This is a postprint version of the following published document:


DOI: 10.1080/0022250X.2017.1323207

© The Authors
Forecasting the Integration of Immigrants

Pierluigi Contucci ¹, Rickard Sandell², Seyedalireza Seyedi ¹ ³.

Abstract. This paper presents and evaluates empirically a theoretical and mathematical framework to forecast immigrant integration. By comparing forecasted integration estimates based on data collected up to specific periods in time, with observed integration quantities beyond the specified period, we show that: Our forecasts turn out to be prompt, i.e. readily available after a short period of time; accurate, i.e. with a small relative error; and finally robust, i.e. able to predict integration correctly for several years to come. The research reported here proves that the proposed model of integration and its forecast framework are simple and effective tools to reduce uncertainties about how integration phenomena emerge and how they are likely to develop in response to increased migration levels in the future.

Keywords. Forecast, Immigrant Integration, Assimilation, Intermarriage mixed partnership, labor force participation.

¹Dipartimento di Matematica, University of Bologna,
²Departamento de Ciencias Sociales, Universidad de Carlos III de Madrid,
³Corresponding Author: seyedalireza.seyedi2@unibo.it,
Introduction

The research reported here addresses the problem of forecasting immigrant integration. The integration of immigrants is and have long been a political priority in countries receiving immigration (European-Commission 2005; Penninx et al. 2008; Jacoby et al. 2013). For example, achieving a minimum of integration is considered a necessity, to avoid friction and conflict between immigrants and natives in the host society (Niessen and Huddleston 2010; Castles and Miller 2009; European-Commission 2011; IOM 2011; European-Commission 2014). Furthermore, and perhaps most importantly, as the effects of low birth rates and aging populations are becoming manifest, high levels of immigrant integration or assimilation is considered by some of the world’s leading economies key in building a competitive and sustainable economy for the future (European-Commission 2010; European-Council 2010; European-Commission 2010; Giovagnoli 2011; Canada-Government 2012).

However, the capacity to formulate effective integration policies hinges on the availability of scientific theories and works generating strong predictions of how the integration phenomenon is likely to unfold with the passing of time, and in the face of changing levels of immigration. While there is a rich demographic, sociological and economic literature on individual integration outcomes (see, for example, Van Tubergen (2006) for a good overview of this literature), the problem of forecasting integrations is largely ignored by past and contemporary research. Hence, by filling an important gap in the literature on immigrant integration, the research reported here constitutes a significant contribution to our knowledge about integration and integration phenomena.
The paper is organized as follows. First we introduce the integration concept we aim at forecasting. Thereafter we introduce the theoretical and mathematical model of the integration phenomena on which we base our forecasting method, followed by data description. Then we present the methodological procedure, produce the forecasts, and conduct the proper evaluation of these forecasts. Finally, we conclude with a brief discussion of the wider implications of the findings.

Immigrant Integration

What is immigrant integration and what is it that we set out to forecast? Instead of engaging in the complex task of defining and operationalize integration, we feel comfortable to adopt a pragmatic approach. For example, it appear as if there exist a high degree of agreement within and across demographic, sociological and economic disciplines that intermarriage/inter-partnerships and immigrant labor force participation are considered core measures of social and socio-economic immigrant integration respectively (see Alba and Nee 1997; Rajzman and Tienda 1999; Van Tuijbergen 2006 for a more extensive discussions on the relevance of these integration quantifiers). Therefore, in the research reported here we set out to develop a quantitative forecasting model capable of predicting inter-partnerships and labor market participation. More specifically, we have suitable data available for the following four quantifiers; 1) rate of intermarriage between native and foreign born, 2) rate of newborns with one native and one foreign parent, 3) rate of indefinite labor contracts given to immigrants, and 4) rate of temporary labor contracts given to immigrants.
(see further in the section Data Description below). The former are example of social integration quantifiers, while the latter are examples of so called socio-economic quantifiers.

**A simple model of a complex phenomenon**

Forecasting integration requires a reliable and efficient model of the integration phenomena. There is a wide theoretical consensus that both individual and contextual factors are important when explaining variation in immigrant integration outcomes (Van Tubergen and Maas 2007). For example, systematic differences in individual integration outcomes are often associated with age, sex, education, language proficiency, and other individual characteristics (Cortina et al. 2008; Hwang et al. 1997; Stevens and Swicegood 1987; Qian et al. 2012). Likewise, contextual characteristics, such as culture, ethnic-group-size, and local heterogeneity levels are influential when assessing intermarriage propensities and labor market participation across different groups (Qian et al. 2012; Blau et al. 1982, 1984; Portes and Zhou 1993).

However, and in difference with main stream research, our aim is not to explain or disentangle the mechanisms that explain variation in individual integration outcomes. Rather, we are interested in forecasting the trend, that is, the systemic evolution of integration quantifiers, in the presence of some noise regardless of its source/origin individual, contextual or other. Hence, our approach calls for a different type of model than what we typically see in contemporary research articles on integration and assimilation in the two dimensions we focus on.
In recent years, there have been considerable efforts to apply ideas and techniques from statistical physics to other areas of science such as economic, finance, social science, and biology. Explaining the social phase transition, modeling collective animal behavior — like natural flocks of birds, and predicting trends and crisis in the financial Markets are some well-known examples of this venture (see Levy 2005; Castellano et al. 2000; Contucci et al. 2008; Mantegna and Stanley 1999; Stanley 2008; Bouchaud and Potters 2004; Ballerini et al. 2008; Bialek et al. 2012).

This genre of models has recently been used to study migration phenomenon (Barra et al. 2014; Contucci and Giardina 2008; Barra and Contucci 2010). More specifically, and of particular relevance for the research reported here, Barra et al. (2014) proposed a theoretical model of an immigrant-native system in which they explain the systemic integration process for social and socio-economic integration quantifiers as a function of immigrant density ($\gamma$). This model, which is inspired by ideas and techniques from statistical physics — the branch of theoretical physics devised to explain thermodynamic laws as emerging average behavior for systems composed of a high number of microscopic interacting elements — showed that while labor participation rates grow proportionally to immigrant density, the rates of mixed partnership have a growth law proportional to the square root of the immigrant density.\footnote{It should be noted that the causal relationship between immigrant group size and integration has received ample attention in Blau et al. 1982 and Blau et al. 1984. Blau et al’s findings are not contradictory to Barra et al. (2014) work, but complementary. However, in difference with Blau et al, Barra et al model the interaction component explicitly.}

To solve for these differences Barra et al. (2014) advanced a theory able to describe all the observed data by means of a unified mathematical framework.

Briefly summarizing the approach; Barra et al. (2014) considered labor contracts,
marriages, and births coupling relations among humans. In a two-group system such as a society composed of immigrants and natives, there can be in-group or cross-group couplings. In other words the choice between, say, marrying or hiring an immigrant over a native is dichotomous. A natural candidate to describe the frequency of cross-group couplings in large populations is McFadden’s Discrete Choice theory (McFadden 2001). A crucial assumption in the discrete choice theory is mutual independence between the involved random variables (Gallo et al. 2009). Consequently, McFadden’s theory would predict linear growth in integration over immigrant density. And indeed, Barra et al. (2014) show that McFadden’s theory works very well when assessing the level of integration in the Spanish labor market (model fit in terms of $R^2$ was over 0.96). Hence, the decisions of contracting an immigrant are made in a mutually independent fashion, regardless of how other actors have decided in this matter.\(^5\)

However, and less surprisingly, the choice of partner is not well described by the classical discrete choice theory, with $R^2$ dropping below 0.85 for intermarriage and 0.79 for mixed births. Drawing on the Weberian notion of social action Barra et al. (2014) argued that this mismatch arises because inter-partnership decisions are not taken independently (Weber 1978). That is, the action of marrying an immigrant or having a child with an immigrant is a decision that is contingent of how others in the environment have acted or not acted in this context before the decision at

\(^5\)Note that in the language of statistical mechanics, the absence of interaction in labor market participation couplings does not mean that social networks are irrelevant. What the data analysis suggests is simply that the propensity of hiring immigrants over natives is unaffected by whether other employers have or have not hired immigrants over natives before. Social networks are still likely to be determinant for the individuals chances of landing the job (see Granovetter 1974 on the latter).
hand (see also, for example, Kalmijn (1998)’s discussion of social determinants in intermarriage). Theories that relax the assumption of independence and cater to this type of social action and interaction, introduced by Brock and Durlauf (2001) and thereafter further developed (Contucci et al. 2008; Contucci and Giardina 2008; Gallo et al. 2009), predict a square root behavior of the probability of cross-group couplings just as observed in the inter-partnership data.

The unifying solution proposed by Barra et al. (2014), is a generalization of the so called monomer-dimer model (Heilmann and Lieb 1970) with the addition of an imitative interacting social network component — with random topology in agreement with the small world-scenario (Watts and Strogatz 1998). Such a model reduces to the classical discrete choice theory with linear growth in situations when imitation is negligible, and to the square root behavior when imitation is prevailing. Barra et al. (2014) showed that when exposed to empirical data their proposed model provides an extraordinary fit across all four integration quantifiers.6

We argue that this nontrivial explanation of the integration process is a useful tool if the goal is to obtain estimates of the development of the general level of integration in a society and how it may change with the passage of time and as immigration levels change, and thus a strong model candidate if the goal is to forecast integration.

In what follows we will draw on this work and extend it by proposing a quantitative forecasting method based on Barra et al. (2014) assumptions. We will conduct a formal test of the theory’s forecast potential. Our objectives are twofold: (1) for a given integration context, we set out to forecast the growth law governing the

6See Barra et al. (2014) for a full account of the model, proofs, and empirical tests.
integration process. That is, we set out to determine which mechanism — social interaction or individual independence — that predominates in the integration process. And (2), once the growth law is determined, establish if and to what extent the methodological framework lends itself to forecast immigrant integration and generate, as a function of immigrant density, precise and robust estimates of integration levels across time.

Data Description

We replicate the data used by Barra et al. (2014). That is, we use immigration data from Spain on the time interval 1999 to 2010. Data on local immigrant densities are compiled as follows. We use the size of the immigrant population and the native population in each municipality as reported in the 2001 Census as our baseline. We then estimate the local immigrant densities for different points in time (quarterly) based on the information contained in the Statistics over residential variation in Spanish municipalities, the so called Estadistica de variaciones residenciales (EVR) as elaborated by Spain’s National Statistical Agency (INE). An exceptional feature

---

This period corresponds to the period in which Spain received most of its current immigrant population. For those unfamiliar with the Spanish immigration context, the following brief information may be useful. In 1999, Spain received fewer than 50,000 new documented and undocumented immigrants. Since then, annual immigration levels have increased dramatically, reaching a peak in 2006 and 2007, with inflows exceeding 800,000. Spain’s documented and undocumented foreign-born population has risen from little more than 1 million to over 6.5 millions in the analyzed period. Its share of the total population has risen from less than 3 % to over 13 % in the same period. Currently there are immigrants from almost all nations in Spain. However, some 20 immigrant origins account for approximately 80 % of Spain’s total immigrant population. Immigrants from Romania form the largest minority in Spain (767,000 at the end of 2008), followed by immigrants from Morocco (737,000 at the end of 2008) and Ecuador (479,000 at the end of 2008). Europe and South America together account for over 70 % of Spain’s total immigrant population.
of the Spanish data is that they also include so-called undocumented immigrants, that is, immigrants who lack a residence permit (see Sandell 2012).

Data on marriages and births are drawn from the local offices of Vital Records and Statistics across Spain (Registro Civil), and have been compounded by the INE. By our definition, intermarriage occurs when a Spanish-born (native) person marries a person born in a foreign country. Similarly, we consider all births with one native and one foreign born parent to be mixed births. Intermarriages and mixed births where the foreign-born spouse or parent is an undocumented migrant is included. We focus our analysis on birth and marriage events that occurred during the period 1999 to 2008. However, data on density, marriages and births are subject to data protection restrictions. An individual residence municipality is only disclosed if its population is larger than 10,000. For this reason, out of approximately 8,000 municipalities in Spain, our analysis focuses on only 735. Still, some 85% of Spanish immigrants reside in the included municipalities.

Data on labor contracts come from Spain’s Continuous Sample of Employment Histories (the so-called Muestra Continua de Vidas Laborales or MCVL). It is an administrative data set with longitudinal information for a 4% non-stratified random sample of the population who are affiliated with Spain’s Social Security. Sampling is conducted on a yearly basis. We use data from the waves 2005 to 2010. The inclusion of an individual in the sample is determined by a sequence in the individual’s social security number that does not vary across sample waves. This means that individuals are maintained across samples. New affiliates with a social security number matching the predetermined sequence are added in each new wave. The
data contain information on contractual conditions such as whether the individuals have a temporary or indefinite labor contract, as well as the contracts start and stop times. Residential data at the level of municipality and information about place of birth are also available. In contrast to the data on densities, marriages, and births, for these data the residence municipality is only disclosed if the population is larger than 40,000.

Operatively, we derive two datasets based on the information described above. One contains data on intermarriages and mixed births, and the other on labor market participation. Both datasets contain spatial and temporal information, such as the municipal code, quarter, year, and the immigration density in the municipalities across time. The data on labor contracts consist of 3,553 entries over the period 2005-2010. The data on marriages and newborns consist of 27,144 (municipality/time) entries spanning the period 1999-2008. In our study, we define the integration quantifiers as a function of immigrant density, \( \gamma = \frac{\text{number of immigrants}}{\text{total population}} \in [0, 1] \), in the form:

\[
J_p = \frac{\text{number of permanent contracts given to immigrants}}{\text{total number of permanent contracts}}
\]

\[
J_t = \frac{\text{number of temporary contracts given to immigrants}}{\text{total number of temporary contracts}}
\]

\[
M_m = \frac{\text{number of mixed marriages}}{\text{total number of marriages}}
\]

\[
B_m = \frac{\text{number of newborns with mixed parents}}{\text{total number of newborns}}
\]
Method and Results

To prepare the data for analysis, we use the following algorithm: we began by aggregating and cleansing the data in which the records should be checked for missing or incomplete data to avoid inaccuracy in our forecasting approach. After that, data on our quantifiers are organized into two datasets. One which contains information on our social integration quantifiers (see equation 1 and 2). The second set contains data on socio-economic quantifiers (see equation 3 and 4). Next, for each dataset, we perform a detailed test of the methodology’s forecast performance for increasing intervals of time: period 1 = (2000 to 2001), period 2 = (2000 to 2002), period 3 = (2000 to 2003), etc.

The objective is to evaluate the forecasting ability of Barra et al’s theoretical model in two areas; 1) uncovering the growth law determining the underlying mechanism for integration (social action or individual independent action) characterizing each of the two types of integration phenomena, and 2) assessing the level of integration in the system under study at different time intervals. More precisely, we set out to evaluate the quality of the subsequent forecasts in terms of promptness — how long is the waiting time before we can identify the underlying mechanism driving integration; accuracy — how well the prediction replicates the observed value; and finally robustness — the forecasting ability over the entire time span.

We start by extracting the growth law of each quantifier in terms of $\Gamma = \gamma(1 - \gamma)^8$ by merging all data entries into a new file containing their $\Gamma$ and the quantifier values, 

---

8The control parameter $\Gamma$ tunes the total number of possible cross-link couplings between immigrant and native populations (see Barra et al. 2014).
regardless of space and time coordinates, and ordering them by increasing values of \( \Gamma \).

Since our quantifier-values are in the shape of fraction, we can compute their averages by obtaining the ratio between the statistical average of nominators and the statistical average of denominators\(^9\). To be more effective in this task, the quantifiers’ averages associated with new files could be efficiently evaluated by grouping the data into bins over \( \Gamma \). In order to avoid noisy result, the bins here are computed with respect to detailed test of different width of bins. The results reveal that 5 to 15 bins optimize 581-3471 entries on the socio-economic quantifiers data-groups in defined successive intervals, whereas 8 to 30 bins optimize 2421-26546 entries for the datasets on social quantifiers in determined consecutive periods. Moreover, each bin on socio-economic quantifier represents the average values of 65-248 data versus 220-982 data for social indicators as time progresses (for more details see the tables 1 and 2).

The growth law of the integration process can then be expressed over the specified periods mathematically by using curve fitting tools over the obtained bins. The curve fitting process reveal that the data set on socio-economic quantifiers can be well predicted by linear model \( r_F \Gamma \) while the nonlinear model \( r_J \sqrt{\Gamma} \) successfully projects the social data-set across the analyzed time sequences, as agreed with Barra et al. (2014).

\(^9\)The average method which is used here is so-called Global Mediant (see Barra et al. 2014)
Table 1: The statistical process on socio-economic integration quantifiers for the period 2005-2010.

### Permanent-Jobs

<table>
<thead>
<tr>
<th>Period</th>
<th>Total data</th>
<th>The number of bins</th>
<th>Population in each bin</th>
<th>$r_F$</th>
<th>$R^2_F$</th>
<th>$r_I$</th>
<th>$R^2_I$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>581</td>
<td>5</td>
<td>116</td>
<td>1.507</td>
<td>0.9758</td>
<td>0.431</td>
<td>0.7189</td>
</tr>
<tr>
<td>2</td>
<td>1165</td>
<td>11</td>
<td>106</td>
<td>1.555</td>
<td>0.9849</td>
<td>0.4658</td>
<td>0.7498</td>
</tr>
<tr>
<td>3</td>
<td>1684</td>
<td>12</td>
<td>140</td>
<td>1.569</td>
<td>0.9929</td>
<td>0.4896</td>
<td>0.7601</td>
</tr>
<tr>
<td>4</td>
<td>2279</td>
<td>10</td>
<td>228</td>
<td>1.560</td>
<td>0.9899</td>
<td>0.4882</td>
<td>0.7437</td>
</tr>
<tr>
<td>5</td>
<td>2875</td>
<td>12</td>
<td>240</td>
<td>1.515</td>
<td>0.9926</td>
<td>0.5014</td>
<td>0.7645</td>
</tr>
<tr>
<td>6</td>
<td>3471</td>
<td>14</td>
<td>248</td>
<td>1.515</td>
<td>0.9935</td>
<td>0.4988</td>
<td>0.7745</td>
</tr>
</tbody>
</table>

### Temporary-Jobs

<table>
<thead>
<tr>
<th>Period</th>
<th>Total data</th>
<th>The number of bins</th>
<th>Population in each bin</th>
<th>$r_F$</th>
<th>$R^2_F$</th>
<th>$r_I$</th>
<th>$R^2_I$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>581</td>
<td>9</td>
<td>65</td>
<td>1.800</td>
<td>0.9264</td>
<td>0.4755</td>
<td>0.6574</td>
</tr>
<tr>
<td>2</td>
<td>1165</td>
<td>12</td>
<td>97</td>
<td>1.818</td>
<td>0.9579</td>
<td>0.5376</td>
<td>0.6770</td>
</tr>
<tr>
<td>3</td>
<td>1684</td>
<td>13</td>
<td>130</td>
<td>1.881</td>
<td>0.9596</td>
<td>0.5788</td>
<td>0.6772</td>
</tr>
<tr>
<td>4</td>
<td>2279</td>
<td>11</td>
<td>207</td>
<td>1.887</td>
<td>0.9613</td>
<td>0.5895</td>
<td>0.6809</td>
</tr>
<tr>
<td>5</td>
<td>2875</td>
<td>13</td>
<td>221</td>
<td>1.861</td>
<td>0.9699</td>
<td>0.6059</td>
<td>0.6976</td>
</tr>
<tr>
<td>6</td>
<td>3471</td>
<td>15</td>
<td>231</td>
<td>1.843</td>
<td>0.9770</td>
<td>0.6098</td>
<td>0.7179</td>
</tr>
</tbody>
</table>

**Notes:** The predictive coefficients $r_F$ & $r_I$, and the coefficient of determinations $R^2_F$ & $R^2_I$ are computed by using curve fitting toolbox of MATLAB.
Table 2: The statistical process on social integration quantifiers for the period 1998-2008.

**Mixed-Marriages**

<table>
<thead>
<tr>
<th>Period</th>
<th>Total data</th>
<th>The number of bins</th>
<th>Population in each bin</th>
<th>$r_F$</th>
<th>$R_F^2$</th>
<th>$r_I$</th>
<th>$R_I^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2421</td>
<td>8</td>
<td>303</td>
<td>1.242</td>
<td>0.6997</td>
<td>0.3402</td>
<td>0.9767</td>
</tr>
<tr>
<td>2</td>
<td>4885</td>
<td>11</td>
<td>444</td>
<td>1.329</td>
<td>0.8703</td>
<td>0.3725</td>
<td>0.9164</td>
</tr>
<tr>
<td>3</td>
<td>7432</td>
<td>11</td>
<td>676</td>
<td>1.210</td>
<td>0.8591</td>
<td>0.3773</td>
<td>0.9373</td>
</tr>
<tr>
<td>4</td>
<td>10039</td>
<td>13</td>
<td>772</td>
<td>1.077</td>
<td>0.8384</td>
<td>0.4007</td>
<td>0.9757</td>
</tr>
<tr>
<td>5</td>
<td>12676</td>
<td>15</td>
<td>845</td>
<td>1.201</td>
<td>0.8539</td>
<td>0.4223</td>
<td>0.9816</td>
</tr>
<tr>
<td>6</td>
<td>15334</td>
<td>18</td>
<td>852</td>
<td>1.162</td>
<td>0.7977</td>
<td>0.4502</td>
<td>0.9904</td>
</tr>
<tr>
<td>7</td>
<td>18060</td>
<td>19</td>
<td>951</td>
<td>1.156</td>
<td>0.8076</td>
<td>0.4738</td>
<td>0.9882</td>
</tr>
<tr>
<td>8</td>
<td>20835</td>
<td>23</td>
<td>906</td>
<td>1.228</td>
<td>0.8476</td>
<td>0.4865</td>
<td>0.9893</td>
</tr>
<tr>
<td>9</td>
<td>23649</td>
<td>27</td>
<td>854</td>
<td>1.215</td>
<td>0.8512</td>
<td>0.5009</td>
<td>0.9930</td>
</tr>
<tr>
<td>10</td>
<td>26546</td>
<td>28</td>
<td>941</td>
<td>1.187</td>
<td>0.8583</td>
<td>0.5242</td>
<td>0.9931</td>
</tr>
</tbody>
</table>

**Newborns**

<table>
<thead>
<tr>
<th>Period</th>
<th>Total data</th>
<th>The number of bins</th>
<th>Population in each bin</th>
<th>$r_F$</th>
<th>$R_F^2$</th>
<th>$r_I$</th>
<th>$R_I^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2421</td>
<td>11</td>
<td>220</td>
<td>0.8164</td>
<td>0.9272</td>
<td>0.3099</td>
<td>0.9637</td>
</tr>
<tr>
<td>2</td>
<td>4885</td>
<td>14</td>
<td>349</td>
<td>0.8595</td>
<td>0.9522</td>
<td>0.319</td>
<td>0.9615</td>
</tr>
<tr>
<td>3</td>
<td>7432</td>
<td>18</td>
<td>411</td>
<td>0.8167</td>
<td>0.9469</td>
<td>0.3333</td>
<td>0.9616</td>
</tr>
<tr>
<td>4</td>
<td>10039</td>
<td>19</td>
<td>523</td>
<td>0.7761</td>
<td>0.8874</td>
<td>0.3027</td>
<td>0.9772</td>
</tr>
<tr>
<td>5</td>
<td>12676</td>
<td>21</td>
<td>582</td>
<td>0.8067</td>
<td>0.9008</td>
<td>0.3012</td>
<td>0.9778</td>
</tr>
<tr>
<td>6</td>
<td>15334</td>
<td>23</td>
<td>659</td>
<td>0.7239</td>
<td>0.7857</td>
<td>0.2823</td>
<td>0.9731</td>
</tr>
<tr>
<td>7</td>
<td>18060</td>
<td>27</td>
<td>667</td>
<td>0.6895</td>
<td>0.7746</td>
<td>0.2828</td>
<td>0.9794</td>
</tr>
<tr>
<td>8</td>
<td>20835</td>
<td>30</td>
<td>690</td>
<td>0.6791</td>
<td>0.7826</td>
<td>0.2812</td>
<td>0.9800</td>
</tr>
<tr>
<td>9</td>
<td>23649</td>
<td>30</td>
<td>780</td>
<td>0.6702</td>
<td>0.7906</td>
<td>0.2809</td>
<td>0.9803</td>
</tr>
<tr>
<td>10</td>
<td>26546</td>
<td>27</td>
<td>982</td>
<td>0.6484</td>
<td>0.8114</td>
<td>0.2831</td>
<td>0.9883</td>
</tr>
</tbody>
</table>

*Notes:* The predictive coefficients $r_F$ & $r_I$, and the coefficient of determinations $R_F^2$ & $R_I^2$ are computed by using curve fitting toolbox of MATLAB.
Those significant findings display that the values of coefficient \( r \) recorded over the indicated periods converge to a fixed number \( \tilde{r} \) and determine a curve which holds the formula:

\[
r(t) = \tilde{r}\left(1 \pm e^{-\frac{t}{K}}\right), \quad t = 2, \ldots, t_f.
\]

(5)

where \( t_f \) is referring to the final period and the parameter \( K \) is varying in the interval \((0, 1)\) fitted on past data. Based on this evidence, we suggest the parameter \( \tilde{r} \) as a predictor to estimate the coefficient \( r \) of subsequent period. Thereafter, we insert the predictor of \( r \) into our models, and proceed by predicting the future outcomes.

Our aim is to predict the quantifier values for new observations given their immigrant densities. It is worth noticing that in our approach, the value of each integration quantifier in the coming years also considers the new immigrants density \( \Delta \gamma \) and thus the new value becomes \( Q_i(\Gamma) + Q_i(\Delta \Gamma), i = 1, \ldots, 4 \), where \( Q \) represents the extracted growth low function for each quantifier and \( \Delta \Gamma = \Delta \gamma(1 - \Delta \gamma) \).

For instance, regarding the equation 5 assuming that the reported values of ”\( r \)” through all these periods of time is a series of numbers, ”\( r \)” will converge to the fixed numbers 0.5212 \& 0.2821 for the mixed marriages and births with mixed parents quantifiers and numbers 1.515 \& 1.861 for the quantifiers measuring the coefficient numbers of permanent and temporary jobs market respectively. Thus, if we use these ”\( r(t) \)” values as the set of points that determines the curve for the four quantifiers, the following formulas apply:
\[
\begin{align*}
\tilde{r}_{J_p}(t) &= \tilde{r}_{J_p}(1 \pm e^{\frac{t}{K_{J_p}}}) \quad \tilde{r}_{J_p} = 1.515, K_{J_p} \approx 0.0000; \\
\tilde{r}_{J_t}(t) &= \tilde{r}_{J_t}(1 \pm e^{\frac{t}{K_{J_t}}}) \quad \tilde{r}_{J_t} = 1.861, K_{J_t} \approx 0.0000; \\
\tilde{r}_{M_m}(t) &= \tilde{r}_{M_m}(1 - e^{\frac{t}{K_{M_m}}}) \quad \tilde{r}_{M_m} = 0.5212, K_{M_m} = 0.05221; \\
\tilde{r}_{B_m}(t) &= \tilde{r}_{B_m}(1 + e^{\frac{t}{K_{B_m}}}) \quad \tilde{r}_{B_m} = 0.2821, K_{B_m} = 0.1058.
\end{align*}
\]

where \( t = 1, \ldots, t_f \). As a result, the values at convergence are introduced into our models:

\[
\begin{align*}
Q_1 &= J_p(\Gamma) = 1.515\Gamma \quad \text{for permanent jobs;} \\
Q_2 &= J_t(\Gamma) = 1.861\Gamma \quad \text{for temporary jobs;} \\
Q_3 &= M_m(\Gamma) = 0.5212\sqrt{\Gamma} \quad \text{for mixed marriages;} \\
Q_4 &= B_m(\Gamma) = 0.2821\sqrt{\Gamma} \quad \text{for newborns.}
\end{align*}
\]

and which then forecast the out of sample. The quality of the subsequent forecasts turns out to be prompt, accurate and robust.

Figure 1 illustrates how the goodness of fit of our forecasts changes as we progressively add information from subsequent time periods. As shown in the Figure 1, this curve fitting exercise tells us the growth law that provides the best fit, when the entire set of data is considered, visibly delivers substantially better forecasts as early on as from period one. Even in the worst case scenario — see upper right panel in the Figure 1 — the forecast obtained by applying the square root growth law deliver more efficient estimates than the linear growth law already from period two and beyond. Thus, the predictability with respect to discerning the underlying mechanism for integration is prompt since the correct behavior emerges from the
dataset as early as the first year and no later than after two years.

Figure 2 shows how the forecasted values $\tilde{r}$ (grey points) based on past data match the values $r$ observed in the successive year. The match turns out to be very accurate — the Adjusted Mean Absolute Percentage Errors $^{10}$ are within 1%.

Figure 3, finally, analyzes the robustness of the forecasting ability with respect to identified models over the entire time span. The fit of the forecasted value vis-a-vis the observed value remain at an exceptionally high level also for large time intervals. The integration quantifiers exhibiting linear growth behavior are identified with high accuracy, with an $R^2$ $^{11}$ that never goes below .98 in the case of permanent contracts and .94 for temporary contracts. As for the two quantifiers evolving like the square root, the lowest $R^2$ is .91 for Mixed marriages, whereas in the case of newborns with mixed parents it is .95.

$^{10}$We used the Adjusted Mean Absolute Percentage Error which is a valid quantity to show the accuracy of the predictions: adjusted MAPE$=\frac{1}{n}\sum_{t=1}^{n} \frac{|f_t-y_t|}{(y_t+f_t)/2}$ where $f_t$ and $y_t$ are observed and forecasting values respectively (see Armstrong 1985).

$^{11}$ $R^2$ indicates how well the observed outcomes are replicated by the statistical model. The measure ranges from 0 to 1 such that the larger numbers representing better fits and also 1 indicates a perfect fit: $R^2 = 1 - \frac{\sum_{t=1}^{n}(y_t-f_t)^2}{(y_t-\overline{y})^2}$ where $y_1, ..., y_n$ are the observed values, $f_1, ..., f_n$ are the forecasts, and $\overline{y}$ is the average of the data (Draper and Smith 1998).
Figure 1: **Forecasts promptness.** Here each circular bullet displays the coefficient of determination computed for square-root fitting ($r_I/\sqrt{T}$) in the relevant period and each triangular bullet represents the analogous value recorded for linear model ($r_F/\Gamma$).

Figure 2: **Forecasts accuracy.** The black points represent the value of the coefficient $r$ (for the suitable growth law identified in Figure 1) obtained by real data collected up to the indicated year, grey points represent the forecasted value of the same coefficient obtained with the equation (5) - the Adjusted Mean Absolute Percentage Errors do are always under 1 %.
Figure 3: **Forecasts robustness.** Forecasts by the equation (5) are tested by the $R^2$ coefficient against real data at increasing time intervals. $R^2$ never goes below .91 even for forecasts 9 years ahead.

**Concluding discussion**

The main objective of this work was to propose a quantitative method capable of elaborating precise forecasts of immigrant integration as a function of immigrant density. This objective has been accomplished. The forecasting ability of the proposed model turns out to be prompt, accurate and robust. **Promptness** is found since the applied method successfully determines and differentiates between integration processes sensitive to social interaction or not, at very early stages (maximum two time periods). **Accuracy** and **Robustness** is found since, once we have determined the growth law, our prediction algorithm provide very accurate forecasts over the entire time sequence. Hence, using this framework it is possible to estimate the future rate
of, for example, intermarriage if the size of the immigrant population rise by say 2-3%.

An important quality of our forecasting framework is that it is capable of uncovering the underlying mechanism driving the integration process — social interaction or independent decision making — at an early stage in the immigration cycle. The capacity to foretell information of this type is by no means trivial since the two mechanisms are likely to demand different policy responses. For example, when integration grow linearly with immigrant density, as our labor participation indicators, effective policy responses should focus on problems such as access to labor markets, to improve integration, whereas integration induced by social action and interaction requires policies targeting the quality and intensity of interaction between immigrants and locals.

Our findings are of particular value to governments and researchers engaged in formulating more effective and precise immigration and integration policies. It is also an excellent research tool for scholars interested in explaining individual integration outcomes since it unveils in a general but powerful way the presence or absence of social network effects in integration phenomena.

Acknowledgments The authors thanks Adriano Barra and Cecilia Vernia for many useful discussions.

References


