INTERNATIONAL MIGRATION PRESSURES IN THE LONG RUN

2017

Rodolfo G. Campos

Documentos de Trabajo
N.º 1734

BANCO DE ESPAÑA
Eurosistema

INTERNATIONAL MIGRATION PRESSURES IN THE LONG RUN

INTERNATIONAL MIGRATION PRESSURES IN THE LONG RUN (*)

Rodolfo G. Campos (**)

BANCO DE ESPAÑA

(*) I am grateful for comments by Iliana Reggio, Jan Stuhler, and an anonymous referee, and by seminar participants at Banco de España, the Informal 4NCBs Workshop at Banque de France, and the FREE seminar. Manuel Montesinos Calvo-Fernández provided outstanding research assistance for this project. The views expressed in this paper are those of the author and do therefore not necessarily reflect the views of the Banco de España or the Eurosystem.
(**) Banco de España, Calle de Alcalá 48, 28014 Madrid, Spain; email: rodolfo.campos@bde.es.

Documentos de Trabajo. N.º 1734

2017

ISSN: 1579-8666 (on line)

Abstract

Using an empirical gravity model, I estimate the contribution of changes in relative labor supply to bilateral migration in the 2000s and apply the resulting estimates to project future bilateral flows based on population forecasts by the United Nations. I extend the work of Hanson and McIntosh (2016) by including non-OECD destinations and project international migration flows for the whole world. In contrast to their findings, and despite of the slowdown of population growth in Latin America, the US will face sustained immigration pressures because of strong population growth in other regions of the world, leading to a projected immigrant stock that grows for decades to come. For the world as a whole, international migrants are projected to increase from 2.8% of total world population in 2010 to 3.5% in 2050, with a substantial increase of migrants originating from India and Sub-Saharan Africa.

Keywords: international migration, gravity model, population growth.

JEL classification: F22, J61.

Resumen

En este trabajo se estima la contribución del crecimiento demográfico a la migración bilateral por medio de un modelo gravitacional y se utilizan los resultados para proyectar flujos migratorios basados en las previsiones de crecimiento demográfico de las Naciones Unidas. Para ello se extiende la metodología de Hanson y McIntosh (2016), y se incluyen países de destino que no pertenecen a la OCDE, para así obtener proyecciones para prácticamente todos los países del mundo. En contraposición con los resultados de Hanson y McIntosh, y a pesar del menor crecimiento demográfico esperado en América Latina, Estados Unidos continuará expuesto a presiones inmigratorias a causa del fuerte crecimiento demográfico en otras regiones del mundo. También se proyecta un incremento del número de migrantes a nivel global, de 2,8 % de la población mundial en 2010 a 3,5 % en 2050, fundamentado en un aumento sustancial de migrantes provenientes de la India y el África subsahariana.

Palabras clave: migración internacional, modelo gravitacional, crecimiento demográfico.

Códigos JEL: F22, J61.

1 Introduction

Both immigration and emigration flows play an important role in shaping a country's population. However, in comparison to the evolution of the native population, which can be projected using birth rates and death rates that evolve smoothly over time, the evolution of migration is more difficult to anticipate, and projections of population often assume excessively simplistic dynamics for the evolution of international migration flows. For example, the widely used population projections by the United Nations (*World Population Prospects. The 2015 Revision*) assume that net migration flows will remain constant at their current level until 2050 and that migration flows then decline gradually to reach 50% of this level in 2100 (United Nations, 2015).

The objective of this work is to provide a framework to project migration flows and stocks in future decades in a richer framework that is based on long-term population pressures and bilateral migration determinants, such as geographical distance and cultural proximity measures. These projections are a valuable input to construct migration and population scenarios that are, in turn, useful to project long-term potential growth rates of particular countries and to evaluate the impact on economic policy of a changing population structure.

Past research suggests that, in addition to economic disparities, population pressure that builds up because of the birth of large cohorts of future workers intensifies outward migration as these workers enter the labor force (Hatton and Williamson, 2003). Hanson and McIntosh (2016) use this idea to project international migration flows for selected OECD countries available in the DIOC database (Dumont and Lemaître, 2005; OECD, 2008).[1] In this paper, I extend their work to virtually all countries in the world. I do so by estimating an enriched bilateral gravity model of international migration that ties migration flows to the relative evolution of population growth rates in origin and destination countries. In comparison to the methodology of Hanson and McIntosh (2016), who project future migration flows for a limited sample of OECD country destinations using cohort-specific bilateral migration data, I show how through equation averaging their method can be generalized for constructing projections for any arbitrary destination country by removing the need for cohort-specific migration observations.

For the United States I find that immigration pressures do not subside, despite of the slowdown of population growth in Latin America, and in particular in Mexico, the primary origin of immigrants. Immigrants from Mexico in the US are projected to be replaced by immigrants from other regions of the world. This result stands in contrast with the findings by Hanson and McIntosh (2016), who project a stagnating immigrant stock in the US. I show that the discrepancy is partly explained by the inclusion of non-OECD countries that lower the estimate of the influence of population growth rates on bilateral migration flows.

[1] DIOC stands for *Database on Immigrants in OECD Countries.*

For the world as a whole, I project international migrants to increase from 2.8% of total world population in 2010 to 3.5% in 2050. The total stock of migrants is projected to grow from 190 million in 2010 to 334 million in 2050. In terms of origins, India and countries in Sub-Saharan Africa explain a large part of the increase in the number of total migrants whereas former USSR countries and countries in Eastern Europe are projected to reduce the number of migrants they supply.

The paper is structured as follows: I present the specification that I estimate to produce the projections and show how it can be derived from the cohort-based specification by Hanson and McIntosh (2016) in Section 2. In Section 3, I describe international migration patterns and highlight the importance of non-OECD countries as hosts of large stocks of international migrants. In Section 4, I report the results of the estimation as well as projections for several countries and for the whole world. In Section 5, I discuss how the methodology of this paper can be applied and highlight some of the caveats of the resulting projections. I conclude in Section 6.

2 Methodology

My baseline specification is directly derived from the specification estimated by Hanson and McIntosh (2016) for OECD destinations. Their specification requires the observation of repeated cohort-specific migration rates, which are only available for OECD countries (from the DIOC database). However, if their specification is averaged over cohort-gender cells, then the need for cohort-specific migration rates disappears. The advantage of this averaged specification is that it requires data that are less granular and can therefore be estimated for a wider set of countries, using the Global Bilateral Migration database from the World Bank.

2.1 Specification

In the baseline specification, the change in the bilateral migration rate is related to changes in birth cohorts, relative GDP levels, and geographical distance variables τ^i in the following way:

$$\begin{aligned} m_{sd,t+1} - m_{sd,t} = \eta_s + \eta_d + \lambda \left(\ln \frac{L^0_{s,t-1}}{L^0_{s,t-5}} - \ln \frac{L^0_{d,t-1}}{L^0_{d,t-5}} \right) + \kappa \left(\ln \frac{GDP_{s,T_0}}{GDP_{d,T_0}} \right) + \gamma_i \tau^i_{sd} \\ + \phi_i \left[\left(\ln \frac{L^0_{s,t-1}}{L^0_{s,t-5}} - \ln \frac{L^0_{d,t-1}}{L^0_{d,t-5}} \right) \times \tau^i_{sd} \right] + \epsilon_{sd}, \end{aligned} \tag{1}$$

where time t is measured in decades. The outcome of interest is the change in the bilateral migration rate after a decade: $m_{sd,t+1} - m_{sd,t}$, where $m_{sd,t}$ is the fraction of the population from the source country s that resided in the destination country d in decade t and $m_{sd,t+1}$ is the fraction of the population from the source country s that resided in the destination country d a decade later, in $t+1$.

On the right hand side, the variables η_s, and η_d are fixed effects for origin and destination countries and they absorb time trends that are specific to a particular country. The variable $\left(\ln \frac{L^0_{s,t-1}}{L^0_{s,t-5}} - \ln \frac{L^0_{d,t-1}}{L^0_{d,t-5}}\right)$ is the difference in the growth rate of birth cohorts over four decades. This difference captures the change in the relative abundance of workers in the source and destination countries. Assuming, as Hanson and McIntosh (2016) do, that virtually all emigration decisions occur between the ages of 15 and 54, the growth of a birth cohort between $t-5$ and $t-1$ measures the growth in the number of people who were born in years that put them into that age group. So, for example, if $t = 2000$ and $t+1 = 2010$, then the growth rate $\ln \frac{L^0_{s,t-1}}{L^0_{s,t-5}}$ measures the growth rate of birth cohorts between $t-5 = 1950$, when those aged 35-54 in 2000 were aged 14 or less, and $t-1 = 1990$, when those aged 15-24 were aged 14 or less.

The ratio $\frac{GDP_{s,T_0}}{GDP_{d,T_0}}$ captures the ratio of per-capita GDP (PPP) across countries in a year T_0 that lies sufficiently far in the past to be plausibly exogenous (I use $T_0 = t-3$) The variables τ^i_{sd} refer to bilateral geographical and cultural distance measures. I use the logarithm of the distance between two countries, a dummy variable indicating whether the countries share a common border, a dummy variable for a colonial relationship, and a dummy variable for sharing a common language. In the variables τ^i_{sd} I also include dummy variables for the presence of prior migration networks, which are known to be important enhancer of bilateral migration (McKenzie and Rapoport, 2007; Beine, Docquier, and Özden, 2011; Docquier, Peri, and Ruyssen, 2014). I rank country pairs according to their migration stock in 1960 and follow Hanson and McIntosh (2016) in constructing indicators for belonging to the top 5%, top 20%, and top 50% of this ranking. The interaction of the relative labor supply $\left(\ln \frac{L^0_{s,t-1}}{L^0_{s,t-5}} - \ln \frac{L^0_{d,t-1}}{L^0_{d,t-5}}\right)$ with these distance and prior migration network indicators captures the intuition that a relative abundance of labor will be more conducive to migration if bilateral migration costs are low.

I estimate the model for $t = 2000$ and $t + 1 = 2010$ and use the estimated coefficients to project migration thereafter. For clarity, and explicitly labeling time variables, the equation that I estimate is:

$$
\begin{aligned}
m_{sd,2010} - m_{sd,2000} = \eta_s + \eta_d + \lambda \left(\ln \frac{L^0_{s,1990}}{L^0_{s,1950}} - \ln \frac{L^0_{d,1990}}{L^0_{d,1950}} \right) \\
+ \kappa \left(\ln \frac{GDP_{s,1970}}{GDP_{d,1970}} \right) + \gamma_i \tau^i_{sd} \\
+ \phi_i \left[\left(\ln \frac{L^0_{s,1990}}{L^0_{s,1950}} - \ln \frac{L^0_{d,1990}}{L^0_{d,1950}} \right) \times \tau^i_{sd} \right] + \epsilon_{sd}. \qquad (2)
\end{aligned}
$$

2.2 Derivation from the specification used by Hanson and McIntosh (2016)

Hanson and McIntosh (2016) divide the population into cohort-gender cells and specify a relationship between the growth of cohort- and gender-specific migration rates, birth cohort ratios, initial GDP ratios, and distance variables. Denote gender by g, a cohort by c, the source country by s, and the destination country by d. Cohorts are defined by ten year intervals, e.g. those aged 15–24, 25–34, 35–44, 45–54 in the year 2000. Time is denoted by t and measured in decades. The specification with notation taken directly from Hanson and McIntosh (2016) is

$$
\begin{aligned}
y_{gcsd,t+1} - y_{gcsd,t} = \delta_{gc} + \delta_s + \delta_d + \alpha \left(\ln \frac{L_{gcs,t}}{L_{gcd,t}} - \ln \frac{L_{gcs,t-1}}{L_{gcd,t-1}} \right) \\
+ \beta \left(\ln \frac{w_{cs,t}}{w_{cd,t}} \right) + \gamma \tau_{sd} \\
+ \rho \left[\left(\ln \frac{L_{gcs,t}}{L_{gcd,t}} - \ln \frac{L_{gcs,t-1}}{L_{gcd,t-1}} \right) \times \tau_{sd} \right] + \varepsilon_{gcsd}. \qquad (3)
\end{aligned}
$$

The dependent variable is $y_{gcsd,t+1} - y_{gcsd,t}$, the change of the gender-cohort-specific emigration rate over ten years. For example, the emigration rate to the US for Mexico-born men who were aged 25–34 in 2010, minus that for those who were aged 15–24 in 2000. The variables δ_{gc}, δ_s, and δ_d are gender-cohort, source country, and destination country dummies. The expression $\left(\ln \frac{L_{gcs,t}}{L_{gcd,t}} - \ln \frac{L_{gcs,t-1}}{L_{gcd,t-1}} \right)$ measures the change in the relative size of gender and cohort-specific birth-cohort ratios. For example, the log Mexico/US birth cohort ratios for men aged 15–24 in 2000 minus the log birth cohort ratio for men aged 15–24 in 1990. The ratio $\frac{w_{cs,t}}{w_{cd,t}}$ captures living standard differences at the start of a cohort's life, which are proxied by relative per-capita GDP measured in the year a cohort was aged 15 or less. Finally, τ_{sd} are distance variables, such as geographical distance, common language, etc.[2]

[2]See Hanson and McIntosh (2016) for additional details on the definition of variables.

Starting from (3), for each gender g, I average over cohorts $c = 1, \ldots C$ to obtain

$$\frac{1}{C}\sum_c y_{gcsd,t+1} - \frac{1}{C}\sum_c y_{gcsd,t} = \frac{1}{C}\sum_c \delta_{gc} + \delta_s + \delta_d + \alpha\frac{1}{C}\sum_c\left(\ln\frac{L_{gcs,t}}{L_{gcd,t}} - \ln\frac{L_{gcs,t-1}}{L_{gcd,t-1}}\right)$$
$$+ \beta\frac{1}{C}\sum_c\left(\ln\frac{w_{cs,t}}{w_{cd,t}}\right) + \gamma\tau_{sd}$$
$$+ \rho\frac{1}{C}\sum_c\left[\left(\ln\frac{L_{gcs,t}}{L_{gcd,t}} - \ln\frac{L_{gcs,t-1}}{L_{gcd,t-1}}\right)\times\tau_{sd}\right] + \frac{1}{C}\sum_c \varepsilon_{gcsd}. \quad (4)$$

Using bars to denote averages, the equation can be expressed as

$$\bar{y}_{gsd,t+1} - \bar{y}_{gsd,t} = \bar{\delta}_g + \delta_s + \delta_d + \frac{\alpha}{C}\sum_c\left(\ln\frac{L_{gcs,t}}{L_{gcd,t}} - \ln\frac{L_{gcs,t-1}}{L_{gcd,t-1}}\right)$$
$$+ \beta\frac{1}{C}\sum_c\left(\ln\frac{w_{cs,t}}{w_{cd,t}}\right) + \gamma\tau_{sd}$$
$$+ \frac{\rho}{C}\sum_c\left[\left(\ln\frac{L_{gcs,t}}{L_{gcd,t}} - \ln\frac{L_{gcs,t-1}}{L_{gcd,t-1}}\right)\times\tau_{sd}\right] + \bar{\varepsilon}_{gsd}. \quad (5)$$

The next step to simplify this expression is to write out the summation $\sum_{c\in\mathbf{C}}\left(\ln\frac{L_{gcs,t}}{L_{gcd,t}} - \ln\frac{L_{gcs,t-1}}{L_{gcd,t-1}}\right)$ and to note that it is equivalent to the difference in gender-specific population growth rates between the sending and the destination countries over a 40-year period.

To exemplify, I focus on $t = 2000$ and $t + 1 = 2010$. In the year 2000, those who were aged 15–24, 25–34, 35–44, 45–54 were aged less than 14 in the years 1990, 1980, 1970, and 1960, respectively. This is exemplified in the following table:

Age in 2000	Age in 2010	Aged 5–14 ("born") in:
15–24	25–34	1990
25–34	35–44	1980
35–44	45–54	1970
45–54	55–64	1960

I use the notation $L^0_{gs,j}$ for the cohort of gender g that was "born" (was aged less than 14) in the year j in country s. As an example, using this notation, the birth cohort of those aged 15–24 in the year 2000 is $L_{g,15-24,s,2000} = L^0_{gs,1990}$ and that of those aged 45–54 in the year 2000 is $L_{g,45-54,s,2000} = L^0_{gs,1960}$.

Summing over all cohorts $c \in \mathbf{C} \equiv \{15-24, 25-34, 35-44, 45-54\}$:

$$\sum_{c\in\mathbf{C}} \left(\ln \frac{L_{gcs,t}}{L_{gcd,t}} - \ln \frac{L_{gcs,t-1}}{L_{gcd,t-1}} \right) = \sum_{j=1960}^{1990} \left(\ln \frac{L^0_{gs,j}}{L^0_{gd,j}} - \ln \frac{L^0_{gs,j-1}}{L^0_{gd,j-1}} \right)$$
$$= \ln \frac{L^0_{gs,1990}}{L^0_{gd,1990}} - \ln \frac{L^0_{gs,1950}}{L^0_{gd,1950}}$$
$$= \ln \frac{L^0_{gs,1990}}{L^0_{gs,1950}} - \ln \frac{L^0_{gd,1990}}{L^0_{gd,1950}} \tag{6}$$

The last expression is simply the difference across countries of growth rates of birth cohorts between the years 1950 and 1990. It does not depend on the cohort structure, just on growth in birth rates over four decades. This implies that, although variation across cohorts is used in the identification of the coefficients, cross-cohort differences in population size (or any other measure of the evolution of the population structure) are not used in the projection step when the variable projected is an average over cohorts.

The elements in $\frac{1}{C}\sum_c \left(\ln \frac{w_{cs,t}}{w_{cd,t}}\right)$ are log-ratios of (PPP-adjusted) per-capita GDP in the sending and destination countries. These ratios are calculated for the year in which a cohort was aged 14 or less. As with the birth cohorts, the year at which GDP ratios are measured for those who in 2000 were aged 15–24, 25–34, 35–44, 45–54 are the years 1990, 1980, 1970, and 1960, respectively. The GDP-term is the average of four GDP ratios:

$$\frac{1}{C}\sum_{c\in\mathbf{C}} \left(\ln \frac{w_{cs,t}}{w_{cd,t}} \right) = \frac{1}{4}\sum_{j=1960}^{1990} \left(\ln \frac{GDP_{s,j}}{GDP_{d,j}} \right) \tag{7}$$

The resulting equation is:

$$\bar{y}_{gsd,2010} - \bar{y}_{gsd,2000} = \tilde{\delta}_g + \delta_s + \delta_d + \frac{\alpha}{C}\left(\ln \frac{L^0_{gs,1990}}{L^0_{gs,1950}} - \ln \frac{L^0_{gd,1990}}{L^0_{gd,1950}} \right)$$
$$+ \beta \frac{1}{4}\sum_{j=1960}^{1990} \left(\ln \frac{GDP_{s,j}}{GDP_{d,j}} \right) + \gamma \tau_{sd}$$
$$+ \frac{\rho}{C}\left[\left(\ln \frac{L^0_{gs,1990}}{L^0_{gs,1950}} - \ln \frac{L^0_{gd,1990}}{L^0_{gd,1950}} \right) \times \tau_{sd} \right] + \bar{\varepsilon}_{gsd}. \tag{8}$$

This equation is still distinguished by gender $g = f, m$. Averaging the equations for females and males yields:

$$
\begin{aligned}
\bar{y}_{sd,2010} - \bar{y}_{sd,2000} = {} & \frac{\bar{\delta}_f + \bar{\delta}_m}{2} + \delta_s + \delta_d \\
& + \frac{\alpha}{2C}\left(\ln\frac{L^0_{fs,1990}}{L^0_{fs,1950}} - \ln\frac{L^0_{fd,1990}}{L^0_{fd,1950}} + \ln\frac{L^0_{ms,1990}}{L^0_{ms,1950}} - \ln\frac{L^0_{md,1990}}{L^0_{md,1950}}\right) \\
& + \beta\frac{1}{4}\sum_{j=1960}^{1990}\left(\ln\frac{GDP_{s,j}}{GDP_{d,j}}\right) + \gamma\tau_{sd} \\
& + \frac{\rho}{2C}\left[\left(\ln\frac{L^0_{fs,1990}}{L^0_{fs,1950}} - \ln\frac{L^0_{fd,1990}}{L^0_{fd,1950}} + \ln\frac{L^0_{ms,1990}}{L^0_{ms,1950}} - \ln\frac{L^0_{md,1990}}{L^0_{md,1950}}\right) \times \tau_{sd}\right] \\
& + \frac{\bar{\varepsilon}_{fsd} + \bar{\varepsilon}_{msd}}{2}, \qquad (9)
\end{aligned}
$$

where $\bar{y}_{sd,2010} - \bar{y}_{sd,2000}$ is the growth of the average of the cohort- and gender-specific sending out rates of the immigrant population between the years 2000 and 2010. In aggregate data I only have access to the sending-out rate of the total population and its growth $m_{sd,2010} - m_{sd,2000}$. Because cohorts are not of the same size, in general, $m_{sd,2010} - m_{sd,2000} \neq \bar{y}_{sd,2010} - \bar{y}_{sd,2000}$. However, these two expressions differ by an error term e_s that is specific only to the population structure in the source country (and does not depend on the destination country), so that

$$
m_{sd,2010} - m_{sd,2000} = \bar{y}_{sd,2010} - \bar{y}_{sd,2000} + e_s \qquad (10)
$$

Using this result, the specification of Hanson and McIntosh (2016) averaged over cohorts and genders turns out to be:

$$
\begin{aligned}
m_{sd,2010} - m_{sd,2000} = {} & \frac{\bar{\delta}_f + \bar{\delta}_m}{2} + (\delta_s - e_s) + \delta_d \\
& + \frac{\alpha}{2C}\left(\ln\frac{L^0_{fs,1990}}{L^0_{fs,1950}} - \ln\frac{L^0_{fd,1990}}{L^0_{fd,1950}} + \ln\frac{L^0_{ms,1990}}{L^0_{ms,1950}} - \ln\frac{L^0_{md,1990}}{L^0_{md,1950}}\right) \\
& + \beta\frac{1}{4}\sum_{j=1960}^{1990}\left(\ln\frac{GDP_{s,j}}{GDP_{d,j}}\right) + \gamma\tau_{sd} \\
& + \frac{\rho}{2C}\left[\left(\ln\frac{L^0_{fs,1990}}{L^0_{fs,1950}} - \ln\frac{L^0_{fd,1990}}{L^0_{fd,1950}} + \ln\frac{L^0_{ms,1990}}{L^0_{ms,1950}} - \ln\frac{L^0_{md,1990}}{L^0_{md,1950}}\right) \times \tau_{sd}\right] \\
& + \frac{\bar{\varepsilon}_{fsd} + \bar{\varepsilon}_{msd}}{2}. \qquad (11)
\end{aligned}
$$

Moreover, because in practice the growth rates of females and males are almost identical, so that $\ln \frac{L^0_{fs,1990}}{L^0_{fs,1950}} \approx \ln \frac{L^0_{ms,1990}}{L^0_{ms,1950}}$ and $\ln \frac{L^0_{fd,1990}}{L^0_{fd,1950}} \approx \ln \frac{L^0_{md,1990}}{L^0_{md,1950}}$, the expression can be further be simplified to

$$\begin{aligned} m_{sd,2010} - m_{sd,2000} = & \frac{\bar{\delta}_f + \bar{\delta}_m}{2} + (\delta_s - e_s) + \delta_d \\ & + \frac{\alpha}{C}\left(\ln \frac{L^0_{s,1990}}{L^0_{s,1950}} - \ln \frac{L^0_{d,1990}}{L^0_{d,1950}}\right) \\ & + \beta \frac{1}{4} \sum_{j=1960}^{1990} \left(\ln \frac{GDP_{s,j}}{GDP_{d,j}}\right) + \gamma \tau_{sd} \\ & + \frac{\rho}{C}\left[\left(\ln \frac{L^0_{s,1990}}{L^0_{s,1950}} - \ln \frac{L^0_{d,1990}}{L^0_{d,1950}}\right) \times \tau_{sd}\right] \\ & + \frac{\bar{\varepsilon}_{fsd} + \bar{\varepsilon}_{msd}}{2}. \end{aligned} \tag{12}$$

This can be written as

$$\begin{aligned} m_{sd,2010} - m_{sd,2000} = & \eta + \eta_s + \eta_d \\ & + \lambda \left(\ln \frac{L^0_{s,1990}}{L^0_{s,1950}} - \ln \frac{L^0_{d,1990}}{L^0_{d,1950}}\right) \\ & + \beta \frac{1}{4} \sum_{j=1960}^{1990} \left(\ln \frac{GDP_{s,j}}{GDP_{d,j}}\right) + \gamma \tau_{sd} \\ & + \phi \left[\left(\ln \frac{L^0_{s,1990}}{L^0_{s,1950}} - \ln \frac{L^0_{d,1990}}{L^0_{d,1950}}\right) \times \tau_{sd}\right] \\ & + \epsilon_{sd}, \end{aligned} \tag{13}$$

where $\eta = \frac{\bar{\delta}_f + \bar{\delta}_m}{2}$, $\eta_s = \delta_s - e_s$, $\eta_d = \delta_d$, $\lambda = \frac{\alpha}{C}$, $\phi = \frac{\rho}{C}$, and $\epsilon_{sd} = \frac{\bar{\varepsilon}_{fsd} + \bar{\varepsilon}_{msd}}{2}$.

Finally, because the average of the per-capita GDP ratio may be subject to reverse causality, it is convenient to replace it with the per-capita GDP ratio for a year T_0 that sufficiently far in the past. The natural candidate is to use the year 1960. However, because of a wider availability of data I use $T_0 = 1970$. This year lies sufficiently far in the past to make the argument that the change in the share of migrants between 2000 and 2010 does not affect GDP 30 or 40 years before.

2.3 Data sources and projection details

I use decennial bilateral migration stocks from the GBM database originally compiled by Özden, Parsons, Schiff, and Walmsley (2011) until 2000 and the World Bank's update for 2010.[3] I estimate the model by weighted least squares with weights given by population in the source country. Past and projected birth cohort sizes are taken from the United Nations' *World Population Prospects. The 2015 Revision.*[4] I use the United Nations' birth cohort estimates for the period 1950–2010 and birth cohort forecasts for the period 2020-2050 (from the zero-migration scenario). Based on the parameters estimated from the empirical model, I project bilateral migration flows into the future until the year 2050 based on the change in birth cohort sizes projected by the United Nations. I then recursively add up bilateral migration stocks and consolidate them either at the source or destination country to obtain projected emigrant or immigrant stocks. To obtain the best possible fit, I estimate an equation that jointly makes use of all distance variables τ^i and their interactions with relative birth cohort growth rates and then project the change in the bilateral migration rate according to the expression:

$$(m_{sd,t+1} - m_{sd,t})\widehat{\ } = \hat{\eta}_s + \hat{\eta}_d + \hat{\lambda}\left(\ln\frac{L^0_{s,t-1}}{L^0_{s,t-5}} - \ln\frac{L^0_{d,t-1}}{L^0_{d,t-5}}\right) + \hat{\kappa}\left(\ln\frac{GDP_{s,t-3}}{GDP_{d,t-3}}\right) + \sum_i \hat{\gamma}_i \tau^i_{sd}$$
$$+ \sum_i \hat{\phi}_i \left[\left(\ln\frac{L^0_{s,t-1}}{L^0_{s,t-5}} - \ln\frac{L^0_{d,t-1}}{L^0_{d,t-5}}\right) \times \tau^i_{sd}\right] + \epsilon_{sd}. \quad (14)$$

I take GDP data from the Penn World Table (version 9.0). I use expenditure-side real GDP ($RGDP^e$), which, according to Feenstra, Inklaar, and Timmer (2015), is ideal to compare living standards across countries and across years. I compute per-capita GDP using total population counts for each country from the United Nations' dataset rather than from the Penn World Table to be consistent with the source for birth cohort sizes. For several countries the coverage of GDP in the Penn World Table starts only in 1990. This is the case for countries that were previously joined in the USSR, Yugoslavia, and Czechoslovakia. Because these countries have important bilateral migration stocks, I impute relative GDP ratios in the pre-1990 years to be able to use them in the projections.[5] Specifically, for country pairs that include one of these countries I set the GDP ratio in 1970 and in 1980 to the value it had in 1990. Bilateral distances and all other geographical data used in the gravity equation are taken from the CEPII gravity dataset by Head, Mayer, and Ries (2010).

[3] These databases are available at `http://data.worldbank.org/data-catalog/global-bilateral-migration-database` and `http://siteresources.worldbank.org/INTPROSPECTS/Resources/334934-1110315015165/T1.Estimates_of_Migrant_Stocks_2010.xls`. Both databases are ultimately based on national sources but differ in methodology. To the best of my knowledge, there has not been a study on the relative data quality of these two sources.

[4] Available at `https://esa.un.org/unpd/wpp/`.

[5] As a robustness exercise, I also exclude countries with missing data and find similar results for countries left in the sample.

3 International migration patterns

In Table 1 I list the top 25 countries according to the number of international migrants residing in that country. The data are from the World Bank's Global Bilateral Migration Database for the year 2010.[6] The United States stands out as the country with the highest number of immigrants, at over 40 million people, representing 20.7% of the total of all migrants in the world. It is noteworthy that there are three countries in the top ten of migration destinations that are not OECD destinations: Russia, Saudi Arabia, and India. In the whole table this number swells to 14 countries out of the 25 top migration destinations.

This raises the issue of the importance of including non-OECD destinations in the estimation of the empirical gravity model. The estimations by Hanson and McIntosh (2016) exclude all non-OECD countries as potential destinations because their dataset (DIOC) is constructed only from censuses of OECD countries. The omission of large destination countries may affect the overall projections, even for countries that are OECD members by biasing the coefficients used in the projections.

The high number of migrants in countries that were formally part of the USSR is noteworthy. Among those countries, Russia, Ukraine, and Kazakhstan exhibit large numbers of immigrants and all three of them show up in the top 25 destinations. The large current migration counts can be traced back to people that were forcibly resettled when the USSR was in existence, people who returned to their ancestral homeland after the fall of the USSR (but who are counted as migrants because of their place of birth), and economic migration due to the booming economies in Russia and Kazakhstan after the collapse of the USSR.

In terms of origin, most countries in the top 25 (Table 2) are developing or emerging countries. Mexico is the largest source of international migrants, followed closely by India and Russia. Migrants of Mexican origin number 11.9 million, representing 10% of Mexico's resident population, and Mexican migrants make up 6.1% of all migrants in the dataset. Virtually all Mexican emigrants reside in the USA.

Table 3 shows the top 25 pairs of source and destination countries. The largest bilateral migration flow is that of Mexicans residing in the USA. At 11.6 million people, the Mexican diaspora in the USA amounts to 6% of the total number of migrants in the world. Table 3 also reveals several bidirectional migration relationships. For example, those between Russia and Ukraine (in positions 2 and 3), Russia and Kazakhstan (positions 6 and 7), and Bangladesh and India (positions 4 and 24). These bilateral relationships are influenced by historical and cultural ties, as well as by geographical proximity, underscoring the appropriateness of a gravity-style model to explain the bilateral migration patterns in the data.

[6] The criterion for being classified as a migrant in that database is according to country of birth, not nationality.

Table 1: *Top 25 destination countries in 2010*

	Destination country	Migrants (thousands)	% of Destination population	% of all Migrants in data
1	United States of America	40,154	13.0%	20.7%
2	***Russian Federation***	11,793	8.2%	6.1%
3	Germany	9,784	12.2%	5.1%
4	***Saudi Arabia***	7,289	25.9%	3.8%
5	Canada	7,027	20.6%	3.6%
6	Spain	6,892	14.8%	3.6%
7	United Kingdom	6,788	10.8%	3.5%
8	France	6,595	10.5%	3.4%
9	Australia	5,441	24.6%	2.8%
10	***India***	5,337	0.4%	2.8%
11	***Ukraine***	4,932	10.8%	2.5%
12	Italy	4,379	7.3%	2.3%
13	***United Arab Emirates***	3,293	39.5%	1.7%
14	***Kazakhstan***	3,071	18.8%	1.6%
15	***Jordan***	2,957	45.4%	1.5%
16	Israel	2,724	36.7%	1.4%
17	***Côte d'Ivoire***	2,366	11.8%	1.2%
18	***Malaysia***	2,301	8.2%	1.2%
19	***Hong Kong***	2,296	32.8%	1.2%
20	Japan	2,176	1.7%	1.1%
21	***Iran***	2,108	2.8%	1.1%
22	***Singapore***	1,879	37.0%	1.0%
23	***South Africa***	1,863	3.6%	1.0%
24	Netherlands	1,599	9.6%	0.8%
25	***Kuwait***	1,572	51.4%	0.8%

Source: own calculations from the Bilateral Migration Matrix (World Bank) for 2010.
Notes: countries that are not members of the OECD are marked in bold.

Table 2: *Top 25 origin countries in 2010*

	Origin country	Migrants (thousands)	% of Origin population	% of all Migrants in data
1	Mexico	11,852	10.0%	6.1%
2	India	11,053	0.9%	5.7%
3	Russian Federation	10,090	7.0%	5.2%
4	China	8,218	0.6%	4.2%
5	Ukraine	6,002	13.1%	3.1%
6	Bangladesh	4,961	3.3%	2.6%
7	United Kingdom	4,548	7.3%	2.3%
8	Pakistan	4,522	2.7%	2.3%
9	Turkey	4,239	5.9%	2.2%
10	Philippines	4,213	4.5%	2.2%
11	Egypt	3,565	4.3%	1.8%
12	Germany	3,402	4.2%	1.8%
13	Italy	3,359	5.6%	1.7%
14	Kazakhstan	3,340	20.5%	1.7%
15	Poland	3,067	8.0%	1.6%
16	Morocco	3,014	9.4%	1.6%
17	Romania	2,765	13.6%	1.4%
18	State of Palestine	2,707	66.5%	1.4%
19	Indonesia	2,321	1.0%	1.2%
20	United States of America	2,272	0.7%	1.2%
21	Viet Nam	2,201	2.5%	1.1%
22	Portugal	2,139	20.2%	1.1%
23	Afghanistan	2,135	7.6%	1.1%
24	Republic of Korea	2,033	4.1%	1.1%
25	Colombia	2,018	4.4%	1.0%

Source: own calculations from the Bilateral Migration Matrix (World Bank) for 2010.

Table 3: *Top 25 cross-country migrant counts in 2010*

	Destination country	Origin country	Migrants (thousands)	% of all Migrants in data
1	United States of America	Mexico	11,636	6.0%
2	Ukraine	Russian Federation	3,684	1.9%
3	Russian Federation	Ukraine	3,647	1.9%
4	India	Bangladesh	3,299	1.7%
5	Germany	Turkey	2,733	1.4%
6	Russian Federation	Kazakhstan	2,648	1.4%
7	Kazakhstan	Russian Federation	2,227	1.2%
8	Hong Kong	China	2,225	1.1%
9	United Arab Emirates	India	2,186	1.1%
10	United States of America	China	1,736	0.9%
11	United States of America	Philippines	1,718	0.9%
12	Iran	Afghanistan	1,704	0.9%
13	United States of America	India	1,654	0.9%
14	United States of America	Puerto Rico	1,651	0.9%
15	Syrian Arab Republic	State of Palestine	1,541	0.8%
16	Saudi Arabia	India	1,453	0.8%
17	Malaysia	Indonesia	1,398	0.7%
18	Côte d'Ivoire	Burkina Faso	1,311	0.7%
19	Australia	United Kingdom	1,208	0.6%
20	United States of America	Viet Nam	1,160	0.6%
21	India	Pakistan	1,151	0.6%
22	United States of America	El Salvador	1,116	0.6%
23	Singapore	Malaysia	1,061	0.5%
24	Bangladesh	India	1,053	0.5%
25	United States of America	Republic of Korea	1,051	0.5%

Source: own calculations from the Bilateral Migration Matrix (World Bank) for 2010.

Table 3 also shows the preeminence of the USA as the main recipient of migrants from many developing around the world. The USA shows up eight times as a destination in the top 25 origin-destination pairs and is in many cases the top destination for emigrants from particular countries. Behind Mexico (position 1 in the ranking), migrants in the USA originate predominantly from China (position 10), Philippines (position 11), India (position 13), and Puerto Rico (position 14). The table also shows that India is important as the provider of large numbers of international migrants to several countries, as it shows up four times in the top 25 list.

4 Results

The gravity equation provides a good fit for changes in bilateral migration. Table 4 contains the outcome of trying out various specifications and exhibits a robust relationship between migration and the distance variables and their interaction with the labor ratio change $\left(\ln \frac{L^0_{s,t-1}}{L^0_{s,t-5}} - \ln \frac{L^0_{d,t-1}}{L^0_{d,t-5}}\right)$ that captures relative population pressures. This Table is the equivalent to Table 2 of Hanson and McIntosh (2016) but estimated for the specification in (2) on migration data from the GBM database. Column 1 in Table 4 shows the estimates of the coefficients in the gravity equation in a specification without any interaction terms. As expected, greater distance reduces migration and a common border and a common language enhance migration flows. A common colonial relationship does not by itself lead to higher bilateral migration. The main variable of interest, the labor ratio change, although not significantly different from zero, has the expected sign. This contrasts with the results by Hanson and McIntosh (2016), who found a counter-intuitive significant negative effect of labor abundance on migration.

Columns 2 through 5 show specifications that add interactions between the change in the labor ratio with the different distance variables. The results indicate that if two countries share a common language or a former colonial relationship, then a change in the relative supply of labor in the source country translates into more bilateral migration to the destination country. The point estimates for geographical distance and a common border also indicate an enhancement of the effect of changes in the labor share in the expected direction, although they are imprecisely estimated. Columns 6 through 8 contain the results of including the indicators for the presence of past migration. Migration networks enhance the effect of changes in the labor force. The effect is stronger for stricter cutoffs in the prior migration indicator. So, the effect of changes in the labor ratio doubles when a source-destination country pair is in the top 20% relative to the top 50% of the ranking of 1960 migration networks. For the top 5%, the point estimator is three times as large as for the top 20%. F-tests at the bottom of the table show that in the specifications with interactions in Columns 4 through 8 the effect of the labor ratio change is significantly positive at the 10% level (and in three cases also at the 5% level).

By including all covariates in a single regression, I estimate the specification that I use to project bilateral migration changes into the future.[7] Armed with these estimates, I first compute the number of immigrants residing in the seven OECD countries studied by Hanson and McIntosh (2016) and plot them in Figure 1. According to these projections, the USA and Spain experience the largest increases in their immigrant counts, followed by Italy, the UK, and France. Germany is projected to experience a decline in its immigrant stock and the projection for Japan is mostly flat. In the case of Spain, the inclusion of the years prior to 2000

[7] The estimated coefficients used for the projection are shown in Table 5, Column (3), in Appendix A.

Table 4: *Interaction Effects of Labor Ratio Growth on Migration Flows*

VARIABLES	(1) Without Interactions	(2) Geographical Interactions	(3) Geographical Interactions	(4) Geographical Interactions	(5) Geographical Interactions	(6) Past migration Interactions	(7) Past migration Interactions	(8) Past migration Interactions
Labor Ratio Chg.	0.012	0.073	0.011	0.009	0.002	0.011	0.009	0.012
	(0.013)	(0.057)	(0.013)	(0.013)	(0.013)	(0.014)	(0.013)	(0.014)
Per-capita GDP ratio	-0.001	0.000	-0.000	-0.001	-0.002	0.000	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Per-capita GDP ratio squared	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log distance	-0.008***	-0.008***	-0.007**	-0.008***	-0.009***	-0.004**	-0.005**	-0.005**
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Common border	0.029*	0.030*	0.043	0.030*	0.028*	0.022*	0.025*	0.024*
	(0.015)	(0.016)	(0.031)	(0.016)	(0.015)	(0.013)	(0.014)	(0.014)
Colonial relationship	-0.003	-0.003	-0.002	0.001	-0.003	-0.006	-0.009	-0.009
	(0.018)	(0.018)	(0.018)	(0.020)	(0.018)	(0.019)	(0.018)	(0.018)
Common language	0.018***	0.018***	0.017***	0.017***	0.017***	0.017***	0.017***	0.018***
	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
Top 5%						0.028	0.019	0.016
						(0.031)	(0.027)	(0.026)
Top 20%						0.014***	0.012**	0.013**
						(0.005)	(0.006)	(0.006)
Top 50%						0.001	-0.000	0.000
						(0.004)	(0.004)	(0.004)
Labor Ratio Chg. × Log distance		-0.007						
		(0.006)						
Labor Ratio Chg. × Common border			0.072					
			(0.086)					
Labor Ratio Chg. × Colonial relationship				0.038**				
				(0.019)				
Labor Ratio Chg. × Common language					0.026***			
					(0.008)			
Labor Ratio Chg. × Top 5%						0.139*		
						(0.072)		
Labor Ratio Chg. × Top 20%							0.034**	
							(0.013)	
Labor Ratio Chg. × Top 50%								0.016***
								(0.005)
Observations	15,122	15,122	15,122	15,122	15,122	15,122	15,122	15,122
R-squared	0.060	0.061	0.062	0.062	0.062	0.071	0.064	0.063
F test: Labor Ratio Chg + Interaction = 0		1.652	0.876	4.343	3.626	4.063	4.956	3.482
F test p-value		0.199	0.349	0.0372	0.0569	0.0439	0.0260	0.0621

Notes: The dependent variable is the bilateral migration flow between two countries. All regressions include origin and destination country dummies (not shown) and are estimated by least squares weighted by population in the source country. Robust standard errors in parentheses. *** represents 1% significance, ** 5% significance, and * 10% significance.

in the figure reveals that the growth in the migrant stock was unusually high between 2000 and 2010, in part because of the booming economy in the years prior to the Great Recession. Because the projections rely on the growth rate of migrants between only these two years, there is reason to suspect that the model is overestimating the growth of migrants in Spain. For the other countries, change between 2000 and 2010 does not fall out of the ordinary when compared to the prior decades.

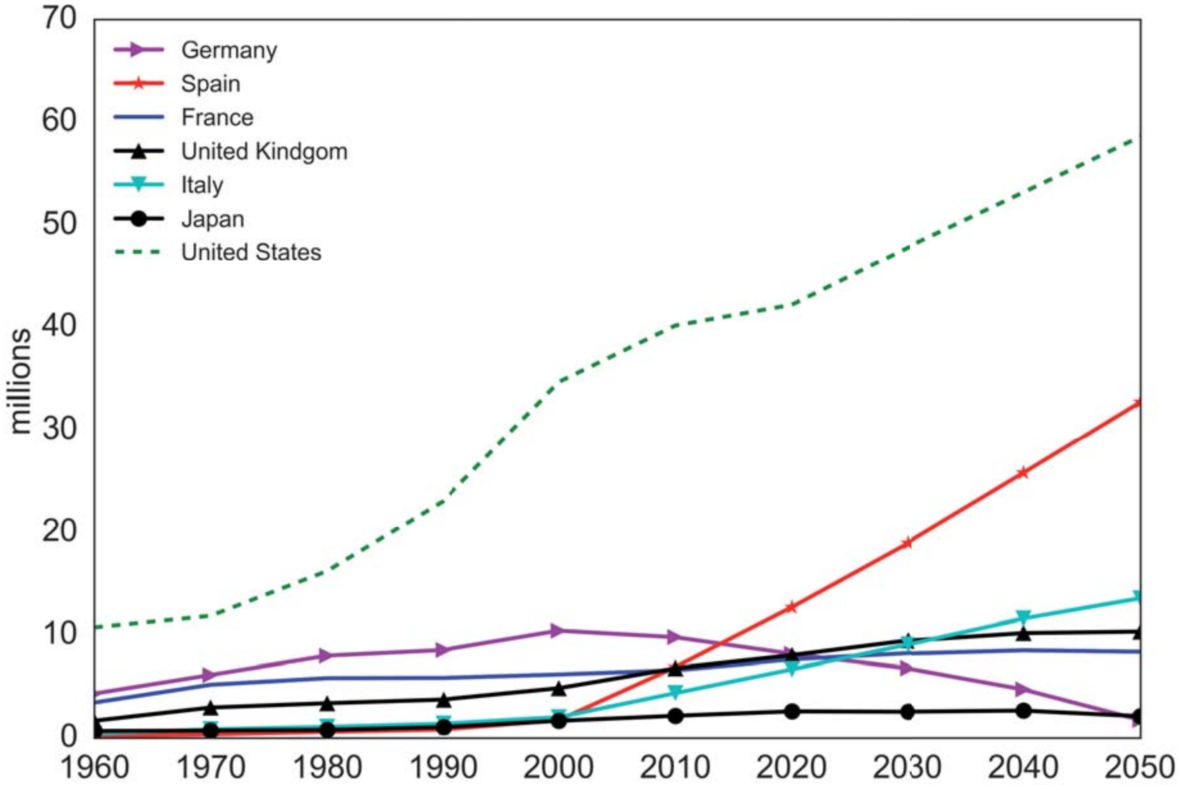
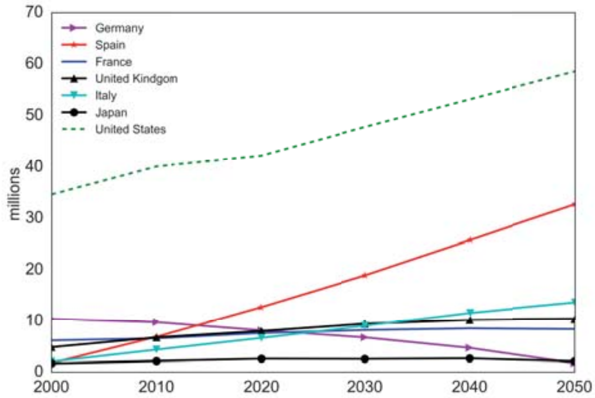


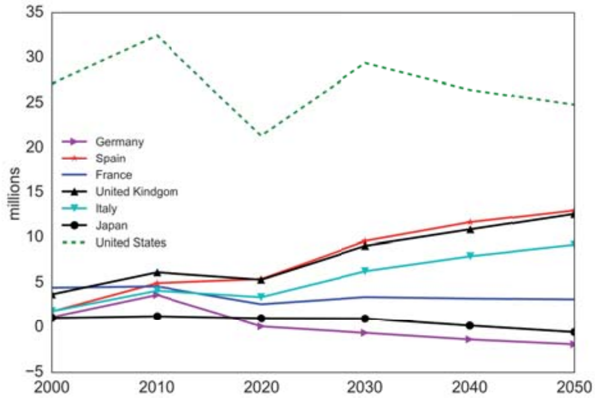
Figure 1: *Past number of immigrants and projections for selected OECD destinations. Data until 2010 are from the GBM database. For later years, the figure plots the projected number of immigrants obtained from the projection equation in* (14) *aggregated by destination country. Projections are obtained for the whole sample but only selected OECD destinations are shown.*

4.1 Comparison with the projections by Hanson and McIntosh (2016)

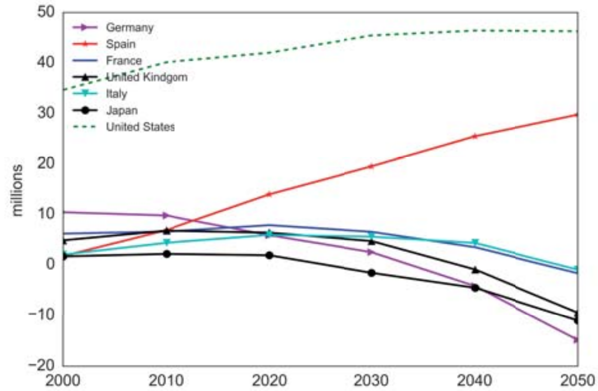
In Figure 2 I compare my projections with those from Hanson and McIntosh (2016), who use only the OECD destination countries available in the OECD's DIOC database. Their definition of the population refers to the population aged 15–54 because their cohort-based methodology allows them to make projections only for this population group. This implies that the scales are not directly comparable. A major qualitative difference that is immediately apparent from the comparison of Figures 2a and 2b is that for the USA the projection of immigrants keeps rising



(a) *GBM Database (World Bank)*



(b) *DIOC Database (OECD)*



(c) *GBM Database restricted to countries in DIOC*

Figure 2: *Past number of immigrants and projections for selected OECD destinations: comparison with the results by Hanson and McIntosh (2016), who use the DIOC database. Data for subfigure (b) are taken directly from Hanson and McIntosh (2016). In subfigures (a) and (c), data until 2010 are from the GBM database. For later years, the figure plots the projected number of immigrants obtained from the projection equation in* (14) *aggregated by destination country. Projections are obtained for the whole sample but only selected OECD destinations are shown. Subfigures (a) and (c) plot the total number of migrants and subfigure (b) only those aged 15–54.*

over the next decades when the GBM Database is used and stagnates with the estimation that uses only the limited set of destination countries included in the DIOC database. The levels of the projections for the other countries are ranked roughly similarly in both figures although they differ in the evolution. Among these differences, Spain exhibits a stronger growth rate in my estimation using the GBM database and the stocks of projected immigrants for the rest of the countries do not fan out as much as time progresses, as they do in the projections by Hanson and McIntosh (2016). The projections for Germany and Japan, although they decrease or stay constant in both figures, do not turn negative in my estimation using GBM data.

There are several factors that might influence the differences in the projections. The first of these factors is in the empirical approach: I rely on variation across country-pairs to identify the effects of the explanatory variables whereas Hanson and McIntosh (2016) rely on variation across country-pairs and cohorts. Second, and related to this, the population for whom Hanson and McIntosh (2016) produce projections is only a subset of the total population. The third factor is in the coverage of countries in the bilateral migration datasets: the DIOC database covers only OECD destinations and the GBM database all countries. Finally, the source data used by the DIOC database do not always coincide with those of the GBM database. It is not possible to give a conclusive answer to how each of these factors affects the projections, and it is also not the purpose of this paper.

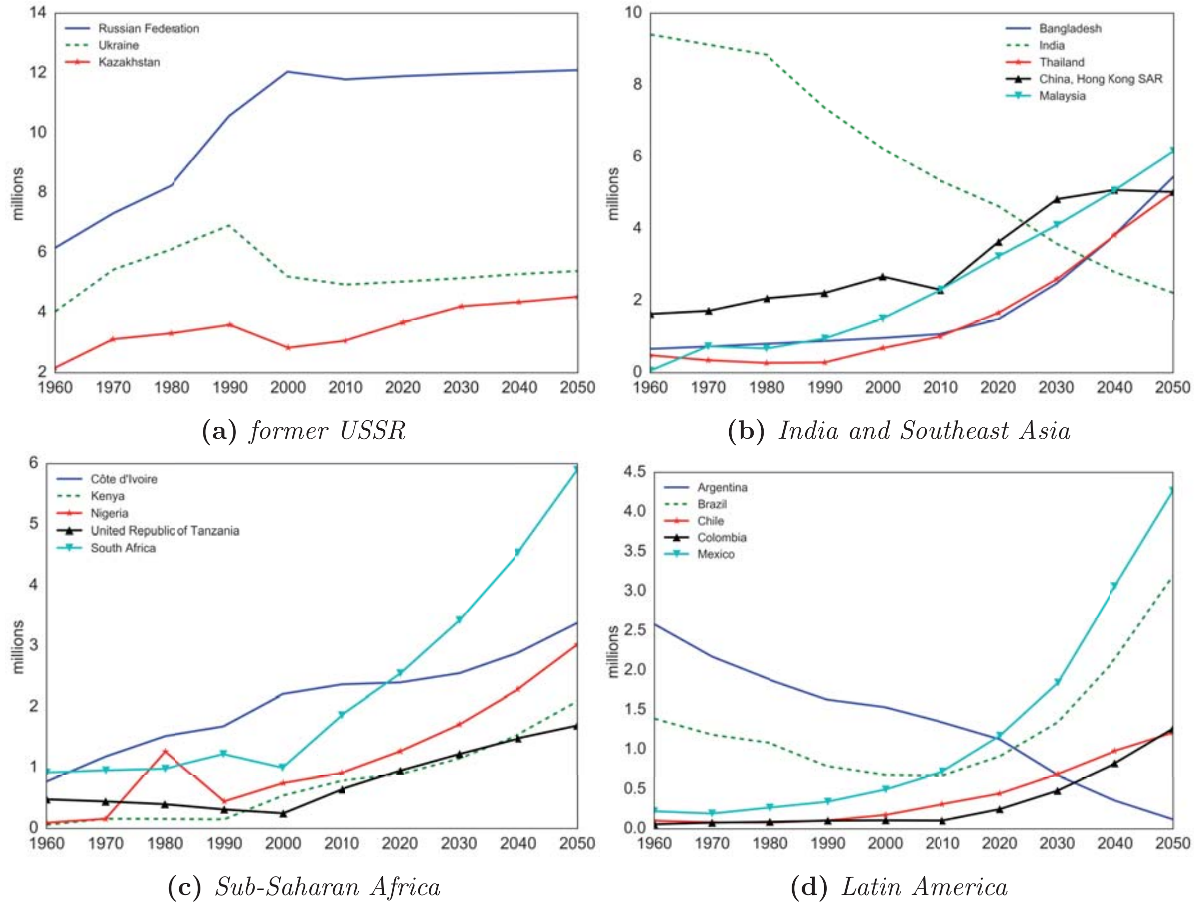
One aspect that can be studied with some detail is the coverage of different destination countries. In Figure 2c I have repeated the projection exercise by estimating the main specification using data from the GBM database but only for destination countries that are available in the DIOC database. With this restriction, immigrants in the USA approach a peak of 46 million in 2030 and stay around that level thereafter. This resembles the evolution of the stock of immigrants in the results by Hanson and McIntosh (2016). For all countries shown, the projected stock of migrants is systematically lower. In fact, the number of immigrants turns negative in some of them, as it does in the results by Hanson and McIntosh (2016): in Japan the number of immigrants turns negative by 2030 and in Germany and the UK by 2040. This suggests that the expansion of the dataset to non-OECD countries produces higher estimates in developed countries. A potential explanation for this result is that the estimation for the reduced set of countries available in DIOC places stronger weight on the change in the labor ratio and its interactions with the geographical and cultural distance parameters.[8] Because birth rates of source countries tend to converge to those of developed countries in the United Nation's projections, the DIOC sample therefore predicts lower flows of migrants in future decades.

[8]These coefficients are shown in Column 4 of Table 5 (Appendix A). Coefficients involving the labor share are uniformly larger (in absolute value) in Column 4 than the coefficients in Column 3 that were estimated using GBM data, with the exception of the top 20% indicator.

4.2 Projections for non-OECD destination countries and total world migrants

Because the GBM sample includes virtually all countries in the world, it is possible to construct projections for non-OECD destinations with important immigrant stocks and also for the whole world.

In Figure 3 I plot the projected stocks of immigrants in several non-OECD countries. Figure 3a depicts the projected number of immigrants in the three former USSR countries that appear as major destinations in Table 1. The stock of immigrants in Russia and Ukraine is projected to remain fairly constant at around 12 million in Russia and 5 million in Ukraine. On the other hand, the stock of immigrants in Kazakhstan is projected to increase by 50% from 3 million in 2010 to 4.5 million by 2050. This is explained by growth in per-capita GDP over the period 1990-2010 as well as low projected birth rates in Kazakhstan.



(a) *former USSR*

(b) *India and Southeast Asia*

(c) *Sub-Saharan Africa*

(d) *Latin America*

Figure 3: *Past number of immigrants and projections for selected non-OECD destinations. Data until 2010 are from the GBM database. For later years, the figure plots the projected number of immigrants obtained from the projection equation in* (14) *aggregated by destination country.*

In Figure 3b I plot the number of projected immigrants in selected Asian countries. Immigrants are projected to increase in all countries except India. In fact, the large projected population increase coupled with an important immigrant presence base in several destination countries, make India one of the main sources of emigrants in the future. Figure 3c shows projected immigrant stocks in Sub-Saharan Africa. South Africa is projected to experience the largest surge of immigrants, although this may be influenced by the particularly large growth of immigrants between 2000 and 2010, the years that are used to fit the estimation. The other countries exhibit more moderate but positive growth of immigrants, many of whom are attracted from other countries in the region. Finally, Figure 3d shows projected immigrant stocks for Latin America. The stock of immigrants in Argentina is projected to fall in future decades, falling almost to zero. The lower birth rates in Argentina's traditional sources if immigrants and its lagging per-capita GDP make this country less likely to receive new immigrant flows. Brazil is projected to turn around the decreasing trend prior to 2010. This is, in part, explained by its lower projected birth rates. In Mexico, growth in its per-capita GDP and the transition to lower population growth make this country an important destination for international migration.

Because all countries are included in the estimation and projection step, it is also possible to sum migrants across destinations to obtain the total number of migrants in the world. In Figure 4 I plot the projection of migrants in millions of people and as a percentage of projected world population. The total stock of migrants is projected to grow from 190 million in 2010 to 334 million in 2050. In terms of the projected world population, the stock of international migrants stood at 2.8% in 2010 and is projected to rise over the next 40 years, to 2.9% in 2020, 3.1% in 2030, 3.3% in 2040, and 3.5% in 2050.

The forecasts can also be cumulated according to region of origin. In Figure 5 I show the projections of emigrants for major regions were migrants originate from. In the panel on the left I plot absolute numbers and in the panel on the right emigrants as a percentage of the population residing in a particular region. From Figure 5b it is apparent that none of the three regions in which emigrants represent close to 10% of the resident population are projected to increase this share significantly. In the case of Mexico and the former USSR countries, the share of emigrants increases only slightly over the years. For Eastern Europe (excluding former USSR countries), the share drops to below 8% by 2050. Sub-Saharan Africa and India are projected to increase the share of emigrants, although from low initial numbers. These two origins, because of their strong projected population growth, feature large increases in the number of migrants. So, emigrants from countries in Sub-Saharan Africa are projected to grow from 19 million in 2010 to 78 million in 2050. Part of them migrate to other countries in Sub-Saharan Africa, such as South Africa. Emigrants from India are projected to increase from 11 million in 2010 to over 40 million by 2050.

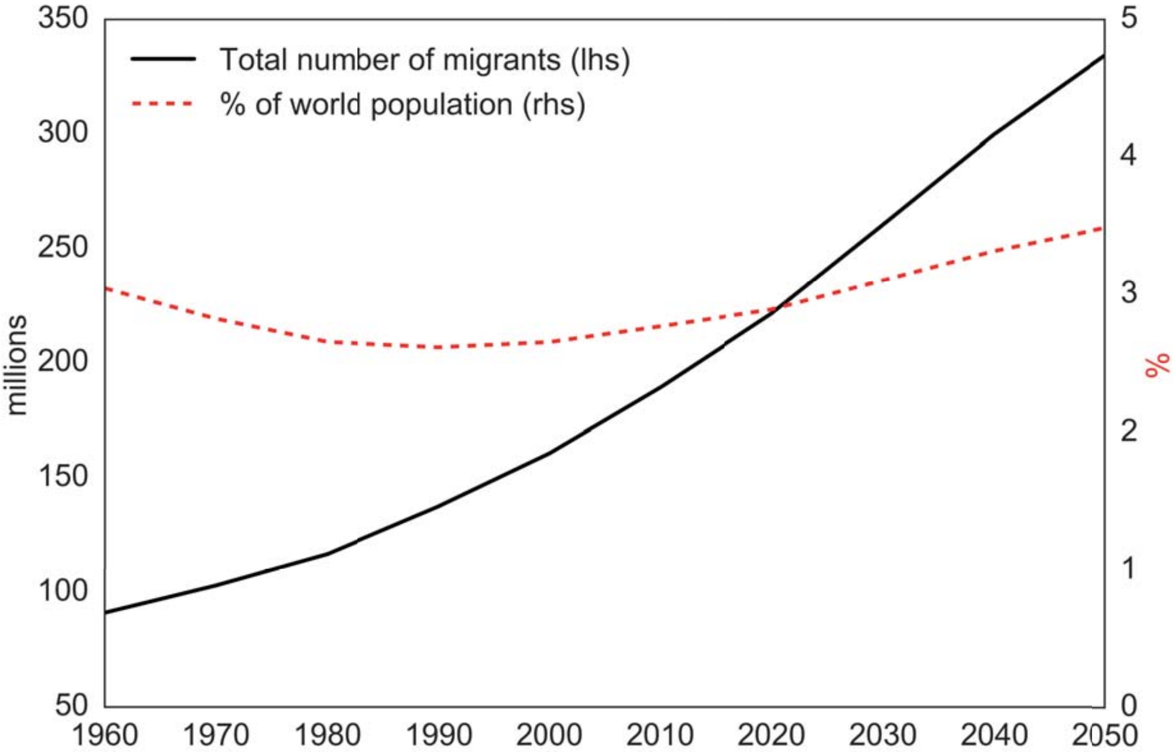
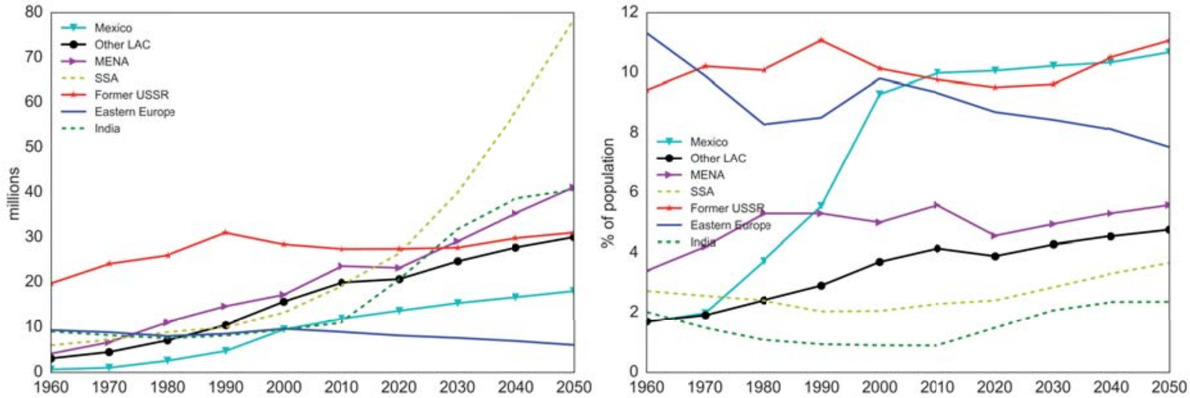


Figure 4: *Number of migrants and their percentage of total world population. Data until 2010 are from the GBM database. For later years, the figure plots the projected number of migrants (and their percentage of world population) obtained from the projection equation in* (14) *aggregated for the whole world.*



(a) *Millions of emigrants* **(b)** *In % of population*

Figure 5: *Past and projected total number and percentage of emigrants for major regions of origin. Data until 2010 are from the GBM database. For later years, the figure plots the projected number of emigrants obtained from the projection equation in* (14) *aggregated by major region of origin in (a) and their percentage of total regional population in (b).*

Emigrants from countries in the Middle East and North Africa (MENA) are projected to decrease until 2020, both as a percentage of population and in total numbers and to increase thereafter. Emigrants from Latin America and the Caribbean (LAC) excluding Mexico are also projected to increase starting in 2020. Although they represent a smaller fraction of the population than in Mexico, emigrants from the rest of Latin America exceed those of Mexico in total numbers. Their growth rate in absolute numbers is also bigger. Emigrants from Mexico are projected to grow from close to 12 million in 2010 to 18 million in 2050 whereas those from the rest of LAC are projected to grow by 50% from close to 20 million to 30 million by 2050.

5 Usage and intepretation of the projections

The projections for international migration are based on population pressures, that are relatively easy to forecast. However, trends in bilateral migration flows are affected not only by population pressures but also by immigration policies and economic and socio-demographic disparities between nations that are not captured by geographical and cultural distance variables (Grogger and Hanson, 2011; Belot and Hatton, 2012). This implies that the projections in this paper should not be interpreted as predictions or forecasts. Rather, they are the best possible projection using a restricted set of variables that capture population pressures and their interaction with geography, and users of these projections should be aware that changes in immigration policies may substantially modify the evolution of migration projected from population pressures.

The upside of including only the limited set of variables describing population pressures is that projections of international migration pressures are based on a minimal set of assumptions on the underlying determinants of migration. In the case of per-capita GDP, the variable is lagged by 40 years, so that projections until 2050 do not require to obtain forecasts on this variable. The only variable needed to project migration is the evolution of population growth, which is projected by the United Nations from fertility and mortality rates.

It could be the case that a user of this methodology wishes to include policy variables, such as immigration policies in receiving countries, in the projections. Provided that it is possible to observe changes in these policy variables between 2000 and 2010, the methodology used in this paper allows for the inclusion of these policy variables by adding them as an additional regressor. Ideally, because of the likely endogeneity of these policy variables, the specification should be estimated by instrumenting the changes in the policy variables with appropriately exogenous variables.

A further caveat of these projections, one that was mentioned in passing in Section 4, is that if a country-pair experienced an atypical high or low growth in the number of migrants between

2000 and 2010, then this may produce persistent high or low flows in the projections. If, as occurred to Spain, the number of immigrants from virtually all origins experiences atypical high positive inflows between 2000 and 2010, then the estimation will place a high value on the destination fixed-effect η_d in the case of Spain. This will in turn imply high projected inflows for Spain in all future years. A useful way to detect such a case is to plot prior years, as I did for the case of Spain, in order to spot sudden changes in the 2000-2010 flow that may be indicative of an atypical observation. Once such an atypical situation is detected there are three possible ways of proceeding. The first is to simply treat the estimation for that particular country with care, noting that the projection is likely to be biased by the atypical value, as I did for Spain and South Africa in Section 4. A second approach is to try to replace migrant flows for the affected country with external estimates of what the number of migrants would have been in absence of the atypical situation. Finally, a third possibility would be to identify affected countries by a dummy variable to attempt to identify the partial effect of the reason behind the atypical growth, such as a booming economy. In the projections, this dummy variable would then be set to zero for all future years. However, for this last approach to work, the effect of a booming economy would need to be comparable across the different countries. Rather than attempting any of these alternative solutions, I leave them for future work and simply report the likely problematic cases.

An additional extension that is left for future work is to try out different population growth scenarios and gauge their impact on projected migration. The United Nations' high and low fertility scenarios, which imply level effects for all countries in the world, could be used. Alternatively, it is also possible to increase or decrease birth cohort sizes used as inputs for a particular country. The methodology in this paper allows to trace how larger population growth in a particular country or region impacts international migration at the global level.

6 Conclusion

In this paper I extend the methodology by Hanson and McIntosh (2016) to construct projections for international migration for all countries in the world, including non-OECD destination countries. I show that non-OECD countries are important destinations for international migration, and this calls for a projection methodology that includes these countries. The specification I estimate is formally implied by that of Hanson and McIntosh (2016). I show that, although they use variation across cohorts to identify the coefficients in the empirical relationship, this variation is not required for the projection phase. Moreover, the absence of cohort-specific parameters in the projection step also implies that the estimation step can be performed without relying on variation across cohorts, and exploiting just variation across country pairs instead. The loss of the cohort dimension in the estimation step is countered by a larger number of destination countries, and therefore more country pair observations, leading to a similar number of observations overall.

One advantage of using a larger set of destination countries is that projections can be made for non-OECD countries and for the world as a whole. A second advantage is that estimates are refined by the inclusion of non-OECD countries that are systematically excluded as destination countries if the DIOC database is used. For example, I find that including these additional destination countries lowers the effect of the labor ratio on migration. Because of this impact on the estimates of coefficients, the inclusion of non-OECD countries also affects the projection for OECD countries. For the US, I find that immigration does not stagnate at its current level but keeps increasing throughout the projection horizon once non-OECD destination countries are taken into account in the estimation.

The projections obtained rely on a very limited set of variables that are relatively easy to project themselves (or are sufficiently lagged to remove the need for projecting them). This makes using the projection methodology in this paper very appealing, especially given the scarcity of data in developing countries. Potential uses of the methodology in this paper include constructing population scenarios and thus calculating long-term potential growth rates in any given country of the world. In doing this, the issues and caveats discussed in Section 5 should be kept in mind when using the projection methodology.

References

Beine, M., F. Docquier, and Ç. Özden (2011): "Diasporas," *Journal of Development Economics*, 95(1), 30–41.

Belot, M. V. K., and T. J. Hatton (2012): "Immigrant Selection in the OECD," *Scandinavian Journal of Economics*, 114(4), 1105–1128.

Docquier, F., G. Peri, and I. Ruyssen (2014): "The Cross-country Determinants of Potential and Actual Migration," *International Migration Review*, 48, S37–S99.

Dumont, J.-C., and G. Lemaître (2005): "Counting immigrants and expatriates in OECD countries," Discussion Paper 25, OECD Social, Employment and Migration Working Papers.

Feenstra, R. C., R. Inklaar, and M. P. Timmer (2015): "The Next Generation of the Penn World Table," *American Economic Review*, 105(10), 3150–3182.

Grogger, J., and G. H. Hanson (2011): "Income maximization and the selection and sorting of international migrants," *Journal of Development Economics*, 95(1), 42–57.

Hanson, G., and C. McIntosh (2016): "Is the Mediterranean the New Rio Grande? US and EU Immigration Pressures in the Long Run," *Journal of Economic Perspectives*, 30(4), 57–82.

Hatton, T. J., and J. G. Williamson (2003): "Demographic and Economic Pressure on Emigration out of Africa," *Scandinavian Journal of Economics*, 105(3), 465–486.

Head, K., T. Mayer, and J. Ries (2010): "The erosion of colonial trade linkages after independence," *Journal of International Economics*, 81(1), 1–14.

McKenzie, D., and H. Rapoport (2007): "Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico," *Journal of Development Economics*, 84(1), 1–24.

OECD (2008): *A profile of immigrant populations in the 21st century: data from OECD countries*. OECD.

Özden, Ç., C. R. Parsons, M. Schiff, and T. L. Walmsley (2011): "Where on Earth is Everybody? The Evolution of Global Bilateral Migration 1960–2000," *The World Bank Economic Review*, 25(1), 12–56.

United Nations (2015): "World Population Prospects: The 2015 Revision, Methodology of the United Nations Population Estimates and Projections," Working Paper ESA/P/WP.242, United Nations, Department of Economic and Social Affairs, Population Division.

Appendix A: Additional tables

Table 5: *Regressions used to predict migration flows*

VARIABLES	(1) GBM	(2) DIOC	(3) GBM	(4) DIOC
Labor Ratio Chg.	0.056***	0.413***	0.015	0.323**
	(0.017)	(0.140)	(0.041)	(0.133)
Per-capita GDP ratio	-0.001	0.002	-0.001	0.003
	(0.002)	(0.011)	(0.002)	(0.008)
Per-capita GDP ratio squared	0.000	0.001	0.000	0.001
	(0.000)	(0.003)	(0.000)	(0.002)
Log distance	-0.008***	-0.022***	-0.004*	-0.007
	(0.002)	(0.008)	(0.002)	(0.008)
Common border	0.029*	0.016	0.029	0.045
	(0.015)	(0.082)	(0.028)	(0.085)
Colonial relationship	-0.003	-0.005	-0.006	-0.037
	(0.018)	(0.042)	(0.021)	(0.043)
Common language	0.018***	0.058***	0.016***	0.043***
	(0.004)	(0.016)	(0.004)	(0.016)
Top 5%			0.028	-0.011
			(0.029)	(0.044)
Top 20%			0.012*	0.026***
			(0.006)	(0.010)
Top 50%			0.001	-0.003
			(0.004)	(0.007)
Labor Ratio Chg. × Log distance			0.001	-0.021**
			(0.004)	(0.010)
Labor Ratio Chg. × Common border			0.035	0.096
			(0.076)	(0.321)
Labor Ratio Chg. × Colonial relationship			-0.004	0.001
			(0.025)	(0.044)
Labor Ratio Chg. × Common language			0.019**	0.037
			(0.008)	(0.025)
Labor Ratio Chg. × Top 5%			0.123*	0.184
			(0.068)	(0.117)
Labor Ratio Chg. × Top 20%			0.007	-0.011
			(0.009)	(0.016)
Labor Ratio Chg. × Top 50%			0.003	0.007
			(0.005)	(0.006)
Observations	15,122	4,384	15,122	4,384
R-squared	0.064	0.079	0.076	0.095

Notes: Column (3) exhibits the coefficients of the specification used to predict bilateral migration flows. Column (4) exhibits the coefficients of the regression that restricts destination countries to those available in the DIOC database. Columns (1) and (2) exhibit regressions without interactions for comparison. All regressions include origin and destination dummy variables (not shown) and are estimated by least squares weighted by population in the source country. Robust standard errors in parentheses. *** represents 1% significance, ** 5% siginficance, and * 10% significance.

BANCO DE ESPAÑA PUBLICATIONS

WORKING PAPERS

1621 ADRIAN VAN RIXTEL, LUNA ROMO GONZÁLEZ and JING YANG: The determinants of long-term debt issuance by European banks: evidence of two crises.

1622 JAVIER ANDRÉS, ÓSCAR ARCE and CARLOS THOMAS: When fiscal consolidation meets private deleveraging.

1623 CARLOS SANZ: The effect of electoral systems on voter turnout: evidence from a natural experiment.

1624 GALO NUÑO and CARLOS THOMAS: Optimal monetary policy with heterogeneous agents.

1625 MARÍA DOLORES GADEA, ANA GÓMEZ-LOSCOS and ANTONIO MONTAÑÉS: Oil price and economic growth: a long story?

1626 PAUL DE GRAUWE and EDDIE GERBA: Stock market cycles and supply side dynamics: two worlds, one vision?

1627 RICARDO GIMENO and EVA ORTEGA: The evolution of inflation expectations in euro area markets.

1628 SUSANA PÁRRAGA RODRÍGUEZ: The dynamic effect of public expenditure shocks in the United States.

1629 SUSANA PÁRRAGA RODRÍGUEZ: The aggregate effects of government incometransfer shocks - EU evidence.

1630 JUAN S. MORA-SANGUINETTI, MARTA MARTÍNEZ-MATUTE and MIGUEL GARCÍA-POSADA: Credit, crisis and contract enforcement: evidence from the Spanish loan market.

1631 PABLO BURRIEL and ALESSANDRO GALESI: Uncovering the heterogeneous effects of ECB unconventional monetary policies across euro area countries.

1632 MAR DELGADO TÉLLEZ, VÍCTOR D. LLEDÓ and JAVIER J. PÉREZ: On the determinants of fiscal non-compliance: an empirical analysis of Spain's regions.

1633 OMAR RACHEDI: Portfolio rebalancing and asset pricing with heterogeneous inattention.

1634 JUAN DE LUCIO, RAÚL MÍNGUEZ, ASIER MINONDO and FRANCISCO REQUENA: The variation of export prices across and within firms.

1635 JUAN FRANCISCO JIMENO, AITOR LACUESTA, MARTA MARTÍNEZ-MATUTE and ERNESTO VILLANUEVA: Education, labour market experience and cognitive skills: evidence from PIAAC.

1701 JAVIER ANDRÉS, JAVIER J. PÉREZ and JUAN A. ROJAS: Implicit public debt thresholds: an empirical exercise for the case of Spain.

1702 LUIS J. ÁLVAREZ: Business cycle estimation with high-pass and band-pass local polynomial regression.

1703 ENRIQUE MORAL-BENITO, PAUL ALLISON and RICHARD WILLIAMS: Dynamic panel data modelling using maximum likelihood: an alternative to Arellano-Bond.

1704 MIKEL BEDAYO: Creating associations as a substitute for direct bank credit. Evidence from Belgium.

1705 MARÍA DOLORES GADEA-RIVAS, ANA GÓMEZ-LOSCOS and DANILO LEIVA-LEON: The evolution of regional economic interlinkages in Europe.

1706 ESTEBAN GARCÍA-MIRALLES: The crucial role of social welfare criteria for optimal inheritance taxation.

1707 MÓNICA CORREA-LÓPEZ and RAFAEL DOMÉNECH: Service regulations, input prices and export volumes: evidence from a panel of manufacturing firms.

1708 MARÍA DOLORES GADEA, ANA GÓMEZ-LOSCOS and GABRIEL PÉREZ-QUIRÓS: Dissecting US recoveries.

1709 CARLOS SANZ: Direct democracy and government size: evidence from Spain.

1710 HENRIQUE S. BASSO and JAMES COSTAIN: Fiscal delegation in a monetary union: instrument assignment and stabilization properties.

1711 IVÁN KATARYNIUK and JAIME MARTÍNEZ-MARTÍN: TFP growth and commodity prices in emerging economies.

1712 SEBASTIAN GECHERT, CHRISTOPH PAETZ and PALOMA VILLANUEVA: Top-down vs. bottom-up? Reconciling the effects of tax and transfer shocks on output.

1713 KNUT ARE AASTVEIT, FRANCESCO FURLANETTO and FRANCESCA LORIA: Has the Fed responded to house and stock prices? A time-varying analysis.

1714 FÁTIMA HERRANZ GONZÁLEZ and CARMEN MARTÍNEZ-CARRASCAL: The impact of firms' financial position on fixed investment and employment. An analysis for Spain.

1715 SERGIO MAYORDOMO, ANTONIO MORENO, STEVEN ONGENA and MARÍA RODRÍGUEZ-MORENO: "Keeping it personal" or "getting real"? On the drivers and effectiveness of personal versus real loan guarantees.

1716 FRANCESCO FURLANETTO and ØRJAN ROBSTAD: Immigration and the macroeconomy: some new empirical evidence.

1717 ALBERTO FUERTES: Exchange rate regime and external adjustment: an empirical investigation for the U.S.

1718 CRISTINA GUILLAMÓN, ENRIQUE MORAL-BENITO and SERGIO PUENTE: High growth firms in employment and productivity: dynamic interactions and the role of financial constraints.

1719 PAULO SOARES ESTEVES and ELVIRA PRADES: On domestic demand and export performance in the euro area countries: does export concentration matter?

1720 LUIS J. ÁLVAREZ and ANA GÓMEZ-LOSCOS: A menu on output gap estimation methods.

1721 PAULA GIL, FRANCISCO MARTÍ, JAVIER J. PÉREZ, ROBERTO RAMOS and RICHARD MORRIS: The output effects of tax changes: narrative evidence from Spain.

1722 RICARDO GIMENO and ALFREDO IBÁÑEZ: The eurozone (expected) inflation: an option's eyes view.

1723 MIGUEL ANTÓN, SERGIO MAYORDOMO and MARÍA RODRÍGUEZ-MORENO: Dealing with dealers: sovereign CDS comovements.

1724 JOSÉ MANUEL MONTERO: Pricing decisions under financial frictions: evidence from the WDN survey.

1725 MARIO ALLOZA: The impact of taxes on income mobility.

1726 DANILO LEIVA-LEON: Measuring business cycles intra-synchronization in US: a regime-switching interdependence framework.

1727 PIERRE GUÉRIN and DANILO LEIVA-LEON: Model averaging in Markov-Switching models: predicting national recessions with regional data.

1728 MÁXIMO CAMACHO and DANILO LEIVA-LEON: The propagation of industrial business cycles.

1729 JAMES COSTAIN: Costly decisions and sequential bargaining.

1730 MARIO ALLOZA: Is fiscal policy more effective in uncertain times or during recessions?

1731 PIERRE GUÉRIN and DANILO LEIVA-LEON: Monetary policy, stock market and sectoral comovement.

1732 HENRIK JENSEN, IVAN PETRELLA, SØREN HOVE RAVN and EMILIANO SANTORO: Leverage and deepening business cycle skewness.

1733 CÉSAR MARTÍN MACHUCA: External stress early warning indicators.

1734 RODOLFO G. CAMPOS: International migration pressures in the long run.

BANCO DE ESPAÑA
Eurosistema

Unidad de Servicios Auxiliares
Alcalá, 48 - 28014 Madrid
E-mail: publicaciones@bde.es
www.bde.es