respective countries and a set of institutional variables which should capture different migration restrictions. The main reason for the varying estimation results is that the different studies use different econometric estimators. All econometric models impose restrictions on the data necessary to estimate the coefficients for the variables of interest. In the case of panel data sets, which have a time dimension and a cross-sectional dimension, econometric models can exploit the variation of the data over time and across countries in order to estimate the coefficients. Different estimators use different sources of this variation and impose different restrictions on the data and thus yield different results.

More specifically, the pooled OLS estimator, which was used in the Ifo–study among others, uses the variation in the data both over time and between countries, but assumes that all cross-sections share a common constant. This implies that country specific fixed effects are ignored. However, country-specific effects such as geography, language, culture, etc. have proven to be very important in the migration context. As a consequence, econometric theory predicts that the pooled OLS estimates are biased, which affects the long-run results in particular. The European Integration Consortium study employed estimators which consider country-specific fixed effects, but exploit only the time-dimension of the data for the estimation of the coefficients.

In this study, we used two data sets for testing the forecasting performance of the different estimators. The first data set comprises 33 years of immigration to Germany from 19 source countries. The second data set comprises the migration between 215 countries over 8 years. In the case of the German sample, which has the large time and small cross-sectional dimension, the traditional estimators which consider the country-specific fixed clearly outperform all other estimators. Conversely, in the case of the European sample, newly developed GMM estimators show a better dynamic forecasting performance within the sample than traditional panel estimators. These newly developed estimators address the problem of estimation bias, which may result if dynamic models are estimated in data sets with a short time dimension. Thus, our results confirm the predictions from econometric theory. The forecasting performance of the estimators has turned out to be much better with the German sample, which may result from the better quality of this data set among other factors. Interestingly enough, the dynamic forecasting performance of the pooled OLS estimator turned out to be the weakest of all estimators in both samples. This result suggests that the pooled OLS estimations of dynamic migration models are heavily biased. This has an especially distorting effect on the long-term forecasts.

Note that these results have not only a technical character. We have also learned something about migration behaviour: beyond macroeconomic variables such as per capita income levels and employment rates, the propensity to migrate is heavily affected by observable and unobservable factors. Those factors such as geographical location, culture, language, the social and political environment seem to affect the propensity to migrate heavily. Ignoring these factors in forecasting migration yields therefore distorted results.

The scale of migration

The quantitative findings in this study are roughly in line with those from the *European Integration Consortium*. On basis of the fixed effects estimators, which show the best forecasting performance, the net increase of the number of foreign residents from the CEEC-10 in Germany – which roughly equals the net migration flow – is estimated to number 180,000 persons immediately after the introduction of free movement. It will reach its peak at around 225,000 persons one year later. The long-run migration potential is estimated at 2.3 million persons. The peak of the foreign population from the CEECs in Germany will be reached around 25 years after free movement has been introduced. Thus, the long-run migration potential is around 200,000 persons below the estimates of the *European Integration Consortium*, while the short-run results for the net increase of the foreign population are very similar. The medium- and long-term simulation results of the other estimators with a good forecasting performance are roughly consistent with those of the fixed effects estimators, although the estimates of the short-term increase of the foreign population are in some cases substantially lower.

Sensitivity of the results

These simulations are based on the assumption that per capita GDP levels converge at an annual rate of 2 per cent and that the employment rates in Germany and the CEECs remain constant at their average level of the 1990s. In order to test the sensitivity of the results, we employed in a high migration scenario a convergence rate of one per cent, a German unemployment rate which is one-third below the average of the 1990s, and an unemployment rate in the CEECs which is one third-above the average of the 1990s. Conversely, in a low migration scenario we employed a convergence rate of 3 per cent, a German unemployment rate which is one-third above the average of the 1990s, and an unemployment rate in the CEECs which is one-third below the average of the 1990s.

In case of the high migration scenario, the initial net increase of the foreign population is expected to be between 35,000 and 50,000 persons higher than in the baseline scenario, while the long-migration potential is estimated to be 500,000 persons higher. Conversely, in the low migration scenario, the initial net migration is estimated to be around 30,000 persons, and the long-run migration potential 330,000 persons below the baseline scenario.

Results of the European Sample

The estimates based on the European sample are much less precise than those based on the German sample. Specifically, the employment variables and most of the distance variables turned out to be insignificant. The simulation based on the GMM estimator with the best within-sample forecasting performance yields an initial increase in the number of foreign workers from the CEEC-10 to the EU-15 of 111,000 persons and a long-run stock of foreign workers of 2.2 million. Given that almost 40 per cent of the foreign residents from the CEECs participate in the labour market as employees and that around 60 per cent of the migrants from the CEECs reside in Germany, the short-run estimates are slightly below, and the long-run figures around one-third above the estimates derived from the German sample.

While the aggregate figures look plausible at first glance, the geographical distribution does not: according to these estimates, Germany's share in the number of migrant workers from the CEECs is expected to fall from 60 per cent at present to between 12 and 30 per cent, depending on the estimator. Given that regional migration patterns show a high persistence over time due to migrant-network effects, these results are certainly not plausible. We thus abstained from basing our simulations of the regional distribution of migrants from the CEEC-10 upon these estimates.

Extrapolation of the German results to the EU-15

An extrapolation of the results from the German sample to the EU-15 yields an initial net increase of residents from the CEEC-10 in the amount of 290,000 persons and a long-run stock of 3.8 million persons. The net increase reaches its peak at around 370,000 persons. Under the assumptions of the high migration scenario, the long-run migration potential could reach 4.5 million persons, and in case of the low migration scenario, 3.2 million persons.

Actual migration flows and stocks may substantially deviate from our simulation results. It is worth noting that country-specific factors that rely on largely unobservable variables have an important impact on the magnitude of migration flows. Extrapolating the coefficients of our regressions to countries outside the sample therefore affects the quality of the projections. Moreover, the extrapolation of our results to the EU-15 relies heavily on the assumption that the present distribution of migrants from the CEECs across individual EU Member States will remain constant. Although we observe a high persistence of regional migration patterns, these need not be necessarily the case in the longer terms.

The impact of transitional periods

The simulation of the transitional periods shows that postponing the introduction of free movement has only a marginal impact on the scale of migration: postponing free movement for seven years or more will reduce initial migration by only a few thousand persons. Thus, postponing free movement does not mitigate migration pressures if policies of zero net migration are pursued during transitional periods. Small reductions in initial migration inflows can, however, be achieved if free movement is postponed a few years for some of the accession countries, since in this case, inflows are distributed across a longer time period.

From an economic perspective, a restrictive use of transitional periods will therefore fail to mitigate possible pressures from migration on the labour market. If the objective of transitional periods is to ease the adjustment process, then transitional periods should be used to implement safeguard clauses or quotas. In this case, potential migration pressures are reduced step by step, such that introducing free movement at a later stage will yield only a moderate influx instead of a migration hump. Note that the results derived from our migration model suggest that migration responds fairly quickly to changes in GDP and employment growth. Thus, international migration contributes significantly to the labour supply's adjustment to fluctuations in economic activity.

1 Introduction

In this study we update the estimation of migration potential that was presented in a study conducted by the European Integration Consortium on behalf of the DG Employment and Social Affairs of the European Commission (Boeri/Brücker et al., 2001). In the European Integration Consortium's study, the annual growth of the migrant population from the CEEC-10 in Germany was estimated at around 220,000 persons p.a. immediately after the hypothetical introduction of free movement in 2002. The long-run stock of the migrant population was estimated there at around 2.5 million persons in the baseline scenario, which corresponds to a share of 2.5 per cent of the population in the CEEC-10. An extrapolation of these estimates to the EU based on the present share of the migrant population from the CEEC-10 in the EU-15 resulted in an annual increase of the foreign population from the CEEC-10 of some 335,000 persons and a long-run stock of migrants of 3.9 million residents (3.9 per cent of the population from the CEEC-10). Less than 40 per cent of the migrant population from the CEEC-10 in the EU are employed at present.

Some minor changes have occurred in the economic conditions affecting migration since the European Integration Consortium's study was carried out. First, the income gap between the present EU and the Central and Eastern European accession candidate countries has diminished slightly according to the purchasing power parity estimates by Eurostat. The growth rates of GDP and the trend of GDP convergence to western European levels has also stabilised in most accession candidate countries in recent years. Second, in some of these countries, unemployment rates have declined substantially, while in others, unemployment has remained at high levels or even increased. Third, the economic conditions in the main host country for migration from the CEECs, Germany, have deteriorated over the last two years. Nevertheless, the income gap between the present EU Members and the CEE accession candidates is still extraordinarily high, such that large economic incentives to migrate persist.

Meanwhile more is known today about the time plan for accession and the application of transitional periods for the free movement after accession of the CEECs: Accession negotiations with eight[1] of the ten Central and Eastern European candidate countries have been completed and these countries will accede in 2004. For the remaining two countries (Bulgaria and Romania), the European Council offered the prospect of acceding in 2007. Moreover, the European Council agreed on transitional periods for free movement lasting up to a maximum of seven years after accession. Although not all EU Members will apply for transitional periods, the two most heavily affected countries, Austria and Germany, have already announced that they plan to adopt such periods. Thus, free movement might be postponed there until 2011 for the eight participants in the first accession round, and until 2014 for the participants in the second accession round. However, it remains unknown whether Austria and Germany will make use of the maximal length of the transitional periods and what kind of immigration policies will be pursued in these countries during the transitional periods. It is also not yet known which other countries will make use of transitional periods.

The analysis in this report has three main objectives. Firstly, we assess whether the recent changes in the economic determinants of migration have affected the magnitude of migration

[1] Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovak Republic, Slovenia.

potential. Secondly, we analyse the quantitative implications of different procedures to estimate migration potential in the light of the recent theoretical and methodological discussion. Specifically, in this study we employ a wide variety of estimators that are used in other studies or are discussed in the literature and analyse their forecasting performance. The purpose of this exercise is not only to improve our forecasts, but also to shed light on the reasons for the widely varying estimates of migration potential discussed in studies on this topic. Finally, we assess whether transitional periods, which effectively restrict migration flows, will also reduce the migration potential in the long run.

An assessment of the migration potential is also seriously hampered by a lack of reliable data on migration stocks and flows in many EU countries. The European integration consortium thus based its forecast on German data, where stocks and flows of migrants are reported for a broad range of source countries since 1967. The results of the German estimates have been extrapolated to the EU-15 on the basis of the present distribution of migrants from the CEECs across the EU-15. In this study, in addition to the German data, we use data on foreign employees derived from the Eurostat Labour Force Survey. These data have the advantage of generally covering all destination countries within the EU-15, but they have the disadvantage of covering only a very brief period of time (nine time series observations at the maximum). Moreover, in many cases the data are distorted by low response rates. Nevertheless, in this study we use this second data source in order to contrast the findings based on the German data set with the results based on the European Labour Force Survey data.

The remainder of this report is organised as follows: In the next section, we summarise key facts on the scale of migration from the CEECs into the EU, the human capital characteristics of migrants from the CEECs and their labour market performance provide background for the study itself (Section 2). The following section describes the development of the key macroeconomic migration determinants in the 1990s and early 2000s (Section 3). In Section 4, we present a brief review of the literature on the estimation of potential migration from the CEECs. In Section 5, we discuss the present estimators and analyse their forecasting performance within the different data sets. In Section 6, we present the simulations of migration potential based on various assumptions on the development of the exogenous variables and the time schedule for free movement. Section 7 summarises our conclusions.

2 What do we know about migration from the CEECs?

Immigration from the ten accession candidates from Central and Eastern Europe into the EU reached its peak with a net immigration of some 300,000 persons immediately after the fall of the Berlin wall in the year 1990. Net immigration declined sharply after the German recession of 1993 and annual net immigration figures from the CEEC-10 into the EU-15 can be estimated to be below 50,000 thousand persons p.a. in the second half of the 1990s and the early 2000s. The number of citizens from the CEEC-10 residing in the present European Union (EU-15) was around 874,000 people or 0.23 per cent of the total population of the EU-15 in 1999.[2] Recent data from national foreigner registers and population censuses indicate that around one million citizens from the CEECs resided in the EU in 2001/2002, although figures for large EU Members such as France and the UK have to be estimated since national statistics do not report the number of foreign residents from the CEECs there (Table 2.1).

Table 2.1 **Residents from the CEEC-10 in the EU-15, 1999 and 2001/2002**

	residents 1999		residents 2001/2002	
	persons	*% of population*	*persons*	*% of population*
Austria	103 000 [1,2]	11.78	78 617	8.04
Belgium	11 233	1.29	13 208	1.35
Denmark	9 167	1.05	11 252	1.15
Finland	12 804	1.46	13 590 [4]	1.39
France	22 000 [1,2]	2.52	25 869	2.65
Germany	547 837	62.67	597 137	61.10
Greece	20 131 [3]	2.30	23 671	2.42
Ireland	200	0.02	235	0.02
Italy	55 791 [3]	6.38	107 419	10.99
Luxembourg	700 [1,2]	0.08	2 654	0.27
Netherlands	11 266	1.29	14 417	1.48
Portugal	819	0.09	963	0.10
Spain	13 819	1.58	16 249	1.66
Sweden	26 394	3.02	26 168	2.68
United Kingdom	39 000 [1,2]	4.46	45 858	4.69
EU-15 (est.)	874 161	100.00	977 307	100.00

Notes: 1) Estimated on basis of employment figures. - 2) 1997. - 3) 1998. - 4) 2001.
Sources: National statistical offices, calculations and estimates of the authors.

[2] See Boeri/Brücker et al. (2001). This estimate is derived from the latest available figures from Eurostat (2001), national statistical offices, and the European Labour Force Survey.

The total number of employees from the CEECs is reported by the Eurostat Labour Force Survey at around 240,000 permanent workers for the year 1995, although this might underestimate the total labour supply from the CEEC-10, particularly with regard to temporary workers from the CEECs. Temporary workers (seasonal workers, project-tied workers) number between 200,000 and 250,000 workers each year in Germany alone at the end of the 1990s (Hönekopp, 2000). The annual employment volume of these temporary migrants equals around one-quarter, such that the annual employment volume of the temporary workers can be estimated at around 40,000-50,000 man years. Altogether, if we assume that between 35 per cent and 38 per cent of the residents from the CEECs participate as employees in the labour market, a reasonable estimate of the total labour supply of citizens from the CEEC-10 in the EU-15 is 430,000 full-time workers.

The reported numbers of border commuters in Austria amount to 5,900 employees and in Germany to 5,700 employees in 1999. All these figures refer to the official statistics and do not cover non-registered migrants and commuters from the CEECs who are employed in the shadow economies of the present EU. The lifting of the visa requirement increased the opportunities of employment in those sectors and might have increased the labour supply substantially there. Unfortunately, to the best of our knowledge, there are no serious estimates about the size of non-registered employees from the CEECs.

Table 2.2 **Stock of Residents from the CEECs in EU countries (latest available date)**

	Austria	Denmark	Finland	Germany	Italy	Luxembourg	Netherlands	Sweden	total EU-8	
year	2001	2003	2001	2001	2002	2003	2002	2001		
	residents by citizenship (persons)								*persons*	*% of population*
Bulgaria	4 217	443	297	38 143	6 758	193	na	805	50 856	0.63
Czech Republic	7 313	412	125	38 504 [4)]	3 468	173	na	471	50 466	0.49
Estonia	54	523	10 839	3 880	205	131	na	1 662	17 294	1.28
Hungary	12 729	457	654	55 978	3 066	417	na	2727	76 028	0.75
Latvia	153	894	227	8 543	467	41	na	780	11 105	0.47
Lithuania	208	1 583	204	11 156	322	32	na	727	14 232	0.41
Poland	21 841	5 410	694	310 432	29 282	885	na	15 511	384 055	0.99
Romania	17 470	1 258	489	88 102	62 262	596	na	2 495	172 672	0.76
Slovak Republic	7 739	216	51	23 004 [4)]	6	114	na	363	31 493	0.58
Slovenia	6 893	56	10	19 395	1 583	72	na	627	28 636	1.44
CEEC-10	78 617	11 252	13 590	597 137	107 419	2 654	14 417	26 168	851 254	0.81
CEEC-8	56 930	9 551	12 804	470 892	38 399	1 865	na	22 868	613 309	0.83

Sources: National statistical offices, authors' calculations.

Geography has a major impact on the regional distribution of migrants from the CEECs across the EU: around 70 per cent of the citizens from the CEEC-10 reside in Austria and Germany, i.e. neighbouring countries which have long borders with the CEECs. The total stock of residents from the CEEC-10 is reported in Austria at 79,000 persons and in Germany at 598,000 persons, corresponding to a share of 1.1 per cent and 0.7 per cent of the total population there in 2001 (Tables 2.1, 2.2). In other countries neighbouring to the CEECs, such as Finland, Sweden and Greece, we observe however migrant shares which are close to the EU average.

Although the high gap in wages and per capita GDP levels does generally also apply to the border regions of the EU to the CEECs, migration to many border regions is mitigated by low population density and less favourable economic conditions relative to national averages. The eastern borders of Austria and Bavaria are a notable exception, however: in Austria, the share of citizens from the CEECs in total employment was reported at 7 per cent in the Burgenland, and at 3 per cent in Vienna in 1998, which is well above the national average (1.9 per cent). In contrast, the share of CEEC-citizens employed in Kärnten, the region bordering on Slovenia, is well below the national average. Along the Eastern German borders to Poland and the Czech Republic, employment shares of CEEC-citizens are at one-third of the national average, while along the Bavarian border to the Czech Republic, the share of citizens from the CEECs in total employment is reported as three times the national average in 1999. Thus, beyond geography, regional migration patterns reflect differences in employment opportunities and wage levels across regions.

The average share of citizens from the CEECs residing in the eight reporting EU countries amounts to 0.9 per cent of the home population, and the share residing in the total EU-15 can be estimated to be around one at 0.9 per cent of the home population in 2001. Among the CEEC-10, the share of citizens residing in EU is highest for Slovenia, Poland, Estonia and Hungary. This pattern reflects different opportunities to emigrate in the 1980s rather than cross-country differences in economic factors which determine migration such as per capita GDP levels or unemployment rates.

The available information on the human capital characteristics of migrants from the CEECs stems from survey data and suffers from low response rates. Data derived from the Eurostat Labour Force Survey suggest that the migrant population from the CEECs possesses a higher share of completed secondary and tertiary education than the native population in the host countries of the EU (Boeri/Brücker et al., 2001). Around 70 per cent of the workers from the CEECs are in the 25-44 age group, while in the EU on average only 55 per cent of all workers belong to this group (Eurostat Labour Force Survey 2001). Low response rates of migrants from the CEECs may, however, distort the picture, such that this information has to be treated with caution. Moreover, the Eurostat Labour Force Survey data may overstate education levels of migrant workers from the CEECs, since temporary workers are covered only very incompletely there.

The labour market performance of migrants from the CEECs does not, however, reflect their presumably high education levels. According to the Eurostat Labour Force Survey data, the activity rate (ratio of labour force to working age population) and the participation rate (ratio of employees to working age population) were, at 63 per cent and 53 per cent respectively, below rates of EU natives (68 per cent and 61 per cent respectively) in 1999. Unemployment rates, at 16.5 per cent, were well above the EU average of 10.3 per cent (Eurostat Labour Force Survey, 2001; Boeri/Brücker et al., 2001). Moreover, the branch structure of employment suggests that workers from the CEECs are – like other foreign workers – concentrated in branches with a high share of relatively less-skilled labour (e.g. construction, private households). Again, all this information may be biased due to low response rates in data based on household surveys.

3 The development of macroeconomic migration determinants

Most models of international migration refer to two main macro-economic variables which affect the migration decision in one way or another: wages and (un-)employment rates. Following the traditional Harris-Todaro (1970) model, wages and employment rates determine expected income in the respective locations. Moreover, the variation in the variables over time and between individuals may affect migration behaviour, depending on the assumptions on risk preferences and the formation of expectations under uncertainty.

Table 3.1 **PPP-GDP and GDP of the accession candidate countries, 2001**

	PPP-GDP[1)]		GDP[2)]	
	EURO	in % of EU-15	EURO	in % of EU-15
Bulgaria	6 500	28.0	1 875	7.8
Czech Republic	13 300	57.2	6 164	25.6
Estonia	9 800	42.2	4 535	18.9
Hungary	11 900	51.2	5 813	24.2
Latvia	7 700	33.1	3 613	15.0
Lithuania	8 700	37.4	3 638	15.1
Poland	9 200	39.6	5 092	21.2
Romania	5 900	25.4	1 982	8.2
Slovak Republic	11 100	47.8	4 229	17.6
Slovenia	16 000	68.9	10 499	43.7
Cyprus	21 118	90.9	15 171	63.1
Malta	..	..	10 553	43.9
Turkey	5 200	22.4	3 213	13.4
CEEC-8	10 675	45.9	5 407	22.5
CEEC-10	9 322	40.1	4 395	18.3
CC-13	7 764	33.4	4 001	16.6
CC-10	10 725	46.2	5 534	23.0
EU-15	23 236	100.0	24 050	100.0

Notes: 1) Eurostat estimate. - 2) At current exchange rates, 1995 prices.
Sources: Eurostat, 2002; OECD, 2002.

Recent data from the prospective source countries in Central and Eastern Europe and the prospective host countries in the EU show that the absolute gap in income levels is still high. This holds particularly true for the wage and GDP per capita gap at current exchange rates. While Eurostat estimates GDP per capita measured in purchasing power parities for the CEEC-10 in 2001 at 40 per cent of that in the EU-15 (Eurostat, 2002b), GDP per capita at current exchange rates amounted to 18 per cent of the average level in the EU-15 (OECD, 2002). GDP levels at purchasing power parities and at current exchange rates of the eight first-round accession candidate countries were, with 46 per cent and 23 per cent of EU-15 levels respectively, above the CEEC-10 average (Table 3.1). Note that a difference between GDP levels measured at purchasing power parities and those measured at current exchange rates is not unusual, since the productivity gap between rich and poor economies is higher in tradable sectors (e.g. manufacturing industries) than in non-tradable sectors (e.g. services, retail trade). This difference can be expected to diminish if productivity of the tradable sectors tend to converge in the long run.

Table 3.2 **PPP-GDP of traditional source countries of EU immigration, 1960-1990**

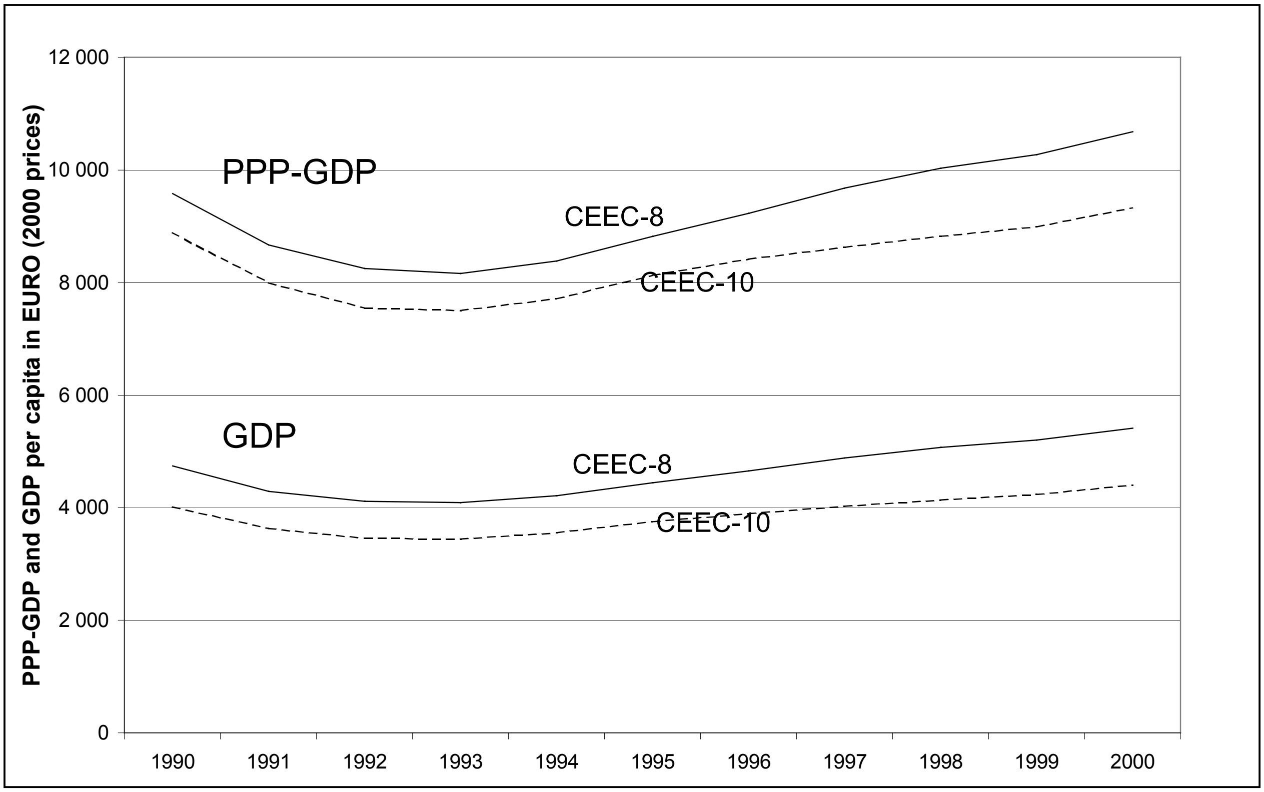
	1960		1970		1980		1990	
	in USD	in % of EU-15	in USD	in % of EU-15	in USD	in % of EU-15	in USD	in % of EU-15
Greece	3 204	38	6 327	53	9 139	59	10 051	54
Italy	5 789	68	9 508	80	13 092	85	15 951	85
Portugal	3 095	37	5 885	49	8 251	54	10 685	57
Spain	3 437	41	7 291	61	9 539	62	11 752	63
Morocco	1 511	18	1 764	15	2 132	14	2 399	13
Turkey	1 801	21	2 437	20	3 129	20	3 989	21
Yugoslavia	2 401	28	3 657	31	5 876	38	5 917	32

Source: Maddison (1995), calculations of the authors.

The income gap between the EU and the accession candidates from Central and Eastern Europe is well above levels of past enlargement rounds: when Greece acceded in 1981, PPP-GDP levels were at 65 per cent of the EU-15, and when Spain and Portugal acceded in 1986 they were 66 per cent and 70 per cent of the EU-15 respectively (Maddison, 1995). However, the income gap between the EU and the accession candidate countries is similar to that between the average of the EU-15 and its later southern Members in the 1960s and early 1970s, i.e. at the times of guest worker recruitment in Germany, France, the Benelux countries and many other Members of the present EU. Moreover, the income gap is smaller than in case of the main source countries of immigration into the EU, i.e. the Mediterranean countries in Northern Africa and South-Eastern Europe (Table 3.2). Thus, the income gap between the EU and the accession candidates from Central and Eastern Europe is hardly a new phenomenon compared with other migration episodes in Europe after Word War II.

Due to the transition recession, average annual GDP growth in the CEEC-10 at current exchange rates and purchasing power parities (PPP) was well below that in the EU-15 in the years 1990-2001, at 1.3 and 1.1 per cent respectively. However, since the end of the transition recession, GDP growth rates are, at 3.5 per cent (GDP at current exchange rates) and 3.2 per cent (PPP-GDP) p.a. in the period 1993-2001, well above those in the EU-15. This tendency toward convergence has proved to be rather robust over recent years. Considerable differences in the growth performance across countries exist. In general, growth rates of the eight first-round accession candidates (CEEC-8) have been, at 4.2 per cent (PPP-GDP) and 3.9 per cent (GDP) in the years 1993-2001, well above those of the total country sample (Figure 3.1, Table A1). The variance of the growth rate of GDP levels in the CEECs tends to decline over time, but is still considerably higher than that of the present EU Members.

Figure 3.1 **PPP-GDP and GDP of the CEECs, 1990-2000**



With respect to unemployment, the picture is more scattered. While unemployment rates in some of the accession countries are similar to levels in the EU-15 and on the decline (Estonia, Hungary, Latvia, Romania), in other countries we observe high and increasing unemployment rates (Slovakia, Poland, Bulgaria, Lithuania). In general, the level of unemployment is not only above the EU average, but also above that in the most affected of the current EU host countries (Table 3.3).

In the potential host countries for CEE migration, the economic incentives to immigrate have deteriorated in the last two years or so. In particular, in the main host country, Germany, we observe a substantial decline in GDP growth rates in 2001 and 2002, and an increase in unemployment rates. Austria is less affected by the economic slow-down, but growth rates have declined there, as well. Since the economic developments in Germany reflect both structural factors and the fluctuation of economic activity in the course of the business cycle, it is not possible to assess whether the slow-down of economic growth in Germany and other continental European countries will have a long-lasting impact on the conditions to migrate at the present stage.

Tab. 3.3 **Unemployment rates in the accession candidate countries, 1990-2001**

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
						unemployment rate in % of labour force						
Slovak Republic	1.6	11.8	10.4	14.4	14.8	13.1	12.8	12.5	15.6	19.2	17.9	18.6
Poland	6.5	12.2	14.3	16.4	16	14.9	13.2	10.3	10.4	13.1	15.1	17.4
Bulgaria	1.8	11.1	15.3	16.4	12.8	11.1	12.5	13.7	12.2	16	17.9	17.3
Lithuania	na	na	3.5	3.4	4.5	7.3	6.2	6.7	6.9	10	12.6	12.9
Slovenia	na	10.1	13.3	15.5	14.2	14.5	14.4	14.8	14.6	13	12	11.8
Czech Republic	0.7	4.1	2.6	3.5	3.2	2.9	3.5	5.2	7.5	9.4	8.8	8.9
Romania	1.3	3	8.2	10.4	10.9	9.5	6.6	8.8	10.3	11.5	10.5	8.6
Hungary	1.7	7.4	12.3	12.1	10.9	10.4	10.5	10.4	9.1	9.6	8.9	8
Latvia	na	na	2.3	5.8	6.5	6.6	7.2	7	9.2	9.1	7.8	7.7
Estonia	na	na	1.6	5	5.1	5	5.6	4.6	5.1	6.7	7.3	7.2
CEEC-8	4.2	9.3	11.0	12.7	12.4	11.9	11.0	9.7	10.2	12.5	13.4	14.6
CEEC-10	3.4	8.1	10.7	12.5	12.1	11.3	10.2	9.8	10.4	12.5	13.1	13.5
memo items												
Cyprus	1.8	3.0	1.8	2.6	2.7	2.6	3.1	3.4	3.4	3.6	3.5	3.5
Malta	3.8	3.6	4.0	4.5	4.1	3.8	4.4	5.0	5.1	5.3	4.5	4.9
Turkey	8.2	7.9	7.9	7.6	8.1	6.9	6.1	6.4	6.8	7.6	6.6	7.9
CC-13	5.0	8.0	9.7	10.7	10.6	9.6	8.6	8.5	9.0	10.6	10.6	11.3
CC-10	4.2	9.2	10.9	12.6	12.3	11.7	10.9	9.6	10.1	12.3	13.3	14.4

Sources: UNECE (2002), OECD (2002), authors' calculations.

Altogether, economic incentives to migrate in the CEECs declined slightly in the second half of the 1990s and the early 2000s: GDP levels in the CEECs tended to converge and growth rates of GDP tended to stabilise. At the same time, economic conditions in major host countries such as Germany deteriorated. However, the gap in per capita income between the present EU Members and the CEE accessions candidates is still extraordinarily high and the stock of the migrant population residing in the present EU is, relative to the income gap, low.

4 Review of the literature

Since 1990, several studies have tried to assess the migration potential associated with the Eastern Enlargement of the EU. Different methodologies have been used: opinion polls; extrapolations from South-North migration; and multivariate analysis of past migration episodes in econometric models. Surprisingly enough, most studies – with some notable exceptions – agree in predicting a long-term migration potential of about 2-4 per cent of the current population of the CEEC-10, that is, about two to four million individuals. It is important to keep in mind the *ad hoc* assumptions on which many of these studies are based and the inherent limitations of their methodologies. In order to provide an overview, we sketch out the main approaches and present the key results.[3]

Survey results

Representative surveys of public opinion suggest that between 10 per cent and 30 per cent of the population in the CEECs have a general preference to migrate to the EU (see, for example, Fassmann and Hintermann, 1997; Wallace, 1998; Hospodárské Noviny, 2001). But only a small fraction of these people will actually move. By assessing the 'seriousness' of the answers, Fassmann and Hintermann estimate that the 'actual migration potential' will be around 2 per cent of the population of the CEECs. This figure does not sound unreasonable in the light of evidence from the German Socio Economic Panel, which shows that only 5 per cent of those East Germans who said they planned to migrate to western Germany in 1991 had actually moved there two years later (Büchel/Schwarze, 1994).

In general, opinion polls face three basic problems, which make it difficult to draw any quantitative conclusions as to the actual migration potential:

- First, they only provide information on the supply side: the propensity of workers to migrate, and not the demand side: the capacity of labour markets to absorb additional workers.
- Second, it is difficult to determine whether somebody who indicates a general propensity to migrate in an opinion poll has serious intentions to move.
- Third, migration from the East is largely a temporary phenomenon. This means that the proportion of the population that will move to another country and perhaps return within a certain *period* of time is much higher than the proportion that will live in a foreign country at a given *point* in time. As a consequence, observed migration stocks might be only a small fraction of the amount of people which have a migration experience in their lives.[4]

Because of these problems, the author of one of the most comprehensive surveys concludes that opinion polls are reliable in relative terms, but cannot be used to estimate migration levels

[3] For surveys of the literature see European Commission (2001), Quaisser et al. (2000).

[4] As an example, for the last five years, around 200,000 temporary workers from the CEECs have been employed each year in Germany with for an average of less than three months. Under the (unrealistic) assumption that each worker from the CEECs has been employed in Germany only once, this figure corresponds to total migration of one million workers in five years. Yet on average, approximately 40,000 temporary workers from the CEECs have been employed in Germany at each point in time during this period.

(Wallace, 1998). Nevertheless, it is possible to draw interesting conclusions as to the human capital characteristics of migrants from these surveys, since the socio-economic characteristics of actual migrants and those who reveal a preference to migrate seem to be similar (Büchel/ Schwarze, 1994). According to the study of Fassmann and Hintermann (1997), 12.2 per cent of the potential migrants possess a university degree, 30.7 per cent an A-level schooling degree (Abitur), 31.4 per cent a polytechnical schooling degree and only 13.7 per cent the minimum obligatory schooling degree. These education levels of potential migrants are well above the country averages (Fassmann/Hintermann, 1997).

Extrapolating South-North migration in the 1960s

Another way of estimating the migration potential is through extrapolation exercises, which take as their point of reference the migration flows from Southern Europe to the West and North European countries in the 1950s and 1960s, and the migration of Mexicans to the United States in the 1970s and 1980s (Layard et al., 1992; Lundborg, 1998; Bauer and Zimmermann, 1999). These studies conclude that less than 3 per cent of the population in the CEECs will migrate to the West within 15 years. This corresponds to an annual immigration of around 200,000 people from all CEECs (including the former Soviet Union) or 130,000 people from the Czech Republic, Hungary, Poland and Slovakia (Quaisser et al., 2000). Note that these figures refer to *gross* inflows, implying that *net* migration and migration stocks will be – at between 50 per cent and two-thirds of the gross figures – substantially lower.

Interestingly enough, the income differentials between the main sending and receiving countries of South-North migration in the 1960s are similar to the income differentials between Eastern and Western Europe today.[5] Moreover, although the free movement of workers was granted to only a few European countries in the 1960s and 1970s, the barriers to labour immigration were *de facto* removed in the main receiving countries during the period of ‘guest-worker’ recruitment program in the 1960s and early 1970s.

But there are also important differences between the conditions of South-North and East-West migration:

- First, labour markets were characterised by full employment and shortages of manual workers in the main receiving countries (Belgium, France, Germany and Switzerland) before the first oil price shock in 1973. Today, unemployment rates are still high in the main receiving countries of East-West migration in the EU.

- Second, the demographic structure of the population has changed: the share of young, mobile cohorts in the population of the CEECs is below that of the Southern European countries in the 1960s and will further decline at the beginning of the next decade.

- Third, the labour force in the CEECs is well educated relative to traditional sending countries.

[5] PPP-GDP per capita levels of the main sending countries ranged between 20 per cent (Turkey) and 55 per cent (Spain) of the main receiving countries of South-North migration in Europe (for example, Belgium, France Germany and Switzerland) in 1965 (Maddison, 1995), while PPP-GDP per capita levels of the CEECs are today between 20 per cent (Bulgaria) and 55 per cent (Slovenia) of the main receiving countries in the EU (Austria and Germany).

- Fourth, the transition process is not yet complete, so rates of structural change and job turnover are higher in the CEECs than in traditional sending countries.

These differences may affect the results of extrapolation exercises in both directions. But the exercises do give a hint as to the magnitude of the migration potential.

Explaining migration rates using econometric models

In order to overcome the shortcomings of simple back-of-the-envelope calculations, a number of econometric studies have tried to exploit the information from post-war migration episodes in Western Europe to assess potential migration after free movement of labour has been introduced. Most of these analyses derive from the traditional Harris-Todaro (1970) model, and explain migration flows or stocks by differences in per capita incomes and employment rates in the respective locations. Moreover, some models include dummy variables in an effort to capture different institutional conditions for migration such as the free movement in the EU or guest-worker recruitment in the 1960s and early 1970s. These studies have produced a wide variety of results and many suffer from several methodological problems.

There are basically three types of models in the literature:

- Cross-sectional regression models, which rely on the assumption that there is a static relationship between migration rates[6] and a set of explanatory variables such as income differentials and unemployment rates.

- Dynamic panel models that take account of the adjustment of migration rates or stocks to changes in the explanatory variables such as income differentials and unemployment rates.

- Error-components models, which 'explain' migration by a set of country-specific and time-specific error components.

Gravity and other cross-sectional regression models

The starting point for the first strand of the literature are results of a number of cross-country studies reported in Barro and Sala-i-Martin (1995), which regresses the migration balance against per capita GDP levels across regions in Europe, Japan and the United States. In this type of model, net migration rates in regions or countries are regressed against GDP per capita levels and some control variables such as temperature and population density for a given point of time. Using the coefficients from Barro and Sala-i-Martin (1995) and a study by Hatton and Williamson (1992) cited there, Franzmeyer and Brücker (1997) estimate the annual net inflow from the CEEC-5 (the Czech Republic, Hungary, Poland, Slovakia and Slovenia) into the EU-15 at 340,000-580,000 people for the per capita GDP levels of 1996. Similar exercises have been

[6] Some of these studies refer also to the migration stock as the dependent variable.

undertaken for Austria by Hofer (1998) and Walterskirchen and Dietz (1998).[7] This simple approach suffers from three shortcomings:

- First, variables that reflect economic factors beyond income differentials (e.g. employment opportunities) are omitted.

- Second, models based on cross-sectional regressions neglect the adjustment of migration to income differentials that takes place over time.

- Third, the migration balance of regions comprises both national and international migration. Since the elasticity of national migration is higher than the elasticity of international migration, using the coefficients from the Barro and Sala-i-Martin (1995) study tends to exaggerate the estimates of the migration potential significantly (see Straubhaar, 1998, and Alecke, Huber and Untiedt, 2000, for a discussion). This does not, however, hold true for the coefficients from the Hatton and Williamson (1992) study, which refers solely to international migration. Nevertheless, the Hatton and Williamson study refers to international migration in the late 19th century, where the elasticities of international migration with respect to differences in income differences have been substantially higher than in the post-war period.

Building on the static approach, other studies have estimated so-called 'gravity models', that is, migration stocks or net migration rates have been regressed against a set of explanatory variables such as per capita GDP levels, (un-)employment rates, population size and distance for a cross-section of sending and receiving countries. On the basis of such a model, Hille and Straubhaar (2001) estimate that for the present income differential, around 340,000 people will migrate from the CEEC-10 to the EU-15, which corresponds to a cumulative figure of around four million people or 4 per cent of the population in the CEEC-10 within 15 years under reasonable assumptions about the convergence of per capita income levels. Note that these figures refer to gross migration flows so that net migration flows will be substantially lower.

On the basis of a model similar to that of Hille and Straubhaar (2001), Orlowski and Zienkowski (1999) estimate that between 1.2 per cent and 3.8 per cent of the Polish population will migrate to the EU within 15 years, depending on different assumptions about the convergence of per capita GDP levels.

Dynamic models

Several dynamic models have been estimated for panels of source countries in order to consider the adjustment of migration to a change in the relevant variables. Dynamic models have the advantage that they make it possible to model the adjustment of migration to changes in the relevant variables. Note that the change of location involves fixed set-up costs, such that economic theory treats migration as an investment in the productive use of human resources. Thus, just like other irreversible investments under uncertainty, the decision to exercise the option to migrate must be based on the option value of waiting for further information (Burda,

[7] Using the same coefficients as Franzmeyer and Brücker (1997), Hofer (1998) expects net migration of 25,000-40,000 people a year into Austria. Relying on other coefficients, Walterskirchen and Dietz (1998) estimate the annual net migration of workers (including commuters) from the CEEC-5 into Austria at 42,000 people in 2005.

1995). Uncertainty not only increases the opportunity costs of migration, but also implies a sluggish adjustment of migration to a change in economic and institutional variables. Neglecting these adjustment processes may therefore bias the results.

Unfortunately, the estimation of dynamic models is seriously hampered by a lack of available data. Most EU Members do not report migration flows and stocks on an annual basis for longer time periods. Most dynamic models therefore refer to German data, where migration stocks and flows are reported by country of origin since 1967. Moreover, estimating dynamic models involves several methodological problems:

- First, the time series analysis of migration requires that the data satisfy certain statistical properties. Since the seminal paper of Nelson and Plosser (1982), who showed that most of the US macroeconomic variables can be better described as difference-stationary ("unit root“) processes rather than (trend-)stationary processes, it is generally acknowledged that it is important to take this distinction into consideration in order to avoid the “spurious“ regression effects as pointed out by Granger and Newbold (1974). However, in the empirical literature on migration the time series properties of the data at hand are not generally addressed at all.[8]

- Second, migration is not only affected by variables that vary over time such as income differentials and employment rates, but also by some constant factors such as geographical distance, language and culture, which affect the pecuniary and non-pecuniary costs of migration. All these variables are usually captured by a country-specific constant term (“fixed effect”) in dynamic panel models, since they cannot be included individually due to perfect collinearity. However, in out-of-sample forecasts, the problem of how to deal with the country specific effects arises. Two approaches are applied in the literature: first, many studies use a common constant for all countries assuming that time-invariant country specific factors are irrelevant (Hille/Straubhaar, 2001; Sinn et al. 2001; Flaig, 2001). Omitting the country specific variables, may however seriously bias the results (see Hsiao, 1986; and Alecke, Huber and Untiedt, 2001; Fertig and Schmidt, 2000; Brücker, 2001, for a discussion in the migration context). The second approach in the literature applies a two-stage procedure (Fertig, 2000): In the first stage, the dynamic model is estimated with fixed effects, and in the second stage, the fixed effects are explained by language, geography and other time-invariant variables. It turns out that the consideration of fixed effects has an important impact on the quantitative results.

- Third, differences in migration behaviour across countries may affect not only the fixed effects, but also the slope parameters. As a consequence, the coefficients can be biased and the migration behaviour from the CEECs may deviate from that in the sample on which the estimates are based.

- Fourth, the estimation of dynamic models with fixed effects may involve an estimation bias, but one that declines with the number of observations over time (Nickell, 1981). Depending

[8] The paper of Hatton (1995) is a notable exception.

on the time-dimension of the respective studies, this bias might distort the results to a different extent.[9]

Based on a flow model of migration into Germany, Fertig (2000) estimates potential migration from the CEEC-10 into Germany at between 35,000 and 70,000 persons per annum. The Fertig study considers a fixed effects model and uses a theoretical framework which states that a long-run equilibrium relation between migration flows and the explanatory variable exists (Hatton, 1995). It does not test whether the data fulfill the appropriate stochastic properties that are necessary for the existence of such an equilibrium.

Building on the approach of Fertig (2000), the study of the European Integration Consortium finds that the hypothesis that a cointegration relation between the migration flows and the explanatory variables such as GDP per capita and employment rates exists is rejected (Boeri/Brücker et al., 2001). However, the data suggest that a cointegration relationship between migration stocks and the explanatory variables does exist. The study considers a fixed effects model, and for the projection of the migration potential uses the two-stage procedure proposed by Fertig (2000). Boeri, Brücker et al. (2001) estimate that between 1.9 per cent and 3 per cent of the population in the CEEC-10 will migrate to Germany in the long run, while the initial growth of the foreign population from the CEECs after the introduction of free movement is estimated at between 200,000 and 300,000 people. An extrapolation of these results to the EU-15 gives a long-run migration potential of between 3.5 per cent and 4.5 per cent of the population in the CEECs, and an initial growth of the foreign population of between 300,000 and 450,000 people in the EU-15 immediately after free movement of labour has been introduced for the CEEC-10. These results have turned out to be relatively robust with regard to changes in the specification of the model and the selection of the sample. Moreover, the dynamic forecast quality of the estimates within the sample is satisfying.

In another study, like the European Integration Consortium the IFO-institute estimates a stock model based on the German data but does not allow for country specific fixed effects (Sinn et al., 2000; Flaig, 2001). The estimation results have produced implausibly high coefficients for long-run migration potentials: for the present income differential of the CEECs, steady-state migrant stocks are estimated at 8 per cent and 11 per cent of the population from the CEECs for Germany alone; and the upper limit of the 95 per cent confidence interval exceeds 50 per cent of the population. However, the hypothesis that a common constant applies for all countries is rejected by all statistical tests such that the estimation results are probably seriously biased. Long-term forecasts in particular can thus not be based on these estimates (Brücker, 2001).

Error-components models

All migration models discussed above impose restrictions on the data, which are necessary in order to identify the coefficients. Fertig and Schmidt (2000) argue that most of these restrictions are not realistic. As an alternative to models that assume migration to be determined by economic factors such as income variables and employment rates, Fertig and Schmidt (2000) estimate

[9] The standard approach to eliminating the bias by instrumented estimation (Anderson/Hsiao, 1981, Arellano/Bond, 1991) has been shown to reduce the forecasting quality within the sample, and thus the non-corrected model has been used for the forecasts (Brücker, 2001).

therefore an error-components model, which explains migration solely by an overall intercept, a random country-specific component and a time-specific component. Moreover, adjustments have been made for changes in the demographic structure of the population. The estimation is based on net migration rates from 17 European countries into Germany.

Fertig and Schmidt expect an average net migration of 15,000-18,000 people a year from four CEECs (Estonia, the Czech Republic, Hungary and Poland) in a baseline scenario, and 49,000-63,000 people according to a "high immigration" scenario. The latter scenario is derived by adding one standard deviation to the estimated coefficients values of the country-specific components. In the first case, the cumulative influx numbers 300,000-400,000 people after 20 years, in the second case, 900,000-1.2 million people. Note that an extrapolation of the figures from the high-migration scenario to the CEEC-10 yields around two-thirds of the corresponding figures in the Boeri/Brücker et al. (2001) estimate.

The approach of Fertig and Schmidt (2000) does not consider any institutional and economic variables (beyond the demographic structure of the population), so it barely answers the question of how an institutional change such as introducing free movement to a number of countries with relatively low incomes may affect migration.

5 Estimation and evaluation of the forecasting performance

Although the empirical literature on potential migration from the CEECs into the incumbent EU Member States tend to converge to estimates of the long-run migration potential between 2 and 4 per cent of the population, the uncertainty surrounding these results is still high. Both from a policy perspective as well as from a methodological perspective, the long-run elasticity of migration with regard to income differences and other explanatory variables thus remains an important unresolved issue. The purpose of this section is to analyze to what degree the differences in the estimates of the long-run elasticities, and, hence, the long-run migration potential, can be attributed to the manner in which a given body of data is analysed by different econometric models. More specifically, we address three questions:

(1) How does the elasticity of migration with respect to income and employment variables differ depending on the estimation approach?
(2) Which estimators yield a better forecasting performance?
(3) What are the quantitative implications of the different estimators for potential migration in the short-run and in the long-run?

The two-dimensional nature of panel data sets – i.e. the time dimension within cross-sections and the dimension between cross-sections – offers a wide variety of estimation options. The researcher can exploit the variation of the data between countries (cross-sections), between different time periods, and both. However, any estimation procedure imposes restrictions on the data which may affect the quantitative results. Our objective is to compare the performance of various estimators in order to assess whether they are appropriate for forecasting purposes.

We use two data sets for the analysis. The first data set is based on the migration to Germany from a panel of 19 source countries during the period 1967 to 2001. The second data set is derived from the European Labour Force Survey and covers foreign employment in the EU-15 countries from 20 source countries during the period 1993 to 2001. The latter data set is – due to numerous missing observations and low response rates – an unbalanced panel and comprises 215 cross-sections. Thus we have one data set with relatively large time dimension and a small number of cross-sections (countries), and another data set with a large number of cross-sections and a small time dimension. The data are described in Annex 1.

In the remainder of this section, we first discuss the theoretical background and specify the empirical model (Section 5.1). We then describe the implications of different estimation procedures (Section 5.2). Finally, we present the estimation results and assess the forecasting performance within the German sample and the European sample (Section 5.3).

5.1 Model specification

The specification of the empirical model used here is consistent with a number of theoretical models based on the so-called human capital approach (Sjaastad, 1962; Harris/Todaro 1970; Banerjee/Kanbur 1981; Hatton, 1995). These have been discussed extensively in the migration literature, so we shall be brief here.

The decision to migrate is understood as an investment in human capital, whose returns are determined by the net present value of expected income streams in the future (Sjaastad, 1962). The costs of migration comprise not only the pecuniary costs of changing the place of residence, but also non-pecuniary costs including all social and psychological costs which result from moving into an unfamiliar environment. An individual will migrate if the expected benefits from moving exceed the expected costs.

Expectations on income in the country of destination are conditioned by the opportunity to find a job on its labour market. Following Harris and Todaro (1970), the average employment rate serves as a proxy for the individual probability to find a job. Similar arguments apply to the expectations on future income in the home countries. Uncertainty on future income levels may hamper migration, even if migrants are risk neutral. Following Hatton (1995), we assume that the uncertainty is in regard to the risk of unemployment rather than to different wage levels. This implies that we expect the coefficients for the employment rates to be higher than those on the wage variables.

The major part of the migration literature assumes a static relation among migration rates, i.e. the share of migrants in the source countries, and explanatory variables such as income and employment differences exists. This hypothesis relies implicitly on the assumptions that individuals are homogenous and that the same decision situation is replicated over time. Here in contrast, we assume that individuals are heterogeneous, that is, that they differ in their preferences and the human capital characteristics that are relevant for the costs and returns of migration. As a consequence, the propensity to migrate declines the higher the share of the population already living abroad. For a given differential in expected income, the stock of migrants will eventually reach a steady state, where its growth is solely determined by the natural rate of population growth and the rate of regularisations. This does not rule out the possibility that chain and network effects affect migration positively. But in the long run, these effects are dominated by declining preferences to migrate in the population. Thus, we understand migration as a disequilibrium phenomenon, which eventually ceases when the equilibrium stock of migrants is achieved.[10] This implies that net migration rates may fall to zero even in cases where non-negligible differences in wages and employment opportunities between countries exist. The proposition that an equilibrium relationship between migration stocks and the explanatory variables exists is tested below.

Given these considerations, we model the migration function as follows:

$$mst_{fht} = f\left(w_{ft}, w_{ht}, e_{ft}, e_{ht}, P_{ht}, Z_{fh}\right), \qquad (1)$$

where mst_{fht} is the share of migrants from country h residing in country f in per cent of the home population, w denotes wage, e is the employment rate and P_h is population in the home country. (In case of the large country sample the variable mst_{fht} reflects the share of workers residing in country f in per cent of the home labour force, and the variable P_h is the home labour force.) The subscript f denotes the foreign country (f = 1 in the German sample and f = 1, …, 15, in the sample based on the European Labour Force Survey) and the subscript h the home country (h = 1, …, 19, in the German sample, and h = 1, …, 20, in the sample based on the European Labour

[10] Note that this is consistent with the traditional Harris/Todaro (1970) model.

Force Survey). The subscript t denotes the year (t = 32 in German sample and t = 8 in the sample based on the European Labour Force Survey). Z_{fh} denotes a vector of time-invariant variables which affect the migration between two countries such as geographical proximity and language.

We assume that the adjustment process can be specified in form of a simple habit-persistence model, i.e. as

$$mst_{fht} - mst_{fh,t-1} = \delta\left(mst^*_{fht} - mst_{fh,t-1}\right) + u_{fht}, \tag{2}$$

where mst_{fht}^* is the share of migrants which desire to reside in the foreign country under the given economic and other conditions. This share is given by

$$mst^*_{fht} = \alpha^* + \beta^*_1 \ln\left(w_{ft}/w_{ht}\right) + \beta^*_2 \ln\left(w_{ht}\right) + \beta^*_3 \ln\left(e_{ft}\right) + \beta^*_4 \ln\left(e_{ht}\right) + \beta^*_5 \ln\left(P_{ft}\right) + Z'_{fh}\gamma^*. \tag{3}$$

This specification can be motivated by the following considerations: Firstly, we follow Hatton (1995) and assume that utility form (expected) is logarithmic, which yields the semi-logarithmic functional form of the model. Of course we can also conceive other specifications of the utility function. However, we observed that the semi-log specifications perform better than log-log specifications of the migration model in terms of the forecasting error within sample.[11] Secondly, we assume that liquidity constraints may limit migration such that the level of home income can positively affect the migration share for a given income difference between the host and the home country. Moreover, taking uncertainty into consideration, income in the host country and income in the home country may enter the model in a non-symmetric way (see e.g. Faini/Venturini, 1994). This implies that the sign of the home wage can be negative. In addition to the income differential, we therefore included home income as an additional variable. Thirdly, we included the population (labour force) of the host country in order to control for the absorptive capabilities of the labour markets in the countries of destination. We included the variable in logarithmic form since the relation between country size and migration is assumed to be non-linear, a presumption which is proved by the results of our regressions.[12]

Substituting equation (3) into (2) yields the following dynamic migration model:

$$\begin{aligned} mst_{fht} = \alpha + (1-\delta)mst_{fh,t-1} + \beta_1 \ln\left(w_{ft}/w_{ht}\right) + \beta_2 \ln\left(w_{ht}\right) \\ + \beta_3 \ln\left(e_{ft}\right) + \beta_4 \ln\left(e_{ht}\right) + \beta_5 \ln(P_{ft}) + Z'_{fh}\gamma + u_{fht}, \end{aligned} \tag{4}$$

where

$\alpha = \delta\alpha^*, \gamma = \delta\gamma^*,\ \beta_j = \delta\beta^*_j$ for $j = 1,...,5$, and $\beta_1 > 0,\ \beta_2 > 0$ or $\beta_2 < 0,\ \beta_3 > 0,\ \beta_4 < 0,\ \beta_5 > 0.$

Finally, we follow the usual convention and specify the error term as a one-way error-component model (Hsiao, 1986):

$$u_{fht} = \mu_{fh} + \nu_{fht}, \tag{5}$$

where μ_{fh} denotes a country-specific effect and ν_{fht} is white noise.

[11] The results are not reported here but available from the authors upon request.

[12] Both the significance of the labour force variable and the explanatory power of the model is higher if the labour force variable enters the model in logarithmic form. The results are not reported here but are available from the authors upon request.

The model in equation (4) forms the basis for our analysis. We assume that it is general enough to allow us to compare various estimators. In the German data sample, we include a further lag of the endogenous variable in order to impose less restrictions on the adjustment process. Further lags of the dependent variable have turned out to be insignificant. In the European sample, we restricted the number of lags to one in order to lose no further time-series observations.

5.2 Discussion of the estimators

The two-dimensional nature of panel data sets allows us to exploit both the variation between countries and time periods in the data for the estimation of the parameters of the migration function. Different estimators applied in the literature use different sources of variation and, hence, yield different results. Here we employ a wide variety of estimators which comprises both estimators which have been already applied in the estimation of migration functions as well as estimators which have not yet been used, but are discussed in the recent literature on dynamic panel estimation. Applying such a variety of different estimators allows us to provide an answer to the methodological question on to which extent the estimated coefficients of the migration functions vary across different estimation procedures as well as to examine the relative (in-sample) forecasting performance of different estimators.

The estimators we employ can be grouped as follows:

1. *Traditional estimators*: pooled OLS, WITHIN, seemingly unrelated regressions (SUR)[13], and several versions of the random effects (GLS) estimators such as those suggested in Wallace and Hussein (1969), Swamy and Arora (1972), and the iterated GLS (Maximum Likelihood) estimator.
2. *Instrumental variable estimator*: the Anderson and Hsiao (1982) estimator for dynamic panels.
3. *GMM estimators*: The GMM-type estimators advocated by Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1998) for dynamic panels.

The pooled OLS estimator employs both sources of variation (time and cross-sectional) in the data, albeit not efficiently. The main feature of the pooled OLS estimator lies in its uniform treatment of all cross-sections. In other words, it neglects the individual heterogeneity amongst the different countries such as culture, language, geography etc. that may have an important impact on migration. On the other hand, the WITHIN and SUR estimators allow for differences amongst the cross-sections captured by the individual specific intercept term. In so doing, these estimators focus only on the within variation in the data as the name of the former estimator suggests. The difference between these two estimators lies in different restrictions imposed on the covariance matrix of the error term. The WITHIN estimator imposes the same covariance matrix for all cross-sections, whereas the SUR estimator relaxes this assumption by allowing for different covariance matrices across the cross-sections with possible correlation between the individual specific error terms.

---

[13] We use a SUR estimator where all slope parmaters are restricted to be equal.

When the model assumptions of the GLS estimators are fulfilled, the GLS estimators are superior in terms of efficiency when compared to the other traditional types of estimators. This is achieved by the optimal weighting attached to the within and between variations in the data. The difference between the various versions of the GLS estimators as in Wallace and Hussein (1969), Swamy and Arora (1972), and the iterated GLS estimator, which is equal to the Maximum-Likelihood estimator, corresponds to different methods of calculating these optimal weights. Below we refer to these GLS estimators as GLS(WALHUS), GLS(SWAR), and GLS(MLE), respectively. Although all these GLS estimators are asymptotically equivalent, in the relatively small samples used here, the estimated coefficients are likely to differ.

In dynamic panel models, which model sluggish adjustment by considering lags of the dependent variable, the traditional estimators are subject to a simultaneous equation bias. This bias is caused by the presence of the lagged dependent variables amongst the explanatory variables. As pointed out in Baltagi, Griffin, and Xiong (2000), this simultaneity bias can arise for the following reasons:

- presence of the individual specific effects, which will definitely be correlated with the lagged dependent variable;
- autocorrelation in the error term of the migration function that results in correlation between the lagged dependent variable and the regression disturbances.

For the fixed size of the time dimension, both the OLS and GLS estimators will yield biased and inconsistent results for either of the two reasons. At the same time, the WITHIN and SUR estimators will be subject to this simultaneity bias only for the latter reason, as these estimators are based on the data transformation that wipes out the individual effects. Nickell (1981) provided the original discussion of the problem and calculated the values that characterise this bias. It is proved that this bias disappears as the time dimension of a panel grows. As a result, the WITHIN and SUR estimators yield credible results for dynamic panels with a relatively large time dimension. Also we expect the elimination of the bias from the GLS estimator to the extent that it emphasises the importance of the within variation in the data.

The rest of the estimators considered in this exercise were initially suggested to circumvent the endogeneity problem caused by the presence of the lagged dependent variable by the instrument variables techniques such as in Anderson and Hsiao (1982) or the Generalised Method of Moments (GMM) technique discussed in Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1998). Historically, Anderson and Hsiao (1982) were the first to suggest a solution to the simultaneous equation bias in the dynamic panel models (henceforth referred to here as the AH estimator). They suggested the following two-step procedure. First, remove the individual effects from the regression by taking the first difference of all the variables. This procedure eliminates the first likely cause of the endogeneity of the lagged dependent variables. The second source of the endogeneity is tackled by instrumenting the (differenced) lagged dependent variables with its own further lagged differences or lagged levels. Arellano and Bond (1991) expanded on the work of Anderson and Hsiao (1982) by pointing out that within the GMM framework there are many more instruments available than used by Anderson and Hsiao (1982). Consequently, the efficiency of the GMM based estimators is greatly enhanced. As Baltagi et. al. (2000) points out, the estimators of Anderson and Hsiao (1982) and Arellano and Bond (1991) may eliminate the simultaneous equation bias, but with a large loss of information.

The application of the first difference transformation destroys the economic structure formed between the levels of the variables across the time series dimension.

Fortunately, the GMM estimator developed by Arellano and Bover (1995) addresses this issue by employing both the first differences as well as the levels equations by specifying the appropriate sets of instruments for either types of equation. Blundell and Bond (1998) reports both Monte Carlo and empirical results which indicate the considerable efficiency improvements achieved by the Arellano and Bover estimator over that of Arellano and Bond (1991). For later use we denote the Arellano and Bond (1991) and Arellano and Bover (1995) estimators as GMM(DIF) and GMM(SYS), respectively, as the former is based on the first *difference* transformation and the latter is based on the *system* of equations specified both for the first difference as well as for the levels of the variables.

Some further comments regarding the use of the GMM estimators are necessary. Firstly, they are carried out as the one- and two-step estimators. The one-step GMM estimator is asymptotically inefficient when compared to the two-step estimator. The standard errors for the two-step estimator are reported after the small sample correction derived in Windmeijer (2000). Adequacy of the GMM estimators is checked by the Sargan test for validity of the employed moment restrictions as well as the autocorrelation tests derived in Arellano and Bond (1991). Lastly, the number of the instruments and hence the moment conditions increases very rapidly as the time dimension grows. In order to avoid the dimension problem, one has to consider only a subset of all available instrumental variables when the time dimension is rather large. Last but not least, in contrast to the GMM(DIF), the GMM(SYS) estimators allow inclusion of the time-invariant variables. When this is the case, we refer to them as GMM(SYS)-TINV for the German sample and do not use the special notation for the GMM(SYS) estimators when considering the European sample.

In summary, given the different properties of the estimators available, we would expect that for the panels with rather small time series dimensions and rather large cross-sectional dimensions, the GMM estimator of Arellano and Bover (1995) will be superior to the rest of the estimators discussed above. In particular, we would expect the traditional estimators to perform rather poorly due to the unresolved simultaneous equation bias. When the opposite is the case, i.e. relatively large time series and rather small cross-sectional dimensions of the panel, the motivation for using the GMM estimators is less obvious as these were primarily designed for the former case. In addition, the inflation of the moment conditions may somewhat worsen their performance in panels typically considered in macroeconomic studies. On the other hand, either the WITHIN or SUR estimator is expected to have a comparative advantage over the rest of the estimators as the Nickell bias is likely to be of considerably smaller magnitude than in panels with a small time dimension. Indeed, the Monte Carlo study of Judson and Owen (1999) supports this view.

5.3 Results[14]

5.3.1 German Sample

First consider the estimation results for the German sample. These panel data are characterised by a relatively large time dimension and a rather small cross-sectional dimension. Hence, we expect the WITHIN and SUR estimators to perform rather well in this sample. The estimated coefficient values for the short-run semi-elasticities are reported in Table 5.1.[15] First consider the results for the traditional estimators. The coefficient estimates for the lagged dependent variable, foreign-to-home wage ratio, home wage, and German employment are well determined and have the expected signs. At the same time, the coefficients of the home employment variable have the expected (negative) signs for most cases but appear to be estimated with less precision for all but the SUR and SWAR estimators. This replicates the results of many empirical studies, which find that (un-)employment rates in the home country or region have no significant impact on migration propensities.

Notice that for the sake of presentation we report the sum of the coefficients of the first- and second lags of migration stock variables with the associated standard error. In all cases but one - GLS(SWAR) - the sum of the coefficients is less than unity, implying the dynamic stability of the models. On the basis of the value of the coefficient estimate on the lagged migration stock we can group the estimators as follows: OLS, GLS(WALHUS), GMM1(SYS), GMM2(SYS) both with and without time-invariant variables produce a value very close to one, implying a very high degree of persistence of the underlying time series. The Arellano and Bond (1991) estimator GMM(DIF) produces the lowest value of the coefficient on the lagged dependent variable of slightly above 0.8. The WITHIN, SUR, and GLS(MLE) produce values of around 0.9, which lies between these two extremes. Note that in order to save space, we do not report the estimation results for the Anderson and Hsiao (1982) estimator due to its very poor performance already at the estimation stage, reflecting the dynamically explosive model and implausible estimated coefficient values.

There are several comments that can be made at this juncture. First, the fact that the OLS estimates for the lagged dependent variable are higher than those for the WITHIN estimator implies the presence of the individual specific effects that cannot simply be ignored. Second, the similarity between the OLS and GLS(WALHUS) coefficient estimates could be explained by the fact that an optimal weighting for the GLS(WALHUS) is based on the OLS residuals, see Doornik, Arellano, and Bond (2002). Third, the similarity between the WITHIN and GLS(MLE) estimates could be explained by the fact that for the importance of the WITHIN variation in the GLS optimal weighting scheme increases with the growing time dimension, see Baltagi (1995).

[14] The OLS, WITHIN, and SUR estimators we employ were obtained in Eviews 4.1, while the GLS, instrumental variable, and GMM estimators were obtained in the DPD Ox package, see Doornik, Arellano, and Bond (2002).

[15] The full regression output is relegated to the appendix in the interests of presentation flow and saving space.

Table 5.1 **Regression Results: Short Run Semi-Elasticities of Migration Stock (German Sample, 1969-2001)**

Model Type	$mst_{h,t-1} + mst_{h,t-2}$	$\ln(w_{ft}/w_{ht})$	$\ln(w_{ht})$	$\ln(e_{ft})$	$\ln(e_{ht})$	$FREE_{ht}$	$GUEST_{ht}$
OLS	0.99 **	0.18 **	0.14 **	1.09 **	-0.09	0.00	0.08 *
	(0.01)	(0.04)	(0.04)	(0.24)	(0.09)	(0.01)	(0.04)
WITHIN	0.88 **	0.35 **	0.18 **	1.08 **	-0.18	0.00	0.06 *
	(0.03)	(0.07)	(0.03)	(0.24)	(0.11)	(0.01)	(0.03)
SUR	0.89 **	0.25 **	0.14 **	0.74 **	-0.10 **	0.00	0.05 **
	(0.01)	(0.02)	(0.02)	(0.10)	(0.01)	(0.00)	(0.01)
GLS(WALHUS)	0.98 **	0.20 **	0.14 **	1.10 **	-0.12	0.00	0.08 **
	(0.01)	(0.04)	(0.04)	(0.25)	(0.10)	(0.01)	(0.02)
GLS(SWAR)	1.002 **	0.11 **	0.10 **	0.94 **	0.05 **	0.00	0.00
	(0.00)	(0.04)	(0.04)	(0.26)	(0.02)	(0.00)	(0.01)
GLS(MLE)	0.90 **	0.35 **	0.17 **	1.07 **	-0.16	0.00	0.06 **
	(0.01)	(0.05)	(0.04)	(0.24)	(0.13)	(0.01)	(0.02)
GMM1(DIF)	0.81 **	0.86 **	0.24 *	1.29 *	-0.41	0.07	-0.40 **
	(0.03)	(0.29)	(0.10)	(0.53)	(0.59)	(0.07)	(0.07)
GMM2(DIF)	0.81 **	0.58	0.24 *	1.51 *	-0.78	0.10	-0.40 **
	(0.03)	(0.35)	(0.11)	(0.58)	(0.81)	(0.11)	(0.07)
GMM1(SYS)	0.98 **	0.24 **	0.18 *	1.40 *	-0.14	0.00	0.11 **
	(0.01)	(0.10)	(0.08)	(0.55)	(0.17)	(0.01)	(0.04)
GMM2(SYS)	0.98 **	0.23 **	0.17 *	1.13 **	0.01	0.00	0.11 *
	(0.01)	(0.10)	(0.08)	(0.44)	(0.16)	(0.01)	(0.06)
GMM1(SYS)-TINV	0.97 **	0.26 **	0.17 *	1.37 *	-0.11	0.02	0.10 *
	(0.01)	(0.11)	(0.07)	(0.54)	(0.18)	(0.01)	(0.04)
GMM2(SYS)-TINV	0.98 **	0.24 **	0.16 *	1.11 *	0.03	0.01	0.10 *
	(0.01)	(0.10)	(0.08)	(0.51)	(0.28)	(0.01)	(0.06)

In parentheses the heteroscedasticity robust standard errors are reported for all but GMM2-step estimators.
The standard errors for GMM 2-step estimators are reported after smalll sample correction suggested in Windmeijer (2000).

Numbers in parenthesis denote standard errors. **,*: rejection of the null hyptohesis at the 1% and the 5% level, respectively.
Apart from self-explanatory OLS, WITHIN, and SUR estimators, GLS(WALHUS) - Wallace and Hussein (1969) estimator, GLS(SWAR) - Swamy and Arora (1972) estimator, GLS(MLE) - iterated GLS, GMM1(DIF) and GMM2(DIF) - one- and two-steps Arellano and Bond (1991) estimator, GMM1(SYS) and GMM2(SYS) - one- and two-steps Arellano and Bover (1995) estimator, GMM1(SYS)-TINV and GMM2(SYS)-TINV - one- and two-steps Arellano and Bover (1995) estimator with time invariant variable included.

The high persistence in the dependent variable implied by the OLS, GLS(WALHUS), GMM1(SYS), GMM2(SYS) estimates implies rather high values for the long-run semi-elasticities when compared to the more moderate values implied by the other estimators as reported in Table 5.2. In any case, these coefficient estimates are the partial effects that capture the response of the dependent variable to the changes in the explanatory variables. In order to assess the long-run implications of the estimated models, all the available information needs to be taken into consideration. In particular, values resulting from individual effects should not be ignored.

Table 5.2 **Regression Results: Long Run Elasticity of Migration Stock (German Sample, 1969-2001)**

Model Type	$\ln(w_{ft}/w_{ht})$	$\ln(w_{ht})$	$\ln(e_{ft})$	$\ln(e_{ht})$
OLS	14.20 *	11.09	87.49	-7.53
	(7.04)	(6.28)	(56.67)	(9.80)
WITHIN	2.84 *	1.44 *	8.86 *	-1.49
	(0.98)	(0.43)	(3.18)	(1.14)
SUR	2.39 **	1.28 **	7.02 **	-0.98 **
	(0.18)	(0.16)	(1.09)	(0.14)
GLS(WALHUS)	9.57 **	6.96 *	53.97 *	-5.98
	(2.72)	(2.54)	(19.15)	(5.40)
GLS(SWAR)	-	-	-	-
GLS(MLE)	3.34 **	1.65 **	10.29 **	-1.58
	(0.51)	(0.37)	(2.57)	(1.26)
GMM1(DIF)	4.64 **	1.29	6.92	-2.19
	(1.21)	(0.67)	(2.99)	(3.09)
GMM2(DIF)	3.11 *	1.29	8.11 *	-4.17
	(1.59)	(0.71)	(3.85)	(4.69)
GMM1(SYS)	10.59 **	7.90 **	61.91 **	-6.22
	(2.74)	(2.71)	(21.45)	(6.41)
GMM2(SYS)	13.41 *	10.20 *	65.77 *	0.86
	(5.52)	(5.07)	(32.26)	(9.73)
GMM1(SYS)-TINV	8.74 **	5.60 **	45.48 **	-3.53
	(1.58)	(1.67)	(13.33)	(5.41)
GMM2(SYS)-TINV	11.23 **	7.75	52.63	1.23
(with time invariant variables)	(4.86)	(4.48)	(28.04)	(13.55)

In parentheses the heteroscedasticity robust standard errors are reported for all but GMM2-step estimators.
The standard errors for GMM 2-step estimators are reported after smalll sample correction suggested in Windmeijer (2000).

Numbers in parenthesis denote standard errors. **,*: rejection of the null hyptohesis at the 1% and the 5% level, respectively.
Apart from self-explanatory OLS, WITHIN, and SUR estimators, GLS(WALHUS) - Wallace and Hussein (1969) estimator, GLS(SWAR) - Swamy and Arora (1972) estimator, GLS(MLE) - iterated GLS, GMM1(DIF) and GMM2(DIF) - one- and two-steps Arellano and Bond (1991) estimator, GMM1(SYS) and GMM2(SYS) - one- and two-steps Arellano and Bover (1995) estimator, GMM1(SYS)-TINV and GMM2(SYS)-TINV - one- and two-steps Arellano and Bover (1995) estimator with time invariant variable included.

Table 5.3 summarises the predicted values for the dependent variable using the long-run coefficient values. The table entries have been calculated for the German data set and are based on the mean (median) of the range of the forecasted values computed for each cross-section. As seen, when compared to the actual values taken from the sample, most of the models display a rather small long-run increase in the foreign migration stocks measured as the percentage of the home population. The OLS results imply the largest long-run migration stocks whereas the GMM1(SYS)-TINV is on the opposite end. It is noticeable that the WITHIN, SUR, and GLS(MLE) estimators produce similar predicted values for the long-run migration stock variable both when measured in terms of mean and median across the range of the individual values calculated for every cross-section.

Table 5.3 **Predicted Long-Run Migration Stocks at the Sample Mean (Median)**
(German Sample, 1969-2001)

Model Type	mean	median
	long-run migration stock in % of home population	
OLS	1.43	0.89
WITHIN	1.00	0.42
SUR	1.01	0.42
GLS(WALHUS)	1.23	0.48
GLS(SWAR)[1)]	-	-
GLS(MLE)	1.01	0.42
GMM1(DIF)	0.92	0.33
GMM2(DIF)	0.96	0.43
GMM1(SYS)	1.19	0.47
GMM2(SYS)	1.37	0.67
GMM1(SYS)-TINV	0.66	0.08
GMM2(SYS)-TINV	0.84	0.23
actual value within sample	0.86	0.38

[1)] Explosive dynmaic behaviour.

For description of estimators see notes in Table 5.1.

As a final evaluation of the different estimators, we have calculated the Root Mean Squared Percentage Error (RMSPE) for static and dynamic evaluation of the in-sample forecasting performance of the models.[16] The static evaluation uses the actual lags of the dependent variables for the calculation of the fitted values from the regression results, while the dynamic evaluation uses the predicted values of the lagged dependent for the calculation of the fitted values. Thus, an estimation bias shows up more strongly in the dynamic evaluation than in the static evaluation. The results of the static and dynamic forecast evaluation are presented in Table 5.4.

The ranking of the estimators differs somewhat, with the GLS(WALHUS) estimator producing the lowest RMSPE for the static forecasts, and the SUR estimator producing the lowest for the dynamic simulation. Note that the OLS estimator, which omits the individual effects, performs the worst in the dynamic simulation. Given the fact that the SUR estimator also performs relatively well for the static forecast evaluation and, more importantly, keeping in mind that the main objective of the present study is to forecast the migration potential of the CEE countries via *dynamic* forecast simulations, we designate the SUR model as the model offering the best

[16] We could also have chosen to base our forecast evaluation procedures on the Root Mean Squared Error (RMSE). However, given that the dependent variable takes different range of values for the various cross-sections, we think that for our purposes, e.g. forecast evaluation in panel data, the RMSPE is more appropriate as it treats the forecast errors in different cross-sections on an equal footing.

forecasting performance. Thus, we select the SUR model as the main object of the current project discussed below.

Table 5.4 **Comparision of Static and Dynamic Forecasting Performance**
(German Sample, 1969-2001)

	Static Simulation		Dynamic Simulation	
Rank	Estimator	RMS Percent Error	Estimator	RMS Percent Error
1	GLS(WALHUS)	0.0920	SUR	0.4117
2	GMM2(SYS)	0.0944	GMM2(SYS)	0.4572
3	SUR	0.0947	GLS(WALHUS)	0.4837
4	GMM2(SYS)-TINV	0.0968	WITHIN	0.5262
5	GLS(SWAR)	0.1036	GLS(MLE)	0.5576
6	GMM1(SYS)	0.1075	GMM1(SYS)	0.5602
7	GMM1(SYS)-TINV	0.1129	GMM2(SYS)-TINV	0.6347
8	OLS	0.1170	GMM1(SYS)-TINV	0.8114
9	GLS(MLE)	0.1217	GMM2(DIF)	1.1196
10	WITHIN	0.1224	GLS(SWAR)	1.1248
11	GMM2(DIF)	0.3101	GMM1(DIF)	1.1783
12	GMM1(DIF)	0.3135	OLS	1.2591

For description of estimators see notes in Table 5.1.

5.3.2 European Sample

The European sample is an unbalanced panel of 15 destination countries and 20 source countries in the time period 1993 to 2001. The sample comprises a total of 215 cross-sections, since not all destination countries report observations for the whole set of source countries. Moreover, several time-series observations are missing. Altogether, this panel is characterised by a small time dimension and a relatively large cross-sectional dimension. Thus, we expect that the traditional estimators may yield seriously biased results in this sample.

Table 5.5 summarises the estimation results obtained from this sample. Observe that when reporting the estimation results we have omitted the employment variables from the migration function. This is done because in regressions where the employment variables have been included in the set of regressors, they turned out to be insignificant and sometimes had counterintuitive signs. The values of the coefficient to the lagged dependent variable are around 0.95 for the OLS, GLS, and GMM(SYS) estimators. On the other hand, the WITHIN, AH, GMM(DIF) yield values that are much lower. Judging from the reported standard errors, only the GMM(SYS) and OLS estimators seem to estimate the model parameters with a high degree of precision. For the other estimators, the estimated coefficients seem to have expected signs, with the exception of the AH and GMM(DIF), which yield the negative sign on the size of the labour force variable.

Table 5.5 **Semi-Elasticity of Migration Stock (European Sample, 1993-2001)**

	Short Run					Long Run			
	$lst_{h,t-1}$	$ln(w_{ft}/w_{ht})$	$ln(w_{ht})$	$ln(lab_{ht})$	$ADJACENT_h$	$ln(w_{ft}/w_{ht})$	$ln(w_{ht})$	$ln(lab_{ht})$	$ADJACENT_h$
OLS	0.95 **	0.03 **	0.04 **	0.01 *	0.03 **	0.71	0.77	0.16	0.56
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)				
WITHIN	0.59 **	0.23	-0.15	0.07		0.57	-0.38	0.18	
	(0.20)	(0.14)	(0.11)	(0.07)					
GLS(WALHUS)	0.95 **	0.03	0.04	0.01 *	0.03 *	0.71	0.77	0.16	
	(0.00)	(0.03)	(0.03)	(0.00)	(0.01)				
GLS(SWAR)	0.96 **	0.02 *	0.02	0.01 **	0.02 **	0.56	0.58	0.13	
	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)				
AH	0.64 **	0.08	0.37	-0.46		0.21	1.04	-1.28	
	(0.05)	(0.11)	(0.27)	(0.35)					
GMM1(DIF)	0.76 **	0.15	0.33	-0.45		0.61	1.39	-1.88	
	(0.06)	(0.15)	(0.24)	(0.35)					
GMM2(DIF)	0.73 **	0.09	0.19	-0.17		0.32	0.69	-0.65	
	(0.06)	(0.08)	(0.11)	(0.14)					
GMM1(SYS)	0.94 **	0.05 **	0.05 **	0.01 **	0.03 **	0.77	0.85	0.15	0.55
	(0.01)	(0.01)	(0.02)	(0.00)	(0.01)				
GMM2(SYS)	0.94 **	0.02 *	0.02	0.00 **	0.01 *	0.34	0.30	0.08	0.20
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)				

Numbers in parantheses denote standard errors. **,*: indicate significance at the 1% and the 5% level, respectively.
In parentheses the heteroscedasticity robust standard errors are reported for all but GMM2-step estimators.
The standard errors for GMM 2-step estimators are reported after smalll sample correction suggested in Windmeijer (2000).

Apart from self-explanatory OLS and WITHIN estimators, GLS(WALHUS) - Wallace and Hussein (1969) estimator, GLS(SWAR) - Swamy and Arora (1972) estimator, AH - Anderson and Hsiao (1982), GMM1(DIF) and GMM2(DIF) - one- and two-steps Arellano and Bond (1991) estimator, GMM1(SYS) and GMM2(SYS) - one- and two-steps Arellano and Bover (1995) estimator.

Given the estimation results and referring to the discussion above on the theoretical properties of the different estimators in the panels with relatively small time series- and rather large cross-sectional dimensions, it seems that the GMM2(SYS) estimator is superior in comparison to the other estimators (also to the OLS estimator despite its seemingly good performance). Observe that the specification tests reported in the appendix in form of the Sargan test of validity of the moment conditions and the test for the presence of the second-order autocorrelation support the correct specification of the model estimated using the GMM2(SYS) estimator. As noted in Arellano and Bond (1991), the consistency of the GMM estimator hinges on the absence of the second-order autocorrelation. Thus our empirical evidence seems to be in line with the theoretical considerations put forward for the Arellano and Bover (1995) estimator.

Table 5.6 compares the forecasting performance of the different estimators and reinforces the conclusions discussed with regard to Table 5.5. The Arellano and Bover (1995) estimator GMM2(SYS) offers superior forecasting performance, both in static and dynamic terms. Also observe that the calculated RMSPE are generally larger than those obtained for the German sample, and that there is much more dispersion between the reported RMSPE values for the European sample than for the German one. This implies that the results for the dynamic simulation are subject to a much greater degree of uncertainty and are therefore less reliable than those associated with the German sample. We interpret this as a strong indication that caution should be observed in making any predictions based on these estimates.

Table 5.6 **Comparision of Static and Dynamic Forecasting Performance**
(European Sample, 1993-2001)

	Static Simulation		Dynamic Simulation	
Rank	Estimator	RMS Percent Error	Estimator	RMS Percent Error
1	GMM2(SYS)	2.593	GMM2(SYS)	3.438
2	GLS(SWAR)	2.883	GLS(SWAR)	4.244
3	GLS(WALHUS)	3.939	GLS(WALHUS)	6.292
4	GMM1(SYS)	5.183	GMM1(SYS)	8.477
5	GMM2(DIF)	5.891	AH	11.233
6	GMM1(DIF)	7.744	GMM2(DIF)	12.103
7	OLS	8.132	GMM1(DIF)	16.051
8	AH	10.737	WITHIN	32.105
9	WITHIN	33.443	OLS	53.326

Apart from self-explanatory OLS and WITHIN estimators, GLS(WALHUS) - Wallace and Hussein (1969) estimator, GLS(SWAR) - Swamy and Arora (1972) estimator, AH - Anderson and Hsiao (1982), GMM1(DIF) and GMM2(DIF) - one- and two-steps Arellano and Bond (1991) estimator, GMM1(SYS) and GMM2(SYS) - one- and two-steps Arellano and Bover (1995) estimator.

6 Simulation of the migration potential

As we have seen in the preceding section, the traditional SUR and WITHIN estimators performed fairly well in the German sample with a relatively large time dimension of the data, while these estimators are clearly outperformed by the GMM-estimators in the European sample with the relatively large cross-sectional dimension. However, the poor forecasting performance of our model within the European sample raises doubts as to its usefulness for our purposes. Thus, in this section, we simulate the migration potential from the CEECs on the basis of the German sample and extrapolate the results to the EU-15. We also use the results from the European sample in order to determine whether they yield reasonable simulations. Note that the purpose of all our simulations is to obtain a general idea as to the magnitude of migration potential: our simulations should therefore not be misunderstood as exact forecasts. The differences we observe in the migration behaviour across countries are too large to allow for an accurate out-of-sample forecast.

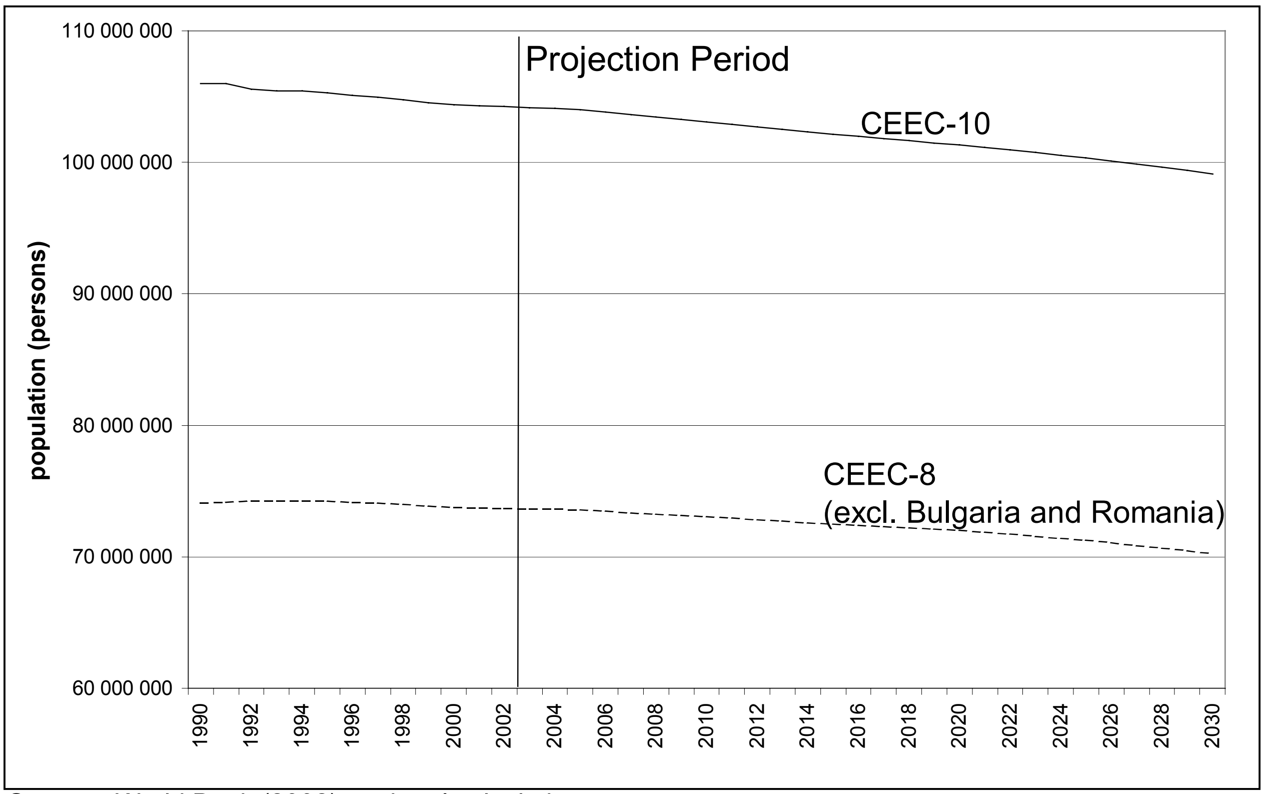
The remainder of this section is organised as follows: in section 6.1 we describe the assumptions of the different scenarios with respect to the exogenous variables. Section 6.2 presents the results of the explanation of the country-specific effects. The explanation of the country-specific effects in second-stage regressions is necessary in order to gain values of the individual effects for our out-of-sample simulations. Section 6.3 reports the simulation results for the different estimators under the assumptions of the baseline scenario. Section 6.4 analyses the sensitivity of the results with respect to the underlying assumptions on income convergence and employment opportunities, and extrapolates the results for Germany to the EU-15. Section 6.5 assesses the impact of transitional periods on the magnitude of the migration potential.

6.1 Assumptions on the exogenous variables

The simulation of the migration potential is based on the following assumptions on the development of the exogenous variables:

- **Population**: Our population scenario for the CEECs is derived from the projection of natural population growth in the World Development Indicators (World Bank, 2002). The population projection is based on the natural rate of population growth, i.e. we excluded the net migration balance, which is exogenous in our scenario. The population scenario is displayed in Figure 6.1: in the absence of migration, the population of the CEEC-10 is expected to decline from approximately 104 million in 2002 to 99 million in 2030, and the population of the CEEC-8 from 73 million to 70 million in the same time period.[17]

Figure 6.1 **Population Scenario for the CEEC-10 (without migration)**

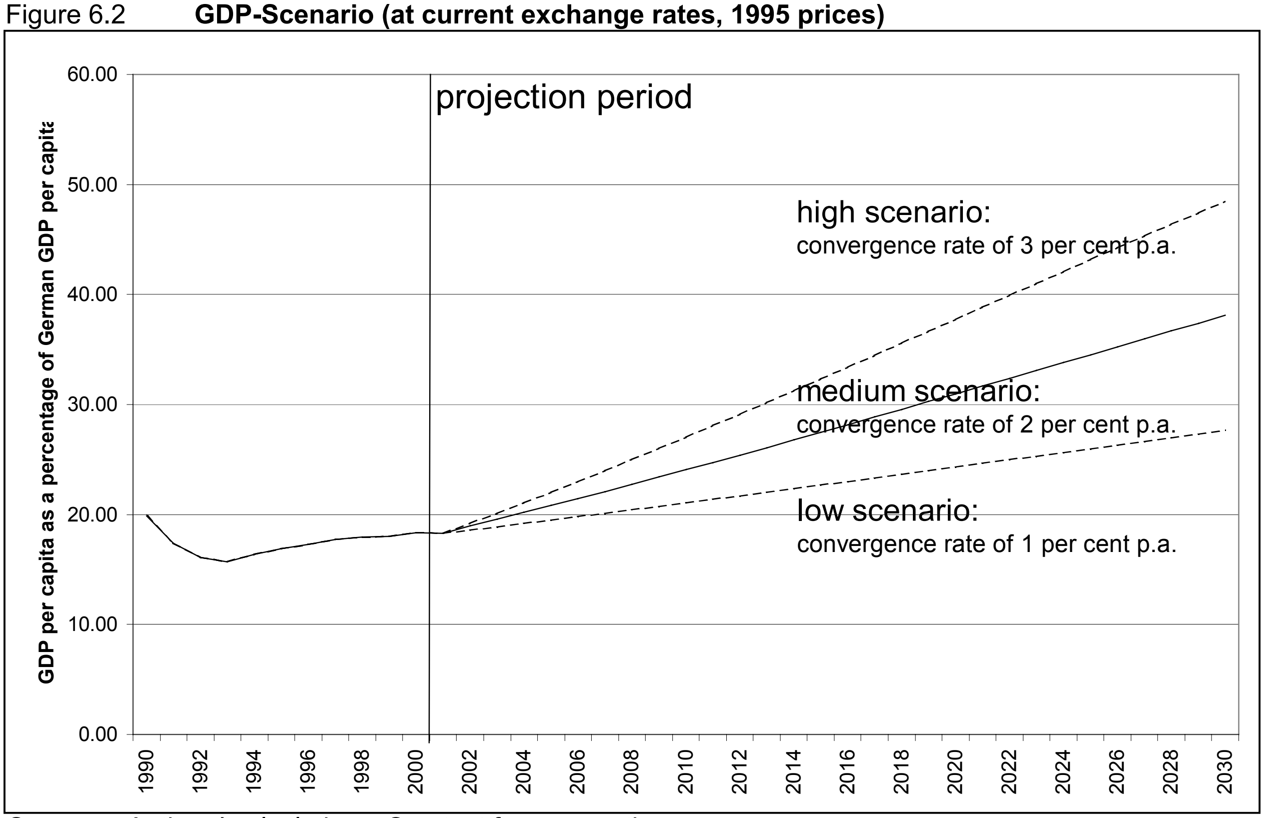


Souces: World Bank (2002), authors' calculations.

[17] In the migration scenario, we subtracted from this population figure the net emigration flow into the EU as estimated by the migration model in order to derive total population figures for our scenarios. This implies that we assume that the net migration balance with the rest of the world is zero, which is reasonable if we consider both net emigration flows from the CEECs to the US and other non-EU countries, and net immigration flows from other Eastern European countries into the CEECs.

- **GDP** and **PPP-GDP**: It is not possible to precisely forecast the long-run development of the GDP variable at current exchange rates and purchasing power parities. While there exists no clear evidence that economies tend to converge at world levels, a number of studies suggest that per capita GDP levels have converged among the EU Member States during the post-war period. Cross-sectional regressions indicate that per capita GDP levels have converged at an annual rate of 2 per cent,[18] which implies that an initial income gap tends to be reduced by half every 35 years. This rate has been found, following the pioneering studies of Barro and Sala-i-Martin (1991, 1995), in a large number of regressions (see Brücker, 2000, for a survey). However, the findings of these convergence regressions have been criticised for a number of reasons. We use the convergence rate of 2 per cent here only as a first indication in our baseline scenario. In an optimistic and a pessimistic convergence scenario we applied convergence rates of 3 per cent and 1 per cent respectively. The growth performance of the CEECs since the end of the transitional recession indicates that actual growth rates have been within the range of these convergence scenarios (see figures 6.2 and 6.3).

Figure 6.2 **GDP-Scenario (at current exchange rates, 1995 prices)**



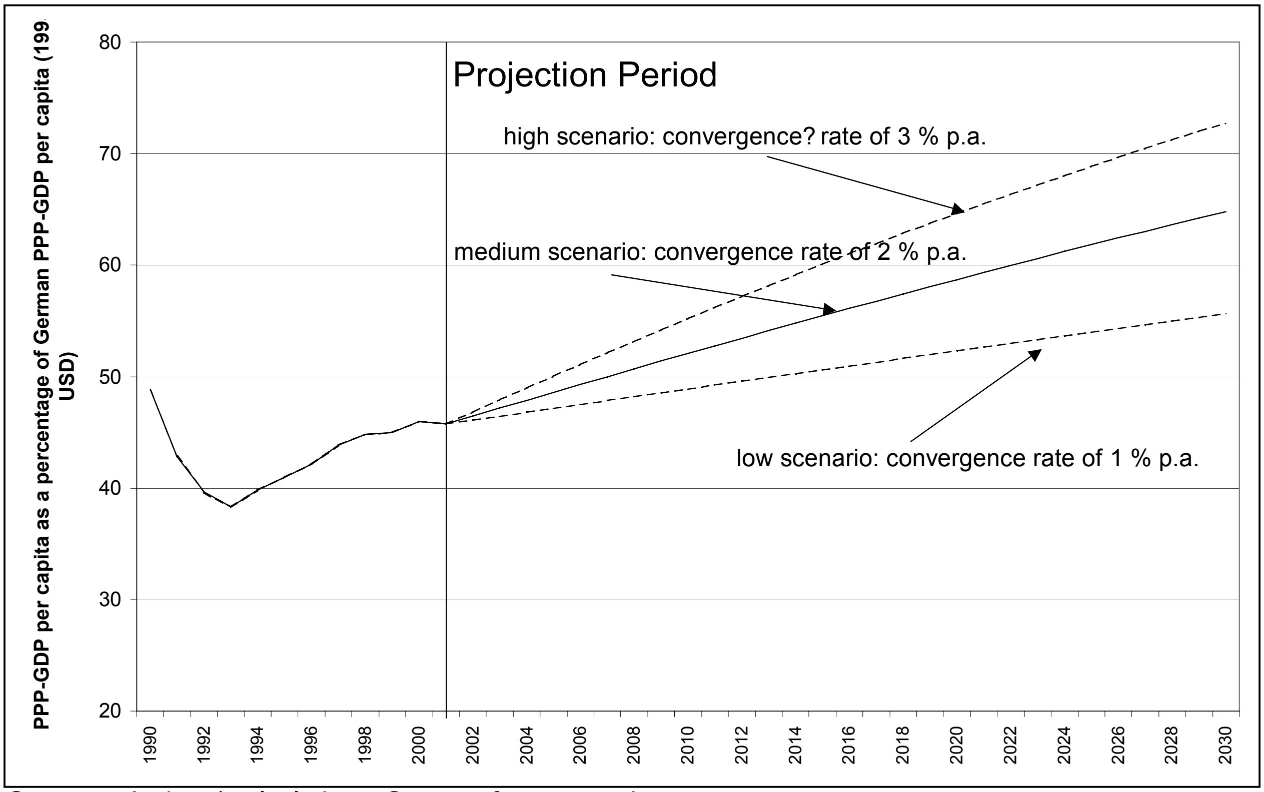
Sources: Authors' calculations. See text for assumptions.

Theoretical arguments suggest that GDP at current exchange rates might converge to EU levels more rapidly than GDP at purchasing power parities, since closing the productivity gap in the tradable sectors might be associated with an appreciation of the exchange rate.

[18] More precisely, the difference in the natural logarithm of GDP levels across countries tends to decline at a rate of 2 per cent p.a.

At the other hand, large current account deficits might be interpreted as evidence that exchange rates in the CEECs are presently overvalued. Moreover, since the end of the transition recessions, we do not observe that GDP at current exchange rates has converged faster in the CEECs than GDP at purchasing power parities. We thus applied the same rates of convergence for the GDP variables at current exchange rates and at purchasing power parities. For Germany an annual growth rate of 2 per cent is assumed, which corresponds to the long-term average during the last two decades.

Figure 6.3 **PPP-GDP Scenario (at purchasing power parity, 1995 prices)**

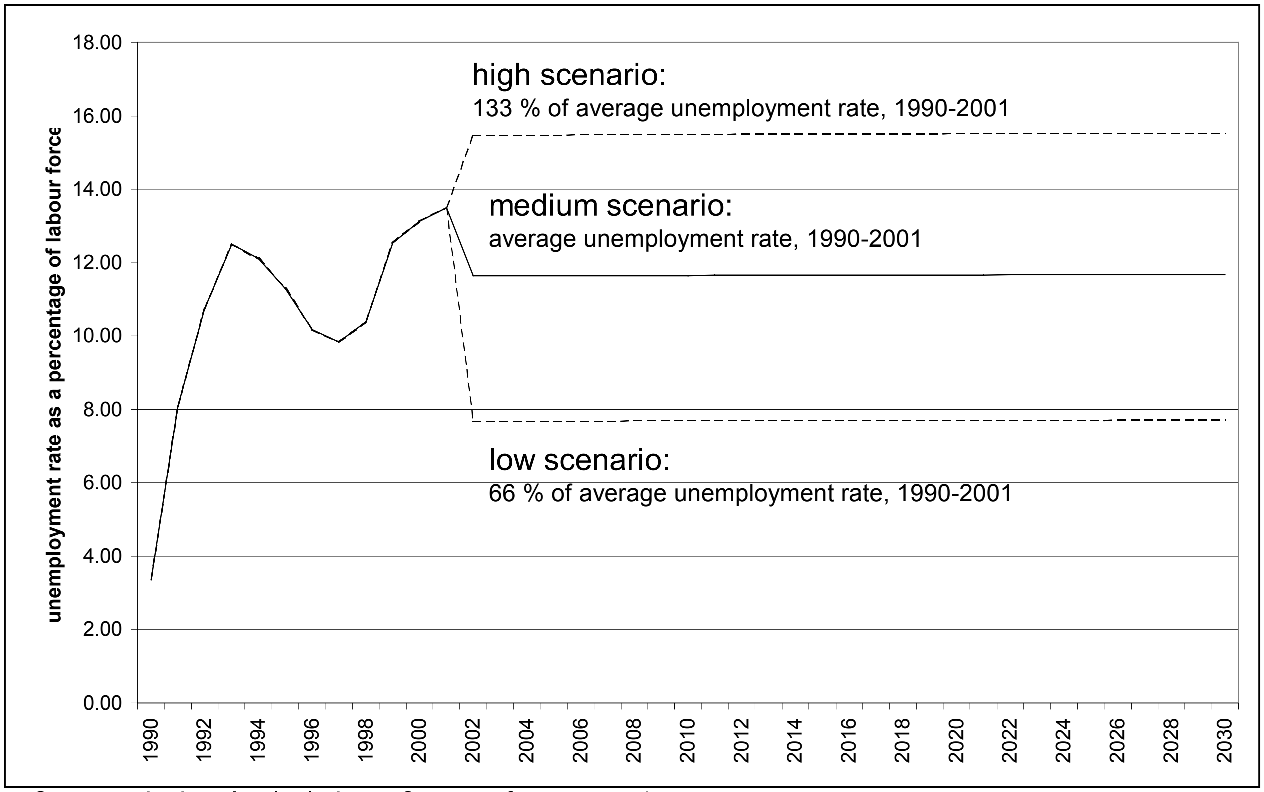


Sources: Authors' calculations. See text for assumptions.

- **Unemployment**: In the CEEC-10 unemployment rates have increased from an initial situation with almost full employment to an average rate of 13.5 per cent in 2001. However, there exist large differences between individual countries. In our baseline scenario we assume the long-run unemployment rate will equal the 1990-2001 average, which yields a rate of 11.6 per cent for the CEEC-10. As a lower bound we assume in an optimistic scenario that the long-run unemployment rate will equal two-thirds of the 1990-2001 average, which yields an average rate of 7.7 per cent for the CEEC-10, and as an upper bound we assume that the long-run unemployment rate will equal 133 per cent of the 1990-2001 average, which yields 15.5 per cent for the CEEC-10 (see figure 6.4).

In Germany, average unemployment amounted to 8.4 per cent during the period 1990-2001. We applied this rate in our baseline scenario, and assumed in the optimistic scenario that it will amount to two-thirds of this average (5.5 per cent) and in the pessimistic scenario that it will increase to 133 per cent of the 1990-2001 average (11.2 per cent).

Figure 6.4 **Unemployment Scenario for the CEEC-10**



Souces: Authors' calculations. See text for assumptions.

We tried to capture all plausible developments of the exogenous variables in Germany and the CEECs in three migration scenarios:

- A baseline scenario, which is based on a convergence rate of 2 per cent for the GDP and PPP-GDP per capita and the average unemployment rates in the CEECs and Germany during the period 1990-2001;
- A high migration scenario, which is based on the assumptions of a low convergence rate of GDP and PPP-GDP (1 per cent), high unemployment in the CEECs (133 per cent of the 1990-2001 average) and low unemployment in Germany (66 per cent of the 1990-2001 average);
- A low migration scenario, which is based on the assumption of a high convergence rate of GDP and PPP-GDP (3 per cent), low unemployment in the CEECs (66 per cent of the 1990-2001 average) and high unemployment in Germany (133 per cent of the 1990-2001 average).

Our simulations on the basis of the European sample are limited to a short-run scenario and the calculation of steady-state values, which are both computed on the basis of the present values of the income and employment variables.

6.2 Explaining the country-specific effects

Our various estimators employed in this study treat the country-specific effects differently. The WITHIN and SUR models treat all country-specific effects (e.g. distance and other geographical factors, language, cultural differences etc.) as fixed. These factors cannot be directly included in the estimation of these models, since they are by definition perfectly collinear. For an out-of-sample forecast, i.e. a forecast of migration flows and stocks from the CEECs, we need further information on the country-specific fixed effects.[19] In order to use the information from the fixed effects we follow the approach of Fertig (2000) and explain the fixed effects in a separate regression. A similar approach has been used in another context by Dickens and Katz (1987).

Country-specific effects also appear in the other models, but are treated differently there. The GLS estimators treat the country-specific effects as random and assume that they are distributed normally. In principle, these estimators allow inclusion of time-invariant variables in the first-stage regressions. We used a variety of variables to approximate geographical distance, common language and common borders as explanatory variables, but they turned out to be insignificant in the GLS regressions. The Arellano-Bond (1991) and Anderson-Hsiao (1982) estimators exclude fixed effects by estimating in first differences, while the Arellano-Bover (1995) GMM estimator, which uses lagged levels as moment conditions, makes it possible to directly include time-invariant variables in the model. With the exception of language, which turned out to be significant at the 10 per cent level, these variables also proved to be insignificant.

Table 6.1 reports the results of the OLS regressions we carried out to explain country-specific effects in the German sample. Against the background of the small number of observations, the explanation of the country-specific effects has to be limited to a small number of variables. In our analysis we employed language and a number of geographical variables: we used distance and distance squared between Frankfurt a.M. and the capitals of the source country in miles in order to capture the impact of distance on the pecuniary and non-pecuniary costs of migration. Moreover, we calculated the travelling time by car between Frankfurt and the capitals of the source countries from a navigation system. Finally we created a dummy variable (ADJACENT) to capture the fact that migrants tend to move to the most adjacent region or country with a high living standard. The last variable clearly outperformed the other distance variables, and as a result we limited our final regression to the ADJACENT and LANGUAGE dummies.

As can be seen in Table 6.1, when using the WITHIN and SUR estimators and also when using the Arellano-Bond (1991) estimators, which eliminate the fixed effects by differencing, the regression can explain between 40 and 50 per cent of the variation in individual effects. In the case of the GLS estimators, the picture is mixed: while the regression has a high explanatory power in case of the GLS(MLE) estimator, which has produced very similar results to the WITHIN estimator, it is almost meaningless in case of the Swami and Arora (1972) estimator. Even in the case of the Arreleno-Bover (1995) estimator, where we considered language as a time-invariant variable in the first regression, the second-stage regression can help explain the remaining variance in the individual effects. However, the coefficients have here rather low values.

[19] In principle, we could use the average value of the fixed effects and set all deviations at null for the simulations. However, in this case we would lose valuable information that may be contained in the fixed effects.

Table 6.1 **Explanation of individual effects (German Sample)**

	constant	ADJACENT	LANGUAGE	adj. R^2	S.E. of regression	F-statistic
WITHIN	-1.66 **	0.13 **	0.12 *	0.48	0.08	8.31 **
	(0.02)	(0.04)	(0.06)			
SUR	-1.28 **	0.12 **	0.09	0.45	0.07	8.49 **
	(0.02)	(0.03)	(0.05)			
GLS(WALHUS)	-1.32 **	0.02 **	0.03 *	0.37	0.02	6.23 **
	(0.01)	(0.01)	(0.01)			
GLS(SWAR)	-0.88 **	0.00	0.00	0.02	0.00	1.20
	(0.00)	(0.00)	(0.00)			
GLS(MLE)	-1.63 **	0.11 **	0.11 *	0.49	0.06	9.66 **
	(0.02)	(0.03)	(0.05)			
GMM1(DIF)	-2.39 **	0.11	0.32 **	0.37	0.14	6.21 **
	(0.05)	(0.07)	(0.11)			
GMM2(DIF)	-2.34 **	0.17 *	0.25 *	0.40	0.13	6.95 **
	(0.04)	(0.06)	(0.10)			
GMM1(SYS)	-1.67 **	0.03 *	0.04 *	0.34	0.02	5.66 *
	(0.01)	(0.01)	(0.02)			
GMM2(SYS)	-1.63 **	0.02 *	0.03	0.34	0.02	5.57 *
	(0.01)	(0.01)	(0.01)			
GMM1(SYS)-TINV	-1.58 **	0.03 *		0.16	0.03	4.48 *
	(0.01)	(0.01)				
GMM2(SYS)-TINV	-1.54 **	0.02 *		0.21	0.02	5.87 *
	(0.01)	(0.01)				

Dependent variable: Individual effect of respective regression.
Number of observations 19
Method: OLS.

Apart from self-explanatory OLS and WITHIN estimators, GLS(WALHUS) - Wallace and Hussein (1969) estimator, GLS(SWAR) - Swamy and Arora (1972) estimator, AH - Anderson and Hsiao (1982), GMM1(DIF) and GMM2(DIF) - one- and two-steps Arellano and Bond (1991) estimator, GMM1(SYS) and GMM2(SYS) - one- and two-steps Arellano and Bover (1995) estimator.

Standard errors in parantheses.

The results for the European sample have much less explanatory power. The *F*-Statistics indicate that with the exception of the GLS estimators, the second stage regression helps explain the individual effects even when using the Arellano-Bover estimator, which considers time-invariant variables. Many coefficients of the explanatory variables do, however, appear to be insignificant, and rather low values of the R-squared indicate that the explanatory power of the second-stage regression is rather small (Table 6.2).

Table 6.2 **Explanation of individual effects (European Sample)**

	constant		ADJACENT		BORDER		COLONIAL		adj. R^2	S.E. of regression	F-statistic	
WITHIN	0.80	**	0.20		0.20		1.23	**	0.17	0.51	15.47	**
	(0.04)		(0.11)		(0.13)		(0.27)					
GLS(WALHUS)	0.00		0.01		0.00		-0.01		0.00	0.04	1.05	
	(0.00)		(0.01)		(0.01)		(0.02)					
GLS(SWAR)	0.00		0.01		0.00		-0.03		0.00	0.03	1.27	
	(0.00)		(0.01)		(0.01)		(0.02)					
AH	0.32	**	0.27		0.23		2.25	**	0.14	0.85	12.90	**
	(0.07)		(0.21)		(0.16)		(0.47)					
GMM1(DIF)	0.60	**	0.13		0.18		1.67	**	0.09	0.77	7.87	**
	(0.06)		(0.19)		(0.15)		(0.43)					
GMM2(DIF)	-0.30	**	0.19	*	0.17	*	1.29	**	0.21	0.41	20.46	**
	(0.03)		(0.10)		(0.08)		(0.23)					
GMM1(SYS)	0.00		0.02	*	0.00		0.05	*	0.05	0.04	4.64	**
	(0.00)		(0.01)		(0.01)		(0.02)					
GMM2(SYS)	0.00		0.03	**	0.02	*	0.04		0.09	0.04	8.28	**
	(0.00)		(0.01)		(0.01)		(0.02)					

Dependent variable: individual effect of respective regression.
Number of observati 215
Method: OLS

Apart from self-explanatory OLS and WITHIN estimators, GLS(WALHUS) - Wallace and Hussein (1969) estimator, GLS(SWAR) - Swamy and Arora (1972) estimator, AH - Anderson and Hsiao (1982), GMM1(DIF) and GMM2(DIF) - one- and two-steps Arellano and Bond (1991) estimator, GMM1(SYS) and GMM2(SYS) - one- and two-steps Arellano and Bover (1995) estimator.

Standard errors in parantheses.

6.3 Comparing the simulation results

Table 6.3 shows the simulation's results for the different estimators in the order of their within-sample dynamic forecasting performance. The simulations have been carried out under the arbitrary assumption that free movement is introduced for all ten CEECs in 2004, and under the assumptions of the baseline scenario for the value of the explanatory variables. As can be seen in Table 6.3, the results vary between 1.8 and 2.4 million residents from the CEEC-10 in Germany ten years after free movement has been introduced, and between 2.2 and 3.0 million residents 20 years after introducing free movement – if we exclude the results from the explosive GLS(SWAR) estimator. The SUR estimator, which has shown the smallest within-sample dynamic forecasting error, predicts a migration stock of 2.1 million residents 10 years and of 2.2 million residents 20 years after the introduction of the free movement. A much higher variance of the different estimators can be observed with regard to the initial increase in the migration stocks. The SUR estimator predicts here an initial increase of some 180 thousand persons, which is slightly below the estimates of the European Integration Consortium (220 thousand persons). It should be noted that many estimators that forecast a small increase of migration stocks in the short run expect these increases to persist for long time periods, thus creating large stocks of migrants in the long run. Note that some of these estimators are biased due to the correlation of the lagged dependent variable with the omitted individual effects (e.g. the OLS estimator) and thus perform poorly in the long run.

Table 6.3 **Short-run and long-run forecasts of various estimators (German Sample)**

Rank	Estimator	increase of mst			mst	
		1st year	2nd year	10th year	10th year	20th year
		in thousand persons				
1	SUR	180	224	44	2 103	2 242
2	GMM2(SYS)	92	117	132	1 981	2 774
3	GLS(WALHUS)	85	113	121	1 919	2 622
4	WITHIN	230	275	40	2 296	2 416
5	GLS(MLE)	220	266	58	2 393	2 579
6	GMM1(SYS)	98	124	131	2 023	2 787
7	GMM2(SYS)-TINV	91	114	117	1 882	2 567
8	GMM1(SYS)-TINV	102	127	113	1 949	2 578
9	GMM2(DIF)	374	396	10	2 434	2 439
10	GLS(SWAR)[1)]	21	35	80	1 215	1 785
11	GMM1(DIF)	473	473	- 3	2 705	2 606
12	OLS	69	92	116	1 757	2 959

Apart from self-explanatory OLS and WITHIN estimators, GLS(WALHUS) - Wallace and Hussein (1969) estimator, GLS(SWAR) - Swamy and Arora (1972) estimator, AH - Anderson and Hsiao (1982), GMM1(DIF) and GMM2(DIF) - one- and two-steps Arellano and Bond (1991) estimator, GMM1(SYS) and GMM2(SYS) - one- and two-steps Arellano and Bover (1995) estimator.

In the European sample, the estimators with the best forecasting performance predict a long-run stock of between 2.2 and 2.7 million workers from the CEECs in the EU-15 for the steady state (at the present income difference). The outlier among the estimators, the WITHIN estimator, is certainly biased. In the short run, the estimators with a good forecasting performance predict that the stock of foreign workers from the CEECs in the EU-15 will increase by between 90 and 120 thousand workers p.a. in the beginning (Table 6.4). Given that around 40 per cent of the residents from the CEECs are employed at present, the short-run results are consistent with the predictions from the German sample, if we assume that the German share in migration from the CEECs remains constant at 60 per cent. The long-run results are around one-third higher than the results derived from the German sample.

While the simulations derived from the European sample are reasonable in the aggregate, the regional distribution is not: the estimators predict that the German share in migration from the CEECs will fall from around 60 per cent at present to between 12 and 30 per cent in the steady state, depending on the estimator used.[20] Given that regional migration patterns show high persistence over time due to network effects and other causes of path dependence, this is not plausible. We thus abstain from using the results from the European sample to develop a migration scenario for the individual EU countries.

[20] The estimator with the best within-sample performance, the GMM2(SYS), predicts Germany's share in steady state to be 13 per cent of the number of foreign nationals from the CEEC-10 working in the EU-15.

Table 6.4 **Short-run and long-run simulation results (European Sample)**

		increase of mst	mst
Rank	Estimator	1st year	steady state value
		in thousand persons	
1	GMM2(SYS)	111	2 188
2	GLS(SWAR)	117	2 661
3	GLS(WALHUS)	91	2 443
4	GMM1(SYS)	150	2 745
5	GMM2(DIF)	289	1 446
6	GMM1(DIF)	565	2 744
7	WITHIN	3 226	8 330
8	OLS	117	2 661

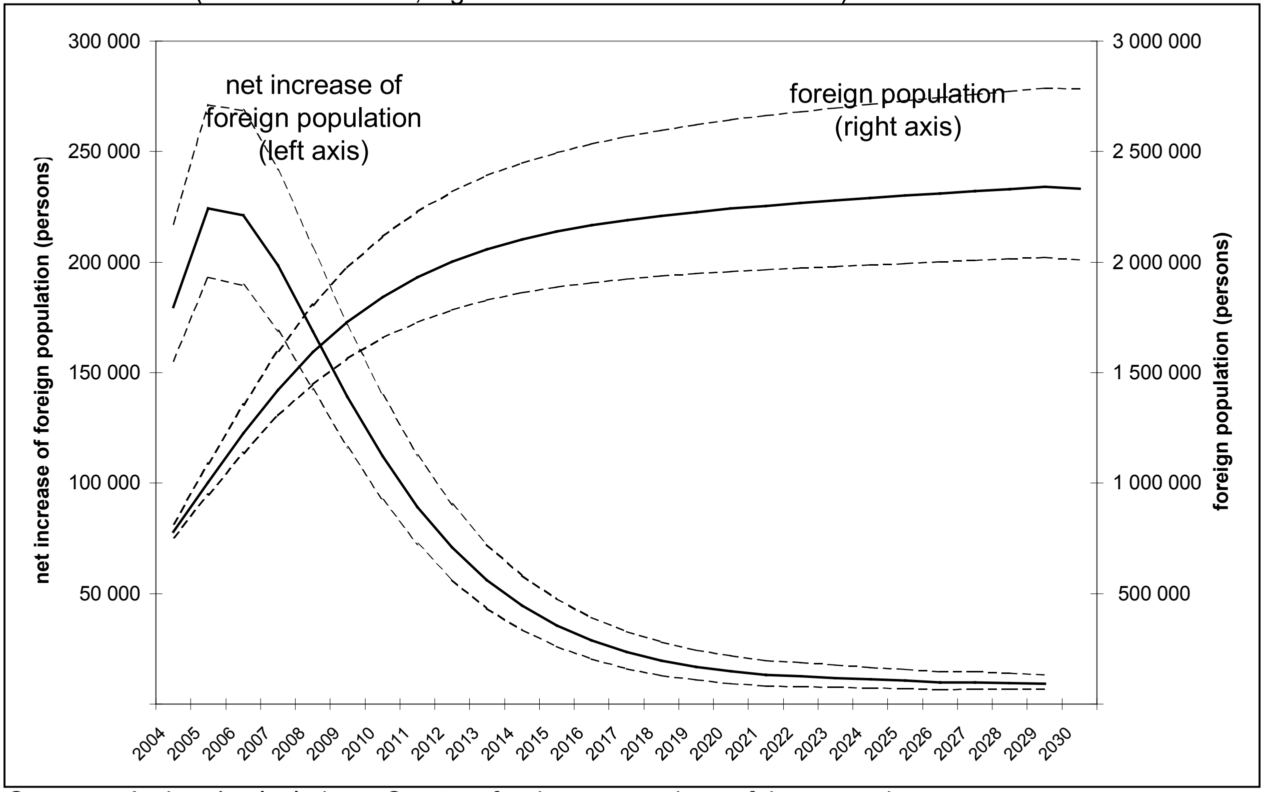
Apart from self-explanatory OLS and WITHIN estimators, GLS(WALHUS) - Wallace and Hussein (1969) estimator, GLS(SWAR) - Swamy and Arora (1972) estimator, AH - Anderson and Hsiao (1982), GMM1(DIF) and GMM2(DIF) - one- and two-steps Arellano and Bond (1991) estimator, GMM1(SYS) and GMM2(SYS) - one- and two-steps Arellano and Bover (1995) estimator.

6.4 Sensitivity analysis

In the baseline scenario, the introduction of free movement in 2004 yields an annual net increase of the foreign population in Germany from the CEEC-10 of approximately 180,000 persons initially, if we employ the assumptions of the baseline scenario. The shape of the curvature of the migration inflow follows a hump, i.e. migration inflows tend to increase initially and then decline. The peak is reached at an annual net increase of some 220,000 persons. The net increase tends to fall to 50 per cent of its initial value within six years after free movement has been introduced. The size of the foreign population from the CEEC-10 reaches its peak at around 2.3 million approximately 25 years after free movement has been introduced and declines slightly in the course of converging per capita income levels.

Under the assumptions of the high migration scenario – low unemployment rate in Germany, high unemployment rate in the CEECs and slow income convergence – the initial net increase in the foreign population from the CEECs is between 35,000 and 50,000 persons higher in the first few years, and the long-run migration stock is, at 2.8 million persons, almost 700,000 persons higher than in case of the baseline scenario. In case of the low migration scenario – high unemployment rate in Germany, low unemployment rate in the CEECs, fast income convergence – the model predicts an initial net increase in the foreign population from the CEECs, which is around 30,000 persons below the baseline scenario and a long-run stock of residents which is 330,000 persons below the baseline scenario (Figure 6.5, Table A.4).

Figure 6.5 **Migration scenarios for Germany**
(Baseline scenario; high and low scenario: dotted lines)



Sources: Authors' calculations. See text for the assumptions of the scenario.

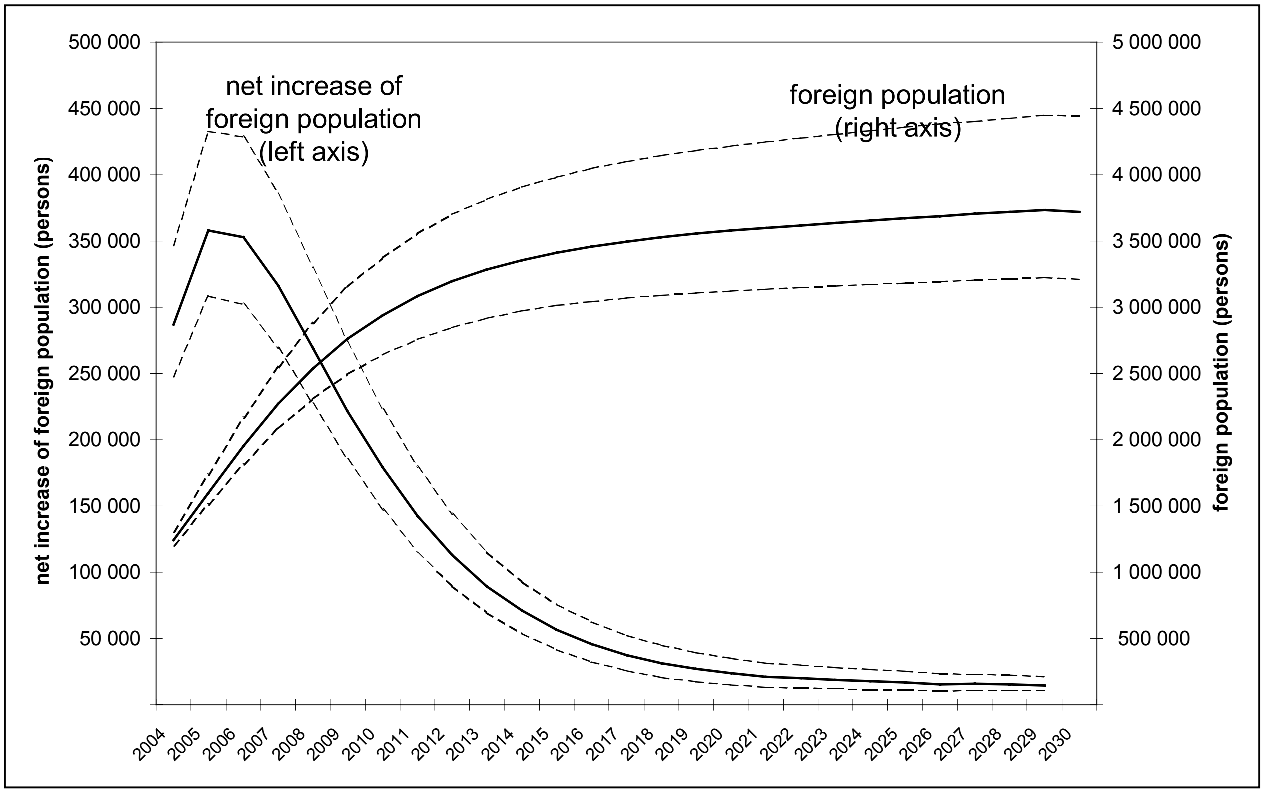
More than 70 per cent of the flows and stocks of the migrants from the CEECs are expected to stem from the eight countries that will presumably participate in the first enlargement round, while less than 30 per cent are expected to stem from Bulgaria and Romania. This corresponds roughly to the population shares. With converging per capita income levels, Bulgaria's and Romania's shares in the migrant population from the CEECs tends to decline in the simulation model (Table A4).

Extrapolation to the EU-15

Our estimates are based on migration into Germany and, hence, allow us to simulate only the development of the foreign population there. However, we observe that the geographical distribution of the migrant population across European countries is fairly stable over time. Based on the present distribution of migrants across the EU-15, we thus extrapolate our findings to the other EU Members in order to obtain a general idea of migration potential. We can safely expect this procedure to yield significantly more plausible results than simulations based on the estimates within the European sample (see above).

Our extrapolation is based on the migrant distribution in 2001 as shown in Table 2.1. Note that we had to estimate the foreign population from the CEECs in some large EU Member States (e.g. France and UK), since recent figures are not available there. As can be seen in Figure 6.6 and Table A.5, an extrapolation of the results from Germany based on the present distribution of the migrant population from the CEEC-10 across the EU-15 yields an initial net increase of the foreign population from the CEEC-10 of 294,000 persons. The net increase reaches its peak at around 370,000 persons. The long-run stock of migrants is estimated in the baseline scenario to number 3.8 million persons. Under the assumptions of the high migration scenario the net increase of the foreign population is between 30,000 and 40,000 persons higher than in the baseline scenario, and the long-run migration potential can increase to almost 4.5 million persons. In contrast, under the assumptions of the low migration scenario, the initial net increase is estimated at 250,000 persons and the long-run migration potential at around 3.2 millions. Note that less than 40 per cent of the foreign residents from the CEEC-10 are employees at present.

Figure 6.6 **Migration Scenario for the EU-15** (Extrapolation from the German Scenario)

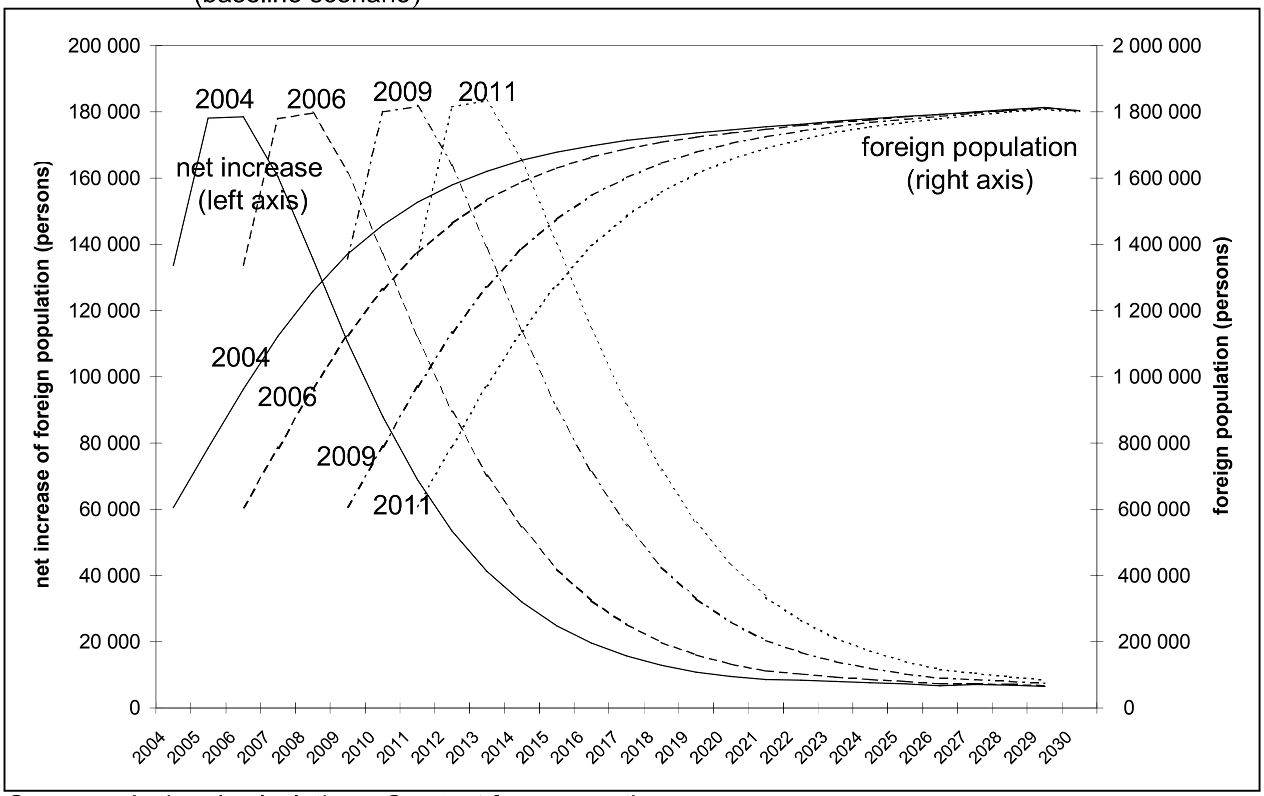


6.5 The impact of transitional periods

The scenarios reported in the previous section rely on the arbitrary assumption of free movement for all ten accession candidates from Central and Eastern Europe being introduced in 2004. This is hardly a realistic policy scenario, however: the accession of Bulgaria and Romania will be postponed at least until 2007, according to a proposal of the European Commission. Moreover, according to the European Council, the present EU Members will be allowed to use transitional periods for introducing free movement, which could postpone it up to seven years. Germany, for example, has already announced that it plans to make use of a transitional period, but it remains unclear whether the maximal period will be used and what kind of migration policies will be pursued during this period.

Thus, we simulated the impact of transitional periods in Germany of two years, five years and seven years for the CEEC-8, assuming accession of this country group in 2004. For the CEEC-2 we simulate the introduction of free movement in 2007, and transitional periods of another five and seven years. The impact of postponing free movement depends largely on the immigration policy pursued during the transitional period. Our simulations are based on the simple assumption that a restrictive policy of zero net migration is pursued, which corresponds closely to actual German immigration policies in the second half of the 1990s and early 2000s.

Figure 6.7 **Simulation of the impact of transitional periods on migration from the CEEC- 8** (baseline scenario)

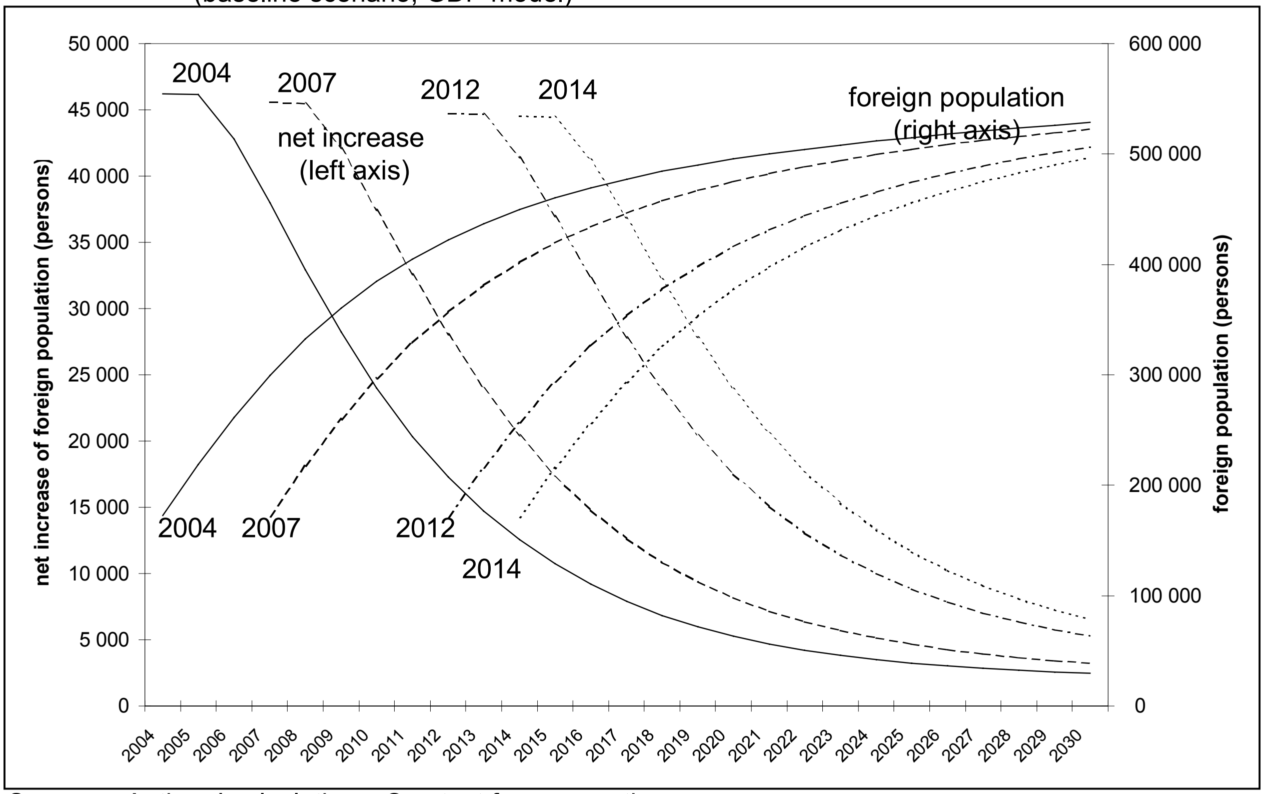


Sources: Authors' calculations. See text for assumptions.

Figure 6.7 shows the impact of transitional periods for the CEEC-8 based on the baseline scenario and the model: under the assumptions of our scenario, postponing free movement from 2004 to 2006, 2009 or even 2011 yields only a marginal reduction in the net increase of migrants after free movement has been introduced. The migrant stocks tend to converge relatively rapidly to their long-run levels in the different scenarios. Thus, postponing free movement neither reduces net migration flows in the initial years after liberalisation nor does it affect the long-run stocks of the foreign population. Note, however, that these results depend heavily on the assumption that a restrictive policy of zero net immigration is pursued during the transitional periods.

From the scenarios for Bulgaria and Romania we obtain similar results: postponing free movement from 2004 to 2014 reduces net inflows only moderately from almost 47,000 persons to some 44,000 persons and net stocks of foreign population tend to converge at some 520,000 persons in the long run (Figure 6.8).

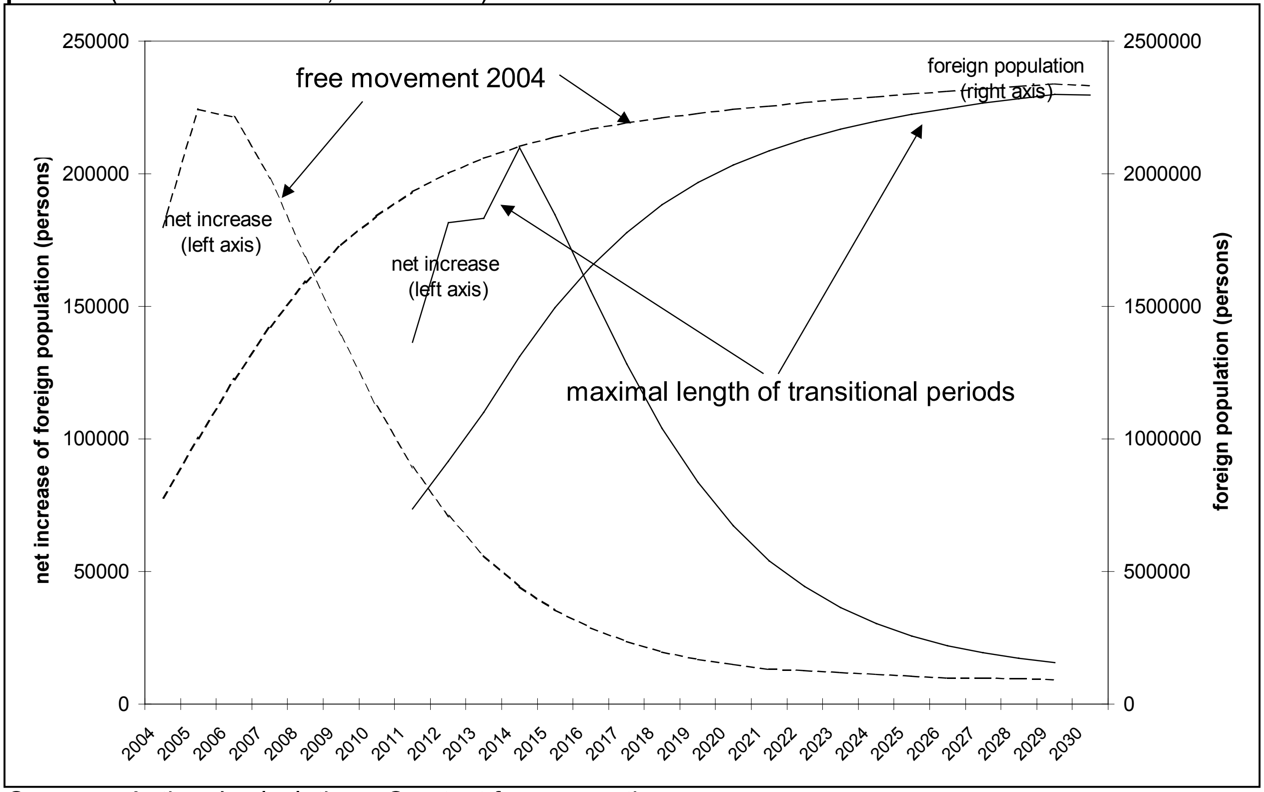
Figure 6.8 **Simulation of the impact of transitional periods on migration from Bulgaria and Romania**
(baseline scenario, GDP model)



Sources: Authors' calculations. See text for assumptions.

Finally, in Figure 6.9 we show the implications of the most restrictive scenario for Germany: free movement is postponed for the eight countries of the first accession round until 2011 and until 2014 for the remaining two countries. In this case, the net increase of the foreign population peaks at approximately 195,000 persons in 2015, one year after free movement has been introduced for Bulgaria and Romania. This is some 50,000 persons less than if free movement were introduced for all the CEEC-10 at the same time in 2004. Thus, distributing introduction of free movement across different dates might mitigate annual inflows slightly, while the long-run stocks of the foreign population remain largely unaffected by transitional periods. Altogether, transitional periods tend to postpone the migration influx, but they will have only a marginal impact on the size of the inflows and, more importantly, on the long-run stocks of the migrant population (Table A.6).

Figure 6.9 Simulation of the impact of the utilisation of the maximal length of transitional periods (baseline scenario, GDP model)



Sources: Authors' calculations. See text for assumptions.

These simulations are, however, highly arbitrary. They not only assume a policy of zero net migration during the whole transitional period, but also ignore third-country effects. The transitional periods can distort the regional distribution of migrants from the CEECs across the EU-15, that is, the diversion migration flows away from countries which restrict immigration and into countries which pursue more liberal migration policies. Note, however, that the regional distribution of migration shows a high persistence over time due to network effects. Moreover, a large migration community in one country increases the opportunities for further immigration through channels such as family reunification and illegal migration, even if a restrictive migration policy is pursued. However, an assessment of the impact of transitional periods on regional migration patterns is beyond the scope of scientific research since the specific conditions for migration from the CEECs into the present EU Member States during the transitional periods are unprecedented.

7 Conclusions

The migration conditions between the EU and the CEECs have changed only slightly since the study carried out by the European Integration Consortium: the income difference in purchasing power parities is estimated by Eurostat to be somewhat below the income difference used in the old scenario, and the conditions for economic growth have tended to stabilise in many of the accession candidate countries. Moreover, economic conditions for immigration from the CEECs in the main host country, Germany, have deteriorated in 2001 and 2002. Nevertheless, since the migration scenario is based on long-term assumptions on growth and convergence of the relevant variables, these minor changes have affected the size of the migrant potential only marginally.

In this study we compared a wide variety of estimators in order to assess their forecasting performance. Our model explains the stock of migrants by income, employment rates and a set of institutional variables. Similar to the approach applied by the European Integration Consortium, our model assumes that a long-run equilibrium relationship exists between the stock of migrants and the explanatory variables -- rather than between the migration rate and the explanatory variables. The equilibrium relation between migrant stocks and the explanatory variables is a result of heterogeneity of individuals with respect to human capital characteristics and preferences affecting the migration decision. The adjustment process is modelled in the form of a simple habit persistence model, i.e. we assume that individuals adjust at a sluggish pace to changes in economic and other explanatory variables.

Panel data sets allow us to exploit the variation in the data over time and across countries in order to identify the coefficients for the variables of interest. Different estimators use different sources of this variation and thus yield different results. Econometric theory suggests that in the case of dynamic panel models, traditional WITHIN and SUR estimators perform relatively well in data sets with a large time dimension, while these estimators are subject to estimation bias in panels with a small time dimension. Recently developed GMM estimators, which address this estimation bias, are therefore expected to outperform traditional estimators in panels with a relatively small time dimension. The pooled OLS estimator, which is widely applied in the empirical literature, yields biased results in data sets with a large time dimension as well if the lagged dependent variable is correlated with omitted country specific effects (e.g. geography, language, culture).

Our results confirm the predictions from econometric theory. We used two data sets for our analysis: a German sample, which comprises 32 years of immigration to Germany from 19 source countries, and a European sample, which comprises eight years of migration between 215 countries. In case of the German sample, the traditional SUR estimator outperforms the other estimators. Conversely, in case of the European sample the GMM estimators show a better forecasting performance than the traditional estimators. Thus, our results suggest that it is advantageous to use different estimators for the two data sets depending on their time dimension. In this context it is worth noting that the forecasting performance of the estimators was much better in the German sample, which is due among other things to the higher quality of this data set.

Interestingly enough, the dynamic forecasting performance of the pooled OLS estimator turned out to be the weakest of all estimators in both samples. This result suggests that the pooled OLS estimations of dynamic migration models are heavily biased. This has a particularly distorting effect on long-term forecasts based on the pooled OLS estimator. This finding has important implications for an assessment of the empirical literature, since the pooled OLS estimator is widely applied there. Specifically, a study carried out by the Ifo institute (Sinn et al., 2000) predicted for Germany a long-run migration potential of between 8 and 10 per cent of the population from the CEEC-10 on basis of a pooled OLS model. Based on similar data, the WITHIN and SUR estimators yield a long-run migration potential between 2 and 3 per cent of the population. Our findings indicate that the extremely high estimates of the long-run migration potential in the Ifo study reflect an estimation bias rather than actual migration pressures.

Our estimation of the migration potential for Germany largely confirms the findings of the European Intergation Consortium. The net increase in the number of foreign residents from the CEEC-10 in Germany – which roughly equals the net migration flow -- is estimated at 180,000 persons immediately after the introduction of the free movement and can be expected to reach its peak at around 225,000 persons one year later. The long-run migration potential is estimated at 2.3 million persons. The peak of the foreign population from the CEECs in Germany will be reached around 25 years after free movement has been introduced. The medium- and long-term simulation results of the other estimators with a good forecasting performance are roughly in line with those of the SUR estimator, although the estimates of the short-term increase in the foreign population are in some cases substantially lower.

These simulations are based on the assumption that per capita GDP levels converge at an annual rate of 2 per cent and that the employment rates in Germany and the CEECs will remain constant at their average levels from the second half of the 1990s. In order to prove the sensitivity of the results, we employed in a high migration scenario a convergence rate of 1 per cent, a low unemployment rate in Germany (1/3 below the average unemployment rate in the 1990s) and a high unemployment rate in the CEECs (1/3 above the average unemployment rate in the 1990s), and in a low migration scenario a convergence rate of 3 per cent, a high unemployment rate in Germany (1/3 above the average unemployment rate in the 1990s) and a low unemployment rate in the CEECs (1/3 below the average unemployment rate in the 1990s). In the high migration scenario, the initial net increase in the foreign population is expected to be between 35,000 and 50,000 persons higher than in the baseline scenario, while long-run migration potential is estimated to be 500,000 persons higher. Conversely, in the low migration scenario, initial net migration is estimated at approximately 30,000 persons fewer than in the baseline scenario, and in the long-run migration scenario, at 330,000 persons fewer.

The estimates based on the European sample are much less precise than those based on the German sample. Specifically, the employment variables and most of the distance variables turned out to be insignificant. The simulation based on the preferred Arellano-Bover (1995) GMM estimator yields an initial increase in the number of foreign workers from the CEEC-10 in the EU-15 of 111,000 persons and a long-run stock of foreign workers of 2.2 million. Given that almost 40 per cent of the foreign residents from the CEECs participate as employees in the labour market and that around 60 per cent of the migrants from the CEECs reside in Germany, the short-run estimates are slightly below, and the long-run figures around one-third above the estimates derived from the German sample. Although the aggregate figures from these estimates are plausible, the geographical distribution is not: they show Germany's share in the number of migrant workers from the CEECs falling from 60 per cent at present to between 12 and 30 per cent, which is certainly not likely. We thus abstained from basing simulations of the regional distribution of migrants from the CEEC-10 on these estimates.

An extrapolation of the results from the German sample to the EU-15 yields an initial net increase of residents from the CEEC-10 of 294,000 persons, which reaches its peak at almost 370,000 persons one year later. The long-run stock is estimated at 3.8 million persons. Under the assumptions of the high migration scenario, the long-run migration potential could be as many as 4.5 million persons, and in case of the low migration scenario as few as 3.2 million persons.

Actual migration flows and stocks may substantially deviate from our simulation results. It is worth noting that country-specific factors that rely on largely unobservable variables have an important impact on the magnitude of migration flows. The extrapolation of the coefficients of our regressions to countries outside the sample therefore affects the quality of the projections. Moreover, the extrapolation of our results to the EU-15 relies on the strong assumption that the present distribution of migrants from the CEECs across individual EU Member States will remain constant. Although we observe a high persistence of regional migration patterns due to network effects, these need not be necessarily remain the case in the longer term.

The simulation of the transitional periods shows that postponing the introduction of free movement has only a marginal impact on the scale of migration: postponing free movement for seven years or more will reduce initial migration by only a few thousand persons. Thus, postponing free movement does not mitigate migration pressures if policies of zero net migration are pursued during the transitional periods. Small reductions in initial migration inflows can, however, be achieved if free movement is postponed a few years for some of the accession countries, since in this case the inflows are distributed across a larger time period.

From an economic perspective, a restrictive use of the transitional periods will therefore fail to mitigate possible pressures from migration on the labour market. If the objective of transitional periods is to ease the adjustment process, then they should be used to implement safeguard clauses or quotas. In this case potential migration pressures are reduced step by step, such that introducing free movement at a later stage will then yield only a moderate influx instead of a migration hump. Note that the results from the estimation of our migration model suggest that migration responds rather quickly to changes in GDP and employment growth. Thus, international migration contributes significantly to adjusting the labour supply to fluctuations in economic activity.

References

Alecke, B., P. Huber, G. Untiedt (2001), What a difference a constant makes. How predictable are international migration flows?, in: OECD (ed.), *Migration Policies and EU-Enlargement, The Case of Central and Eastern Europe*, OECD, Paris, pp. 63-78.

Anderson, T.W., C. Hsiao (1982), Estimation of dynamic panel models with error components, *Journal of the American Statistical Association*, 76, pp. 47-82.

Arellano, M., S. Bond, (1991), Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *Review of Economic Studies*, 58, pp. 277-97.

Arellano, M., O. Bover (1995), Another look at the instrumental-variable estimation of error-components models, *Review of Economic Studies*, 58, pp. 277-97.

Baltagi, B. H. (1995), *Econometric Analysis of Panel Data*. Chichester: John Wiley and Sons.

Baltagi, B. H., J. M. Griffin, and W. Xiong (2000), To pool or not to pool: homogeneous versus heterogeneous estimators applied to cigarette demand, *The Review of Economics and Statistics*, 82(1), 117-126.

Banerjee, B., S.M.R. Kanbur (1981), On the specification and estimation of macro rural-urban migration functions with an application to Indian data, *Oxford Bulletin of Economics and Statistics*, 43, pp. 7-29.

Barro, R., X. Sala-i-Martin (1991), Convergence across States and Regions, *Brookings Papers on Economic Activity*, No. 1, pp. 107-182.

Barro, R., X. Sala-i-Martin (1995), *Economic Growth*, New York et al.: McGraw-Hill.

Bauer, T., K.F. Zimmermann (1999), *Assessment of Possible Migration Pressure and ist Labour Market Impact Following EU Enlargement to Central and Eastern Europe*, Research Report No. 3, IZA, Bonn.

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

Boeri, T., H. Brücker et al. (2001), *The Impact of Eastern Enlargement on Employment and Labour Markets in the EU Member States*, Report for European Commission, DG Employment and Social Affairs, Brussels.

Brücker, H. (2000), A Second Economic Divide in Europe? in: Hoffmann, L., F. Möllers (eds.): *Ukraine on the Road to Europe*, Springer Publ., Berlin.

Brücker, H. (2001), Die Folgen der Freizügigkeit für die Ost-West-Migration. Schluss-folgerungen aus einer Zeitreihenanalyse der Migration nach Deutschland, 1967 bis 1998, *Konjunkturpolitik*, Supplement 52, pp. 17-54.

Büchel, F., J. Schwarze (1994), Die Migration von Ost- nach Westdeutschland – Absicht und Realsisierung, *Mitteilungen zur Arbeitsmarkt- und Berufsforschung (MittAB)*, 1/94, 43-52.

Burda, M.C. (1995), Migration and the Option Value of Waiting, *Economic and Social Review*, 27 (1), pp. 1-19.

Doornik, J. A., M. Arellano, and S. R. Bond (2002), *Panel Data estimation DPD for Ox*, Manual, available online: http://www.nuff.ox.ac.uk/users/doornik/

Engle, R.F., C. Granger (1987), Co-integration and error correction: interpretation, estimation, and testing, *Econometrica*, 66, pp. 251-76.

Eurostat (2002a): *Yearbook 2002*, Eurostat: Luxembourg.

Eurostat (2001): *Yearbook 2001*, Eurostat: Luxembourg.

Eurostat (2000): *Yearbook 2000*, Eurostat: Luxembourg.

Eurostat (2002b): *The GDP of the Accession Candidate Countries*, Statistics in Focus, Eurostat: Luxembourg.

Eurostat Labour Force Survey (2001), Eurostat: Luxembourg.

Faini, R., A. Venturini (1995),: *Migration and Growth: The Experience of Southern Europe*, CEPR Discussion Paper No. 964.

Fassmann, H., C. Hintermann (1997), *Migrationspotential Ostmitteleuropa*, ISR Forschungsberichte 15, Institut für Stadt- und Regionalforschung, Vienna.

Fertig, M. (2000), The economic impact of EU-enlargement: assessing the migration potential, *Empirical Economics*, 26, pp. 707-20.

Fertig, M., C.M. Schmidt (2000), *Aggregate-Level Migration Studies As a Tool for Forecasting Future Migration Streams*, Universität Heidelberg, Department of Economics, Discussion Paper Series No. 324.

Flaig, G. (2001), Zur Abschätzung der Migrationspotentiale der osteuropäischen EU-Beitrittsländer, *Konjunkturpolitik*, Supplement 52, pp. 55-76.

Franzmeyer, F., H. Brücker (1997), Europäische Union, Osterweiterung und Arbeitskräftemigration, *DIW-Wochenbericht,* 64 (5), pp. 89-96.

Granger, C. W. J. and P. Newbold (1974), Spurious regression in econometrics, *Journal of Econometrics*, 2(2), pp. 111-20.

Hamilton, J. (1994) *Time series analysis*, Princeton, NJ: Princeton University Press.

Harris, J., M. Todaro (1970), Migration, unemployment and development: a two-sector analysis, *American Economic Review*, 60, pp. 126-142.

Hatton, T.J, J.G. Williamson (1992), *What Drove the Mass Migration from Europe in the Late Nineteenth Century*, Unpublished Paper, Harvard University, October.

Hatton, T.J. (1995), A model of UK migration 1870-1913, *Review of Economics and Statistics*, 77 (3), 407-415.

Hille, H., T. Straubhaar (2001), *The Impact of EU-Enlargement on migration movements and economic integration: results of recent studies*, in: OECD (ed.), Migration Policies and EU-Enlargement, The Case of Central and Eastern Europe, OECD, Paris, 79-100.

Hönekopp, E. (2000): *The Impact of Migration, Background Report to 'The Impact of Eastern Enlargement on Employment and Labour Markets in the EU Member States'*, mimeo, Nuremberg: Institut für Arbeitsmarkt- und Berufsforschung.

Hospodarske Noviny (2001), *Práce v cizine neláká jen výdelkem*, April, 30, p. 4.

Hsiao, C. (1986), *Analysis of Panel Data*, Cambridge University Press, Cambridge MA.

Im, K. S., Pesaran, M. H., and Y. Shin (2002), Unit root testing in heterogenous panels, *Journal of Econometrics*, forthcoming.

Judson, R.A., A.K. Owen (1999): Estimating dynamic panel data models: a guide for macroeconomists, *Economic Letters*, 65, pp. 9-15.

Layard, R., O.J. Blanchard, R. Dornbusch, P. Krugman (1992), *East-West Migration: The Alternatives*, MIT-Press, Cambridge MA.

Lundborg, P. (1998), The Free Movement of Labour between Sweden and the New EU Members, in: Committee on the Economic Effects of EU Enlargement (eds.), *A Bigger and Better Europe? – Final Report*, Government Official Reports, Stockholm.

Maddison (1995), *Monitoring the World Economy 1820-1992*, OECD: Paris.

McCormick, B., H. Brücker, G.S. Epstein, G. St.Paul, A. Venturini, K.F. Zimmermann (2002), *Managing Migration in the European Welfare State*, Fondazione DeBenedetti, Milano.

Nelson, C. R. and C. I. Plosser (1982) Trends and random walks in macroeconomic time series: some evidence and implications, *Journal of Monetary Economics*, 10(2), pp.139-62.

Nickell, S. (1981), Biases in Dymamic Models with Fixed Effects, *Econometrica*, 49, pp. 1417-1426.

OECD (2002a), *Historical Statistics*, CD-ROM, Paris.

OECD (2002b), *Main Economic Indicators*, CD-ROM, Paris.

Orlowski, W.M., L. Zienkowski (1999), *Potential size of the migration from Poland after joining the EU*, The Vienna Institute Monthly Report, 1999(2), 8-12.

Pedroni, P. (1999) Critical values for cointegration tests in heterogenous panels with multiple regressors, *Oxford Bulletin of Economics and Statistics*, 61, pp. 653-70.

Quaisser, W., M. Hartmann, E. Hönekopp, M. Brandmeier (2000), *Die Osterweiterung der Europäischen Union: Konsequenzen für Wohlstand und Beschäftigung in Europa*, Friedrich Ebert Foundation, Bonn.

Sinn, H.-W./G. Flaig/M. Werding/S. Munz/N. Düll/H. Hofmann (2000), *EU-Erweiterung und Arbeitskräftemigration: Wege zu einer schrittweisen Annäherung der Arbeitsmärkte*, ifo-Institut für Wirtschaftsforschung, Munich.

Sjaastad, L. (1962), The Costs and Returns of Human Migration, *Journal of Political Economy*, 70, 80-93.

Statistisches Bundesamt (1989), *Wirtschaft und Statistik*, No. 9, Wiesbaden.

Statistisches Bundesamt (1999), *Fachserie 1 (Bevölkerung)*, R2 (1998/99), Wiesbaden.

Swamy, P.A.V.B., and S.S. Arora (1972), The exact finite sample properties of the estimators of coefficients in the error component regression models, *Econometrica*, 40, 261-275.

UNECE (2002), *Economic Survey of Europe*, No. 1, 2002, Economic Commission for Europe of the United Nations, Geneva.

Wallace, C. (1998), *Migration Potential in Central and Eastern Europe*, International Organization for Migration (IOM), Geneva.

Wallace, T. D., and A. Hussein (1969), The use of error components models in combining cross-section and time-series data, *Econometrica*, 37, 55-72.

Walterskirchen, E., R. Dietz (1998), *Auswirkungen der EU-Osterweiterung auf den österreichischen Arbeitsmarkt*, Österreichisches Institut für Wirtschaftsforschung (WIFO), Vienna.

Windmeijer, F. (2002). A finite sample correction for the variance of linear two-step GMM estimators. IFS working paper W00/19: http://www.ifs.org.uk , London: The Institute for Fiscal Studies.

World Bank (2002), *World Development Indicators*, CD-Rom, Washington.

Annex 1 Description of the Data

In our econometric analysis we use two data sets: a German sample, where the migration figures stem from the German central register of foreign nationals, and a European sample, which refers to the number of foreign workers which are derived from the European Labour Force Survey.

In our econometric analysis of the German sample we pool the data for 19 European source countries (Austria, Belgium, Denmark, Finland, France, Greece, Iceland, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, United Kingdom, (former) Yugoslavia). This country sample covers the European source countries of migration in Germany almost completely, with the exception of the countries of the former COMECON. The COMECON countries have been excluded since the "iron curtain" effectively prevented migration for the main period of analysis.

The dependent variable is the share of foreign citizens residing in Germany as a percentage of the home population. Foreign nationals are defined by their citizenship. Note that citizenship is granted on basis of the concept of ethnicity in Germany, such that the large majority of second- and third-generation migrants still possess foreign citizenship. Data on the foreign-born population are not available in the German statistics. The data on foreign residents stem from the Federal Statistical Office (Statistisches Bundesamt, Fachserie 1). Foreign residents have been reported in Germany since 1967 on an annual basis by the local municipalities, and have been counted by the central register of foreign nationals (Ausländerzentralregister) in Cologne since 1972. In general, the foreigner statistics in Germany tend to over-report the number of legal migrants slightly, since return migration is not completely recorded in the official figures.

In the sample period, we observe two statistical breaks: First, the transition of paper-based counting of foreign nationals by the local municipalities to computer-based counting by the central register of foreigners in 1972 produced a minor statistical break in case of some source countries (Statistisches Bundesamt, 1999, p. 5). The second break emerged after a revision of the foreigner statistics in the course of the population census of 1987, which reduced foreigner figures significantly for a period of three years (Statistisches Bundesamt, 1989, p. 594). After three years, the statistics were based again on the non-revised figures of the central register of foreigners, however. In order to control for the first break we included a dummy variable in the regressions, but this turned out to be insignificant. We thus decided to ignore this break. With respect to the second break, we recalculated the number of foreign residents on the basis of net migration figures for the three years affected by the revisions of the Federal Statistical Office.[21]

The migration stock variable is normalised by the population of the home countries. Population figures are taken from the World Bank (2002). The dependent variable in the econometric analysis is the change in the migration stock as a percentage of the home population. By definition, this deviates from the net migration rate by the rate of natural population growth of the migrant population relative to that of the home population and the rate of naturalisations, i.e.

[21] To be precise, we divided the difference in the number of foreign residents between 1986 and 1989 by total net immigration in this time period, and multiplied this factor by annual net immigration in order to calculate the change in the number of foreign residents in each year.

$$\Delta mst_t = m_t + \frac{n_f - n_h - \delta}{1 + n_h} mst_{t-1},$$

where *mst* is defined as the ratio of the stock of residents to the home population, *m* as the ratio of net migration to the home population, n_f as the rate of natural population growth in the migrant population, n_h as the ratio of natural population growth in the home population, δ the rate of naturalisations in the migrant population. Thus, the change in the migrant stock equals the net migration rate if the difference of the rates of natural population growth in the migrant population and the population of the home countries and the rate of naturalisations is zero. In our sample the difference between the net migration rate and the change in migration stocks is moderate.

The explanatory variables in our model are per capita income and employment rates in Germany and the source countries. Consistent wage variables are not available for our country sample. Following the literature we thus used GDP per capita as an approximation for income levels. There is a long discussion in the literature as to whether income variables in purchasing power parities or at current exchange rates are appropriate measures for income. On the one hand, income measured at purchasing power parity reflects differences in the costs of living which affect real wages and the consumption of migrant households. On the other hand, part of the income is saved or remitted, and, hence, consumed in the home countries, such that GDP at current exchange rates might be an appropriate measure particularly if we consider optimising behaviour of migrants. In view of the large difference between the per capita GDP measured at purchasing power parity and that measured at current exchange rates in case of the CEECs, we used both variables to assess whether the choice of the income variable might affect our quantitative results.

The per capita GDP variables at current exchange rates stem from the OECD Historical Statistics and the OECD Main Economic Indicators (OECD 2002a, OECD 2002b) and are complemented by national sources for countries not covered by the OECD series (i.e. the (former) Yugoslavia). The per capita GDP at purchasing power parity (PPP-GDP) series is taken from Maddison (1995) for the period 1967-1994 and has been extrapolated with the real growth rate of the PPP-GDP per capita. The latter has been taken again from the OECD Main Economic Indicators and Historical Statistics.

The employment rate in the econometric analysis is calculated as one minus the unemployment rate. The ILO-definition for the unemployment rates have been used; time series for the unemployment rates stem from the OECD (2002a; 2002b) and are complemented by national statistical sources.

The data for the European sample refers to foreign workers as the migration stock variable instead to foreign residents in the German sample. The dependent variable is here the share of workers residing in another country as a percentage of the home labour force. The data are derived from the European Labour Force Survey, which is a household survey which is carried out in the EU-15 countries on an annual basis since 1993. We include all EU-15 countries as countries of destination, and 20 source countries (the EU-15 countries, Algeria, Morocco, Tunesia, (former) Yugoslavia, Turkey). However, not all countries of destination report observations for all source countries, such that our sample is limited to 215 cross-sections. Moreover, a large number of time-series observations are missing in the sample. For Austria, Finland and Sweden exist only observations since 1995, i.e. since their accession to the EU. There are however many other time-series observations missing in the data set. Finally, in many cases low response rates yield results which are not representative. We nevertheless decided to work with the largest number of possible observations rather than to exclude observations in an ad hoc manner.

As explanatory variables we used the same set of variables from the same data sources as in the German sample. In case of Algeria, Morocco and Tunesia we used additional data on employment and GDP from national statistical sources. As time-invariant variables we employed several distance variables (distance between capitals in miles, road distance in kilometers, travelling time by car), dummy variables for geographical location (*ADJACENT*), common language (*LANGUAGE*) and common border (*BORDER*). The *ADJACENT* variable captures the fact that migrants tend to move into the first geographical regions which offers a relatively high income, i.e. that the Portugese tend to move to France rather than to Germany while the Greeks tend to move to Germany rather than to France although the distance between Germany and France is relatively small. Note that this dummy variable produces significant results in many regressions, while simple distance variables do not.

The descriptive statistics for the variables in both the German and the European sample is available from the authors upon request.

Annex Tables

Table A1 **GDP and PPP-GDP per capita in the accession candidate countries, 1990-2000**

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
						GDP per capita					
Bulgaria	2 282	2 091	1 938	1 910	1 946	2 001	1 797	1 672	1 730	1 772	1 875
Czech Republic	6 171	5 452	5 427	5 427	5 552	5 877	6 133	6 083	6 014	5 989	6 164
Estonia	4 886	4 397	3 775	3 456	3 387	3 541	3 679	4 067	4 269	4 243	4 535
Hungary	5 379	4 738	4 593	4 565	4 704	4 771	4 838	5 055	5 301	5 524	5 813
Latvia	5 782	5 181	3 377	2 872	2 894	2 866	2 961	3 220	3 344	3 377	3 613
Lithuania	5 322	5 019	3 952	3 313	2 989	3 088	3 231	3 467	3 643	3 500	3 638
Poland	3 553	3 304	3 392	3 521	3 702	3 963	4 200	4 489	4 706	4 895	5 092
Romania	2 360	2 055	1 875	1 905	1 980	2 120	2 205	2 070	1 970	1 947	1 982
Slovak Republic	4 011	3 427	3 204	3 089	3 237	3 455	3 673	3 900	4 060	4 138	4 229
Slovenia	8 739	7 960	7 522	7 741	8 150	8 492	8 787	9 186	9 538	10 033	10 499
CEEC-8	4 736	4 294	4 115	4 093	4 216	4 441	4 650	4 886	5 066	5 199	5 408
CEEC-10	4 014	3 625	3 456	3 445	3 552	3 746	3 897	4 026	4 138	4 231	4 395
memo items:											
Cyprus	10 078	10 139	11 132	11 210	11 872	12 596	12 835	13 143	13 800	14 435	15 171
Malta	6 537	6 949	7 275	7 603	8 036	8 534	8 875	9 301	9 618	10 012	10 553
Turkey	2 223	2 429	2 451	2 598	2 806	2 652	2 843	3 042	3 270	3 371	3 213
CC-13	3 429	3 244	3 144	3 187	3 331	3 399	3 561	3 714	3 868	3 964	4 005
CC-10	4 793	4 361	4 197	4 178	4 309	4 541	4 752	4 990	5 176	5 316	5 534
						PPP-GDP per capita					
Bulgaria	7 910	7 248	6 718	6 622	6 744	6 935	6 230	5 795	5 995	6 143	6 500
Czech Republic	13 313	11 764	11 710	11 710	11 979	12 680	13 233	13 125	12 977	12 923	13 300
Estonia	10 558	9 501	8 157	7 468	7 318	7 652	7 950	8 789	9 226	9 168	9 800
Hungary	11 010	9 698	9 401	9 344	9 630	9 766	9 903	10 348	10 850	11 307	11 900
Latvia	12 322	11 041	7 197	6 119	6 167	6 107	6 311	6 862	7 125	7 197	7 700
Lithuania	12 728	12 004	9 450	7 923	7 147	7 384	7 726	8 292	8 713	8 371	8 700
Poland	6 419	5 969	6 128	6 361	6 688	7 160	7 588	8 111	8 503	8 844	9 200
Romania	7 023	6 116	5 580	5 669	5 893	6 309	6 562	6 160	5 863	5 796	5 900
Slovak Republic	10 528	8 994	8 411	8 109	8 498	9 070	9 642	10 236	10 657	10 862	11 100
Slovenia	13 319	12 130	11 464	11 797	12 420	12 942	13 391	14 000	14 536	15 290	16 000
CEEC-8	9 581	8 669	8 249	8 158	8 378	8 816	9 227	9 684	10 032	10 270	10 678
CEEC-10	8 884	7 994	7 549	7 497	7 711	8 126	8 416	8 619	8 818	8 985	9 324
memo items:											
Cyprus	12 289	12 363	13 575	13 670	14 476	15 359	15 651	16 027	16 828	17 602	18 500
Malta											
Turkey	3 597	3 932	3 967	4 205	4 542	4 292	4 601	4 923	5 292	5 456	5 200
CC-13	7 060	6 583	6 297	6 336	6 585	6 748	7 029	7 262	7 513	7 667	7 767
CC-10	9 560	8 661	8 259	8 170	8 395	8 836	9 243	9 698	10 048	10 290	10 701

Sources : OECD (2002), UNECE (2002), Eurostat (2002), authors' calculations.

Table A2 **Regression Results (German Sample, 1969-2001)**

	OLS		WITHIN		SUR		GLS(WALHUS)		GLS(SWAR)		GLS(MLE)	
	coefficient	p-value	coefficient	p-value	coefficient	p-value	coefficient	p-value	coefficient	p-value	coefficient	p-value
constant	-1.27	0.00	0.000									
$mst_{h,t-1}$	1.42	0.00	1.30	0.00	1.34	0.00	1.40	0.00	1.60	0.00	1.31	0.00
$mst_{h,t-2}$	-0.43	0.00	-0.42	0.00	-0.45	0.00	-0.43	0.00	-0.60	0.00	-0.42	0.00
$\ln(w_{ft}/w_{ht})$	0.18	0.00	0.35	0.00	0.25	0.00	0.20	0.00	0.11	0.01	0.35	0.00
$\ln(w_{ht})$	0.14	0.00	0.18	0.00	0.14	0.00	0.14	0.00	0.10	0.02	0.17	0.00
$\ln(e_{ft})$	1.09	0.00	1.08	0.00	0.74	0.00	1.10	0.00	0.94	0.00	1.07	0.00
$\ln(e_{ht})$	-0.09	0.29	-0.18	0.11	-0.10	0.00	-0.12	0.22	0.05	0.00	-0.16	0.20
$FREE_{ht}$	0.00	0.74	0.00	0.80	0.00	0.10	0.00	0.87	0.00	0.95	0.00	0.88
$GUEST_{ht}$	0.08	0.03	0.06	0.09	0.05	0.00	0.08	0.00	0.00	0.79	0.06	0.00
WAR_{ht}	0.40	0.08	0.43	0.03	0.43	0.00	0.40	0.00	0.42	0.00	0.41	0.00
$PEACE_{ht}$	-0.30	0.01	-0.17	0.24	-0.20	0.00	-0.31	0.00	-0.10	0.00	-0.21	0.00
$LANGUAGE_{h}$												
Sample:	1969-2001		1969-2001		1969-2001		1969-2001		1969-2001		1969-2001	
Included observations:	33		33		33		33		33		33	
Number of cross-sections used:	19		19		19		19		19		19	
Total panel (balanced) observations:	627		627		627		627		627		627	
No of parameters	11		10		10		11		11		11	
R-squared	0.995		0.996		0.996		0.992		1.000		0.961	
S.E. of regression	0.087		0.078		0.079		0.086		0.094		0.078	
sigma							0.086		0.094		0.078	
average thetha							0.198		-9.181		0.844	

The p-values are reported on the basis of the heteroscedasticity as well as heteroscedasticity and autocorrelation robust standard errors are reported for OLS, WITHIN, SUR and GLS, GMM 1-step estimators, respectively. The standard errors for GMM 2-step estimators are reported after smalll sample correction suggested in Windmeijer (2000).

Table A2 (cont.) **Regression Results (German Sample, 1969-2001)**

	GMM1(DIF)		GMM2(DIF)		GMM1(SYS)		GMM2(SYS)		GMM1(SYS)-TINV		GMM2(SYS)-TINV	
	coefficient	p-value	coefficient	p-value	coefficient	p-value	coefficient	p-value	coefficient	p-value	coefficient	p-value
constant												
$mst_{h,t-1}$	1.14	0.00	1.16	0.00	1.35	0.00	1.35	0.00	1.34	0.00	1.34	0.00
$mst_{h,t-2}$	-0.33	0.00	-0.35	0.00	-0.38	0.00	-0.36	0.00	-0.37	0.00	-0.36	0.00
$\ln(w_{ft}/w_{ht})$	0.86	0.00	0.58	0.10	0.24	0.02	0.23	0.03	0.26	0.02	0.24	0.02
$\ln(w_{ht})$	0.24	0.02	0.24	0.03	0.18	0.02	0.17	0.03	0.17	0.03	0.16	0.03
$\ln(e_{ft})$	1.29	0.02	1.51	0.01	1.40	0.01	1.13	0.01	1.37	0.01	1.11	0.03
$\ln(e_{ht})$	-0.41	0.49	-0.78	0.34	-0.14	0.42	0.01	0.93	-0.11	0.56	0.03	0.93
$FREE_{ht}$	0.07	0.33	0.10	0.33	0.00	0.96	0.00	0.73	0.02	0.21	0.01	0.57
$GUEST_{ht}$	-0.40	0.00	-0.40	0.00	0.11	0.01	0.11	0.07	0.10	0.02	0.10	0.08
WAR_{ht}	0.08	0.00	0.04	0.13	0.38	0.00	0.37	0.00	0.38	0.00	0.42	0.00
$PEACE_{ht}$	0.19	0.07	0.19	0.11	-0.37	0.00	-0.35	0.00	-0.36	0.00	-0.36	0.00
$LANGUAGE_{h}$									0.06	0.03	0.04	0.13
Sample:	1971-2001		1971-2001		1970-2001		1970-2001		1970-2001		1970-2001	
Included observations:	31		31		32		32		32		32	
Number of cross-sections used:	19		19		19		19		19		19	
Total panel (balanced) observations:	589		589		608		608		608		608	
No of parameters	10		10		11		11		12		12	
Wald-Test	$\chi^2(10)$	27610 **	$\chi^2(10)$	5619 **	$\chi^2(10)$	8E+05 **	$\chi^2(10)$	183200 **	$\chi^2(11)$	709400 **	$\chi^2(10)$	70850 **
S.E. of regression (levels)	0.068		0.068		0.088		0.088		0.088		0.087	
σ (transformed)	0.096		0.096									
Sargan-Test	$\chi^2(31)$	293.8 **	$\chi^2(31)$	9.662	$\chi^2(95)$	1438 **	$\chi^2(95)$	8.352	$\chi^2(95)$	1418 **	$\chi^2(95)$	7.887
AR(1) N(0,1)		-2.425 *		-2.183 *		-2.278 *		-2.017 *		-2.279 *		-2.043 *
AR(2) N(0,1)		1.504		1.752		1.137		1.126		1.08		1.215

The p-values are reported on the basis of the heteroscedasticity as well as heteroscedasticity and autocorrelation robust standard errors are reported for OLS, WITHIN, SUR and GLS, GMM 1-step estimators, respectively. The standard errors for GMM 2-step estimators are reported after smalll sample correction suggested in Windmeijer (2000).

Table A3 **Regression Results (European Sample, 1994-2001)**

	OLS		WITHIN		GLS(WALHUS)		GLS(SWAR)	
	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
constant	-0.43 *	0.17			-0.43	0.30	-0.28	0.15
$lst_{h,t-1}$	0.95 **	0.01	0.59 **	0.20	0.95 **	0.00	0.95938 **	0.00
$\ln(w_{ft}/w_{ht})$	0.03 **	0.01	0.23	0.14	0.03	0.03	0.02279 *	0.01
$\ln(w_{ht})$	0.04 **	0.01	-0.15	0.11	0.04	0.03	0.02362 *	0.01
$\ln(lf_{ht})$	0.01 **	0.00	0.07	0.07	0.01 **	0.00	0.00535 **	0.00
$ADJACENT_h$	0.03 *	0.01			0.03 *	0.01	0.02052 **	0.01
Sample period:	1994-2001		1994-2001		1994-2001		1994-2001	
Maximum observations:	8		8		8		8	
Minimum observations:	2		2		2		2	
Number of cross-sections used:	215		215		215		215	
(Unbalanced) panel observations:	1464		1464		1464		1464	
R-squared	0.98		0.39		0.98		0.99	
RSS	58.08		46.57		58.08		64.40	
Wald-Test			$\chi^2(4)$	15.61**	$\chi^2(5)$	72060**	$\chi^2(5)$	268900**
sigma (levels)	0.20		0.19		0.20		0.21	
avg. theta					0.00		-1.20	

The p-values are reported on the basis of the heteroscedasticity as well as heteroscedasticity and autocorrelation robust standard errors are reported for OLS, WITHIN, SUR and GLS, GMM 1-step estimators, respectively. The standard errors for GMM 2-step estimators are reported after smalll sample correction A49 suggested in Windmeijer (2000).

Apart from self-explanatory OLS and WITHIN estimators, GLS(WALHUS) - Wallace and Hussein (1969) estimator, GLS(SWAR) - Swamy and Arora (1972) estimator.

Table A3 (cont.)

	AH		GMM1(DIF)		GMM2(DIF)		GMM1(SYS)		GMM2(SYS)	
	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
constant							-0.59	0.19	-0.21	0.11
$lst_{h,t-1}$	0.64	0.05	0.76	0.06	0.73	0.06	0.94	0.01	0.94	0.01
$\ln(w_{ft}/w_{ht})$	0.08	0.11	0.15	0.15	0.09	0.08	0.05	0.01	0.02	0.01
$\ln(w_{ht})$	0.37	0.27	0.33	0.24	0.19	0.11	0.05	0.02	0.02	0.01
$\ln(lf_{ht})$	-0.46	0.35	-0.45	0.35	-0.17	0.14	0.01	0.00	0.00	0.00
$ADJACENT_h$							0.03	0.01	0.01	0.01
Sample period:	1996-2001		1996-2001		1996-2001		1996-2001		1996-2001	
Maximum observations:	6		6		6		6		6	
Minimum observations:	1		1		1		1		1	
Number of cross-sections used:	211		211		211		211		211	
(Unbalanced) panel observations:	1034		1034		1034		1245		1245	
RSS	64.52		72.64		70.51		49.30		49.47	
Wald-Test	$\chi^2(4)$	326.50	$\chi^2(4)$	581.6**	$\chi^2(4)$	138.4**	$\chi^2(5)$	24280**	$\chi^2(5)$	26580**
sigma (levels)	0.18		0.19		0.19		0.20		0.20	
Sargan-Test	$\chi^2(2)$ 98.69**		$\chi^2(6)$ 152.9**		$\chi^2(6)$	5.391	$\chi^2(13)$ 504.8**		$\chi^2(13)$	19.38
AR(2) N(0,1)			-1.01		-1.87		-2.202 *		-2.133 *	
AR(1) N(0,1)			-1.793		-1.013		-1.009		-1.005	

The p-values are reported on the basis of the heteroscedasticity as well as heteroscedasticity and autocorrelation robust standard errors are reported for OLS, WITHIN, SUR and GLS, GMM 1-step estimators, respectively. The standard errors for GMM 2-step estimators are reported after smalll sample correction suggested in Windmeijer (2000).

Apart from self-explanatory OLS and WITHIN estimators, GLS(WALHUS) - Wallace and Hussein (1969) estimator, GLS(SWAR) - Swamy and Arora (1972) estimator, AH - Anderson and Hsiao (1982), GMM1(DIF) and GMM2(DIF) - one- and two-steps Arellano and Bond (1991) estimator, GMM1(SYS) and GMM2(SYS) - one- and two-steps Arellano and Bover (1995) estimator.

Table A.4 **Germany: Migration Scenerio (based on SUR-Model)**

	2004	2005	2006	2007	2008	2009	2010	2015	2020	2030
					net increase of foreign population p.a. (persons)					
					High Scenario					
CEEC-10	216 955	271 119	268 516	241 953	206 954	171 595	139 595	47 543	21 865	- 3 181
CEEC-8	160 435	213 934	214 982	193 937	164 824	135 070	108 111	32 092	13 263	- 7 690
CEEC-2	56 519	57 184	53 534	48 017	42 130	36 525	31 484	15 450	8 601	4 509
					Baseline Scenario					
CEEC-10	179 787	224 308	221 215	198 405	168 739	138 939	112 078	35 556	14 870	- 7 323
CEEC-8	133 593	178 134	178 426	160 417	135 772	110 703	88 071	24 818	9 609	- 9 786
CEEC-2	46 195	46 174	42 788	37 988	32 967	28 235	24 008	10 738	5 261	2 463
					Low Scenario					
CEEC-10	154 953	193 182	189 504	168 822	142 346	115 948	92 293	26 016	9 318	- 9 772
CEEC-8	116 429	154 982	154 504	138 186	116 207	94 003	74 059	19 104	6 678	- 10 987
CEEC-2	38 525	38 200	35 000	30 636	26 139	21 945	18 234	6 913	2 640	1 215
					foreign population (persons)					
					High Scenario					
CEEC-10	815 704	1 086 822	1 355 338	1 597 292	1 804 246	1 975 841	2 115 436	2 496 246	2 642 632	2 783 974
CEEC-8	632 939	846 874	1 061 856	1 255 792	1 420 616	1 555 686	1 663 797	1 940 633	2 032 742	2 115 447
CEEC-2	182 764	239 949	293 482	341 499	383 629	420 155	451 639	555 614	609 889	668 527
					Baseline Scenario					
CEEC-10	778 536	1 002 844	1 224 058	1 422 463	1 591 203	1 730 141	1 842 220	2 138 397	2 242 136	2 332 446
CEEC-8	606 097	784 230	962 657	1 123 074	1 258 846	1 369 549	1 457 620	1 678 115	1 746 648	1 803 908
CEEC-2	172 440	218 613	261 402	299 390	332 357	360 592	384 600	460 283	495 488	528 538
					Low Scenario					
CEEC-10	753 702	946 885	1 136 389	1 305 211	1 447 557	1 563 506	1 655 798	1 887 476	1 956 922	2 011 484
CEEC-8	588 933	743 915	898 419	1 036 605	1 152 812	1 246 815	1 320 874	1 499 493	1 549 024	1 588 557
CEEC-2	164 770	202 970	237 970	268 606	294 745	316 690	334 924	387 983	407 898	422 927

CEEC-10: Bulgaria, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovak Republic, Slovenia. -
CEEC-8: CEEC-10 without Bulgaria and Romania. - CEEC-2: Bulgaria, Romania.

Sources : Authors' calculations. See text for assumptions of the scenario.

Table A.5 **Extrapolation of the migration scenario to the EU-15 (Baseline-Scenario)**

	2004	***2005***	***2006***	***2007***	***2008***	***2009***	***2010***	***2015***	***2020***	***2030***
					net increase of foreign population (persons)					
Austria	23 670	29 532	29 124	26 121	22 216	18 292	14 756	4 681	1 958	- 964
Belgium	3 977	4 961	4 893	4 388	3 732	3 073	2 479	786	329	- 162
Denmark	3 388	4 227	4 168	3 739	3 180	2 618	2 112	670	280	- 138
Finland	4 092	5 105	5 035	4 515	3 840	3 162	2 551	809	338	- 167
France	7 789	9 717	9 583	8 595	7 310	6 019	4 855	1 540	644	- 317
Germany	179 787	224 308	221 215	198 405	168 739	138 939	112 078	35 556	14 870	- 7 323
Greece	7 127	8 892	8 769	7 865	6 689	5 508	4 443	1 409	589	- 290
Ireland	71	88	87	78	66	55	44	14	6	- 3
Italy	32 342	40 351	39 794	35 691	30 355	24 994	20 162	6 396	2 675	- 1 317
Luxembourg	799	997	983	882	750	618	498	158	66	- 33
Netherlands	4 341	5 416	5 341	4 790	4 074	3 354	2 706	858	359	- 177
Portugal	290	362	357	320	272	224	181	57	24	- 12
Spain	4 892	6 104	6 020	5 399	4 592	3 781	3 050	968	405	- 199
Sweden	7 879	9 830	9 694	8 695	7 395	6 089	4 912	1 558	652	- 321
United Kingdom	13 807	17 226	16 989	15 237	12 959	10 670	8 607	2 731	1 142	- 562
EU-15 (est.)	294 250	367 114	362 052	324 720	276 168	227 394	183 433	58 193	24 338	- 11 986
					foreign population (persons)					
Austria	102 499	132 031	161 155	187 277	209 492	227 784	242 540	281 534	295 192	307 082
Belgium	17 220	22 182	27 075	31 463	35 196	38 269	40 748	47 299	49 594	51 591
Denmark	14 670	18 897	23 065	26 804	29 983	32 601	34 713	40 294	42 249	43 951
Finland	17 718	22 823	27 858	32 373	36 214	39 376	41 926	48 667	51 028	53 083
France	33 728	43 445	53 028	61 624	68 934	74 953	79 808	92 639	97 133	101 046
Germany	778 536	1 002 844	1 224 058	1 422 463	1 591 203	1 730 141	1 842 220	2 138 397	2 242 136	2 332 446
Greece	30 862	39 754	48 523	56 388	63 077	68 584	73 027	84 768	88 880	92 460
Ireland	306	395	482	560	626	681	725	842	882	918
Italy	140 051	180 402	220 196	255 887	286 242	311 235	331 397	384 676	403 338	419 584
Luxembourg	3 460	4 457	5 440	6 322	7 072	7 690	8 188	9 504	9 965	10 367
Netherlands	18 797	24 212	29 553	34 343	38 417	41 772	44 478	51 628	54 133	56 314
Portugal	1 256	1 617	1 974	2 294	2 566	2 790	2 971	3 449	3 616	3 762
Spain	21 185	27 289	33 308	38 707	43 299	47 080	50 130	58 189	61 012	63 469
Sweden	34 117	43 947	53 641	62 336	69 730	75 819	80 731	93 710	98 256	102 213
United Kingdom	59 789	77 015	94 003	109 240	122 199	132 869	141 476	164 221	172 188	179 124
EU-15 (est.)	1 274 195	1 641 309	2 003 361	2 328 081	2 604 249	2 831 644	3 015 077	3 499 818	3 669 602	3 817 409

Sources : Authors' calculations. See text for assumptions.

Table A6 **Simulation of different accession scenarios (Germany, baseline assumptions)**

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2020	2025	2030
								net increase of foreign population (persons)							
								CEEC-8							
free movement in 2004	133 593	178 134	178 426	160 417	135 772	110 703	88 071	68 934	53 509	41 317	31 918	24 818	9 609	7 286	- 9 786
transitional period of															
2 years	..	..	133 730	177 904	179 541	161 811	137 191	112 025	89 316	70 026	54 341	41 979	13 297	7 947	- 9 657
5 years	..	..				135 431	180 037	181 705	163 848	138 995	113 568	90 540	25 899	10 369	- 9 223
7 years	..	..						136 548	181 565	183 273	165 256	140 218	43 288	14 084	- 8 556
								CEEC-2							
free movement in 2004	46 195	46 174	42 788	37 988	32 967	28 235	24 008	20 345	17 312	14 731	12 557	10 738	5 261	3 248	2 463
transitional period of															
2 years	..	..		45 573	45 482	42 088	37 439	32 564	28 032	23 954	20 405	17 377	8 174	4 660	3 211
5 years	..	..							44 731	44 660	41 409	36 949	17 524	8 823	5 268
7 years	..	..									44 547	44 453	23 926	11 623	6 560
								CEEC-10							
free movement in 2004	179 787	224 308	221 215	198 405	168 739	138 939	112 078	89 278	70 821	56 048	44 476	35 556	14 870	10 533	- 7 323
transitional period of															
2 years	..	..	133 730	223 477	225 023	203 898	174 630	144 589	117 348	93 980	74 747	59 357	21 471	12 607	- 6 446
5 years	..	..	..	..	..	135 431	180 037	181 705	208 579	183 655	154 977	127 490	43 424	19 192	- 3 955
7 years	..	..	..	..	..	..	..	136 548	181 565	183 273	209 804	184 671	67 214	25 706	- 1 996
								foreign population (persons)							
								CEEC-8							
free movement in 2004	606 097	784 230	962 657	1 123 074	1 258 846	1 369 549	1 457 620	1 526 553	1 580 062	1 621 378	1 653 297	1 678 115	1 746 648	1 786 395	1 803 908
transitional period of															
2 years	472 504	472 504	606 234	784 138	963 680	1 125 490	1 262 681	1 374 706	1 464 022	1 534 048	1 588 389	1 630 368	1 737 269	1 784 452	1 803 311
5 years	472 504	472 504	472 504	472 504	472 504	607 935	787 972	969 677	1 133 525	1 272 520	1 386 088	1 476 628	1 704 404	1 778 082	1 801 765
7 years	472 504	472 504	472 504	472 504	472 504	472 504	472 504	609 052	790 617	973 889	1 139 146	1 279 364	1 656 171	1 768 627	1 799 799
								CEEC-2							
free movement in 2004	172 440	218 613	261 402	299 390	332 357	360 592	384 600	404 944	422 257	436 988	449 545	460 283	495 488	514 957	528 538
transitional period of															
2 years	126 245	126 245	126 245	171 818	217 299	259 387	296 826	329 390	357 422	381 377	401 782	419 159	475 052	504 061	522 517
5 years	126 245	126 245	126 245	126 245	126 245	126 245	126 245	126 245	170 976	215 636	257 045	293 994	415 906	474 239	506 482
7 years	126 245	126 245	126 245	126 245	126 245	126 245	126 245	126 245	126 245	126 245	170 792	215 245	377 221	455 685	496 853
								CEEC-10							
free movement in 2004	778 536	1 002 844	1 224 058	1 422 463	1 591 203	1 730 141	1 842 220	1 931 498	2 002 318	2 058 366	2 102 842	2 138 397	2 242 136	2 301 352	2 332 446
transitional period of															
2 years	598 749	598 749	732 479	955 956	1 180 979	1 384 877	1 559 507	1 704 096	1 821 444	1 915 424	1 990 171	2 049 527	2 212 321	2 288 514	2 325 828
5 years	598 749	598 749	598 749	598 749	598 749	734 180	914 217	1 095 922	1 304 501	1 488 156	1 643 133	1 770 622	2 120 309	2 252 322	2 308 247
7 years	598 749	598 749	598 749	598 749	598 749	598 749	598 749	735 297	916 862	1 100 134	1 309 938	1 494 609	2 033 393	2 224 312	2 296 652