The evolution of population at the local level in Argentina, 1895-2010.

Diminishing returns, agglomeration economies and congestion

Esteban A. Nicolini
Departamento de Ciencias Sociales – Universidad Carlos III de Madrid
INVECO – Universidad Nacional de Tucumán
CIEDH - Universidad del Norte Santo Tomás de Aquino

Valentina Fuentes Mortensen
INVECO – Universidad Nacional de Tucumán

Abstract

The relevance of agglomeration economies for the regional economic and demographic evolution of Argentina in the long run is estimated for the first time. We present a completely novel data set that combines a consistent set of 360 comparable quasi-departments recently elaborated by Talassino, Nicolini and Aráoz (2022) and a newly assembled data set with demographic and georeferenced information for nine benchmarks between 1895 and 2010. With this data set we estimate a polynomial version of a standard OLS regression and a non-linear kernel regression and robustly confirm a non-linear relationship between initial density and population growth: for quasi-departments with relatively low density (mostly from the initial periods) the relationship between population growth and initial density is negative (convergence in densities), for quasi-departments with intermediate density the relationship is positive (divergence) and for quasi-departments with very high density the relationship is negative again. Our findings suggest the existence of diminishing marginal returns for land abundant districts at the beginning, agglomeration economies in the intermediate case and significant congestion effects in the more densely populated districts, in particular those close to the city of Buenos Aires.

JEL classification: J0, N9, R12.
Keywords: Argentina, agglomeration economies, population growth

1. Introduction

Spatial economic and demographic heterogeneities in Argentina are huge. Average income in richest provinces is more than eight times the average income in poorer regions (Arázo, Nicolini and Talassino 2020). While population density in Patagonia is extremely low, the Gran Buenos Aires is one of the largest metropolitan areas in the world.

One of the obvious factors explaining the current spatial distribution of population and economic activity in Argentina is the extreme geographical variability and its associated dispersion in access to natural resources and productivity. Over this geographic context, several economic mechanisms interacted with each other in different ways since colonial times. After a clear orientation toward the mining centers in the current areas of Bolivia in the 16th and 17th centuries with a predominance of the provinces in the center and North of the country (Assadourian 1982), in the 19th century there was an increasing integration into the Atlantic economy with growing importance of the City of Buenos Aires and the area of the Pampa Húmeda, specialized on temperate crops (Cortés Conde 1979); in the 20th century, the inward looking economic strategy stimulated some industries in the large urban areas (reinforcing the already paramount relative importance of the city of Buenos Aires) and promoted the expansion of sectors like oil production in Patagonia. The combination of these processes defined a regional pattern of economic development marked by an intensification of a divide of a poor north and a rich south with the highest incomes per capita in the City of Buenos Aires and provinces in Patagonia (Aráoz, Nicolini and Talassino 2020).

This variability in the regional characteristics of economic development came together with specific demographic processes as well. In the second part of the 19th century, population in Argentina increased very rapidly because of a relatively early decline of mortality and a significant contribution of international immigration. Between 1890 and 1915 population growth was extraordinarily high with immigration accounting for up to 70% of this growth (Lattes 1974a). The expanding importance of the city of Buenos Aires in economic terms was paralleled by its role in the geographic distribution of population; until 1914 the capital of the country accumulated an increasing share of the population of the country and in the central decades of the 20th century this process of accumulation spilled out to the neighboring districts in the province of Buenos Aires. The other relevant process was the relatively high population growth in the areas only lately incorporated to the national economy (Patagonia and the North-East) that increase their share in the national population from very low initial levels at the end of the 19th century (Lattes 1974b).

The analysis of the spatial variation in population dynamics is important for several reasons. First, it can be a valuable input for our understanding of the evolution of living standards; on the one hand, many times migration (internal or international) is a consequence of regional differences in some dimension of standards of living; on the other hand, population, or density or both could be important determinants of productivity, wages, and many dimension of quality of life. More specifically, the patterns of population growth have been used to suggest possible explanations of productivity differentials through the interaction between diminishing marginal productivity in land and potential agglomeration economies (Desmet and Rappaport 2017, Beltrán-Tapia et al. 2018). In these approaches, identifying a negative relationship between population growth and initial population (or density) level (a kind of convergence) is taken as an indication of diminishing marginal productivity (Desmet and Rappaport 2017) while a positive relationship between these variables (divergence) is interpreted as a signal of agglomeration economies (Beltrán Tapia et al. 2018).

So far, the evidence is mixed. Beltrán Tapia et al. (2018) find that the effect of size was negligible in Spain the second half of the nineteenth century and centripetal forces only started consistently to induce population concentration after 1910. On the other hand, Desmet and Rappaport (2017) findings suggest that in the US between 1800 and 2000 the westward expansion into an open frontier economy generated a kind of U-pattern with negative correlation between population growth and initial population for districts with a low population level and positive for districts with larger populations.

A detailed analysis of the evolution of population and density levels in Argentina since the end of the 19th century provides an excellent opportunity to better understand the regional patterns of population growth in the long run and to test the existence of diminishing marginal returns or agglomeration effects. The resource endowment of Argentina is, in many senses, relatively like the one of the US with abundant fertile land and a process of recent occupation of previously unproductive regions but, at the same time, the expansion into an open frontier economy was a much rapid process that accelerated only at the end of the 1880s and was basically completed few years later (Cortés Conde 1979). The whole current area of the country was incorporated into the national economy, at least nominally in 1895, which is the starting point of our analysis.

The main contributions of this paper are, first, the presentation of a new data set with the evolution of population in homogenous and comparable 360 administrative units in nine benchmarks between the end of the 19th century and the beginning of the 21st century (1895, 1914, 1947, 1960, 1970, 1980, 1990, 2001, 2010) ; second, the incorporation in these data set of several geo-referenced variables that can eventually be relevant for explaining productivity and population growth (like soil characteristics, suitability for agricultural production, and proximity to big cities, or population of contiguous administrative units). Third, the estimation of two different econometric models that identify a robust non-linear relationship between population growth and initial population density: for land-abundant, low-density departments (most of them in Patagonia and in most of the cases at the end of the 19th and beginning of the 20th century), there is convergence, meaning faster population growth in low density departments. On the other hand, for departments with intermediate density the relation between population growth and initial density is positive suggesting the existence of agglomeration economies. Finally, we identify the emergence of a new type of congestion for departments with very high density implying that among denser departments, population density is detrimental for population growth, probably because of urban disamenities and housing prices.

After a discussion of the historical context and the presentation of a selection of the relevant literature in section 2, we present our methodology and our data set in section 3. The last two sections present the results and the conclusions. At the end of the paper we include an appendix with some ancillary information.

2. Historical context and literature review

Until the 18th century, most of the economic activity of the southern cone was oriented toward the mining center of Potosi and the more important areas of the were the ones located in what is the current north and north-west part of Argentina (Assadourian 1982, Maloney and Valencia Caicedo 2016). The administrative changes of the bourbon dynasty and increasing importance of the Atlantic economy expanded the roles of the city of Buenos and its surroundings. This tendency was reinforced after independence and the consolidation of Buenos Aires as the most important administrative and commercial center of the country.

The period of the first globalization and the integration of Argentina into the international division of labor induced a set of important changes in the regional profile of economic activity of the country: (i) Increasing profitability of land in the Pampa Húmeda area (ii) confirmation of the importance of Buenos Aires and (iii) Expansion of production, mainly sheep raising in Patagonia at the end of the 19th century. Even though the dominant economic model in this period is usually characterized as being focused on exports of agropastoral goods according to comparative advantages in the international trade arena, the production of non-tradable or semi-tradable goods implied a relatively large manufacturing sector located mostly on urban centers (Fajgelbaum and Redding 2020). Additionally, the expansion of the export sector created complementarities with a very specialized services sector in banking, finance, trade, and transportation mainly concentrated in the city of Buenos Aires (Aráoz and Nicolini 2020). Finally, in most of the large and medium size cities, the high rates of population growth and urban expansion pushed the building sector (Cortés Conde 1979).

The model of economic development after 1930 changed substantially with the Import Substitution Industrialization and the expansion of the secondary and tertiary sectors. The participation of manufactures in the total value added of the country increased until the 70s when a new process of globalization with a new boom of commodities interrupted the process and initiated a new expansion of the primary and tertiary sectors.

Many features of the demographic evolution of Argentina can be properly characterized by standing in an intermediate position between the more advanced Western economies and the relatively backward neighbor countries in Laitn America. Since the end of the 19th century, both population growth and changes in the regional composition were mainly determined by migration with the Pampean region (including the city of Buenos Aires) increasing its share of population between 1895 and 1914 until a 70 % of the total and then staying at this level during most ot the 20th century. The other relevant changes in the regional participation in total population are the relative growth of the Gran Buenos Aires (this is the area including the city of Buenos Aires and some surrounding districts in the province of Buenos Aires[1]) from 19.8% in 1895 to 35.3% in 1970, the relative expansion of the areas that started with very low density like Patagonia (Lattes 1974b).

In this context, a better understanding of the processes of economic development and population growth and how these processes were distributed across regions can shed more light on the mechanisms behind the interaction between the productivity differentials associated with abundance of natural resources and those linked to of population concentrations. The causes and implications of the uneven distribution of people and economic activity across space are in the core of the research on agglomeration economies. The seminal paper by Ciccone and Hall (1996) tries to explain differences in labor productivity across counties in the US by estimating two models in which spatial density results in aggregate increasing returns through three main mechanisms: (i) constant returns to scale at the firm level but transportation costs from one stage of production to the next, (ii) externalities associated with physical proximity, and (iii) possibilities of higher degree of specialization. Their main empirical approach is regressing the log of nominal wage (proxy for workers' productivity) on the log of density; given that in their case there are potential risk of endogeneity[2] they use several versions of instrumental variables.

Many other attempts have been made to check the existence and to estimate the magnitude of the agglomeration economies. One branch of the literature focused on the contemporaneous relationship between wages and density and, like Ciccone and Hall (1996) were forced to look for credible instruments to solve the problem of endogeneity (Combes

[1] For a more precise definition of what is the Gran Buenos Aires, see section 3.
[2] In this case the risk of endogeneity is related to the fact that it is unrealistic to assume that fundamentals differences in productivity across states are unrelated to population density.

et al. 2010, Combes and Gobillon 2015). When the focus is more on the long run it has been quite usual to use population growth as dependent variable under the explicit or implicit assumption that relatively higher rates of population growth between two periods are a kind of proxy for relatively higher productivity; if this is the case, population or density convergence would imply diminishing marginal returns (growth is inversely related to initial levels) and divergence would be a signal of agglomeration economies.

The use of population growth as dependent variable in the studies concentrated in the long run reduces the risk of endogeneity given the time structure of the data but the potential problem of biases because of omitted variables remains important.[3] Desmet and Rapapport (2017) try to understand the relationship between initial population and population growth in the US counties between 1800 and 2000. The period under study allows them to incorporate areas with very low density in areas in the West of the country only gradually incorporated to the national economy. They estimate (i) kernel regressions with continuous non-linear approximations of growth versus initial population and (ii) continuous, piecewise-linear splines and with the two methodologies find that there is a negative correlation between growth and levels (convergence) in the 19th and early 20th centuries but a moderate divergence in the 20th century. The initial convergence would be related to new and low-density counties in the West and their growth until their long run level of population. The subsequent divergence is explained by a decrease in net congestion arising from a relative decline from land-intensive production and an increase in the returns to agglomeration associated to the introduction of several general-purpose technologies during the twentieth century (Desmet and Rappaport 2017).

Beltrán et al. (2018) study district population in Spain between 1860 and 1991 to test the hypothesis that initial population affect subsequent population growth assuming that a significant positive relationship would imply the existence of significant agglomeration economies. They find that the positive effect of initial population on subsequent growth increased between 1910 and 1970 because standard agglomeration economies but it decreased in the 1970s and the 1980s because congestion costs associated to the slow-down of rural-urban migration and rising housing prices and other cost related to urban concentrations.

The connection between structural change and variation of population growth rates across districts is questioned by Eckert et al. (2023) because their finding that structural change in the US between 1880 and 1940 was primarily a local phenomenon and most sectoral reallocations took place within counties.[4] This is not necessarily contradictory with the results by Desmet and Rappaport (2017): Eckert et al. clarify that there was spatial reallocation but that this migration between industrialized and rural states was not systematically related to sector switching; instead, the most significant sectoral shifts did not occur via the expansion of incumbent cities but rather through the birth of new cities and towns in the rural hinterland. In this sense, spatial agglomeration forces were essential because workers clustered in newly formed cities specialized in manufacturing. At the same time, agglomeration effects were not strong enough to attract workers from afar to existing industrial hubs.

[3] TAKE CARE HERE… SEE Combes and Gobillon (2015, p. 285); the suggest there that endogeneity may arise because an omitted variable correlated with both density and wages… IS THIS ENDOGENEITY??

[4] [this goes against our hypothesis …Some information on urban population within districts…]

Regarding the evolution of the spatial distribution of demographic and economic growth in Argentina, Aráoz and Nicolini (2020) use novel estimations of subnational GDPs in 1914 to hypothesize that in the first half of the 19^{th} century regional growth responded to a combination of low population density and land abundance in some districts in Patagonia and high density and agglomeration economies mainly around of the city of Buenos Aires. Also for Argentina, Talassino and Herrera (2021) identify the relevance of local spillovers by finding that the central decades of the 20^{th} century, conditional to some controls, growth of the departments (a very small spatial-administrative unit) was faster/slower because of the per capita income (growth rates) of their neighboring departments.

3. Methodology and Data

3.1. Methodology

The estimation of agglomeration economies and congestion effects requires first the choice of the correct independent variable. The main hypothesis in the seminal paper by Ciccone and Hall (1996) is that spatial *density* results in increasing returns[5]; empirically they use as instruments both population density and population levels, accepting both as measures of agglomeration (p. 61). However, they explore the size versus density effects and conclude that "density externalities are more important than size externalities at the county level" (p. 67-68). Desmet and Rappaport (2017) use initial population levels. Beltrán et al. (2018) also use initial population levels as independent variables, while additionally employing land area as a control variable (appendix table 3, p. 110); the parameter of land area is negative and statistically significant, but it is smaller in absolute value than the parameter associated with population, suggesting a kind of double effect. Briant et al. (2010), in line with Ciccone and Hall (1996), use workers' density (equation 2, p. 295).

As is evident from the aforementioned studies, it is necessary to consider the size of the spatial unit of reference when assessing agglomeration economies and congestion effects. This consideration can be addressed by either incorporating it, along with the population, into a single variable—namely density—or by including it as a control variable in the model. The mechanisms underlying the benefits (or costs) of agglomeration do not solely depend on the size of the population within a specific administrative unit area, but rather on how tightly or closely individuals are located. With an understanding of this aspect, we will employ population density as the independent variable, in line with the majority of the previous studies mentioned[6].

The coexistence of two types of mechanisms influencing the relationship between population growth and initial population levels (agglomeration effects and congestion) suggests the possibility of a non-linear relation between those two variables. We will use two estimations that can potentially capture these non-linearities: the simplest approach is a kind

[5] They mention that previous studies looking for increasing returns have focused on city size while they "… believe that density rather than city size is the more accurate determinant" (p. 55).

[6] In any case, the idea is present, as an extension of this work, to analyze the results of the proposed models using the initial population size as the main independent variable to attempt to explain the subsequent growth of administrative units. Other potential extensions for further research could involve using, for example, urban population or population aged 14 years and older (this definition resembles that of the economically active population). However, these definitions are not available at high levels of disaggregation (departments) in all censuses. Regarding urban and rural population data for departments, genders, and nationalities, we only have information for the years 1895 and 1914. For 'Population aged 14 years and older, classified by gender, living area, department, and illiteracy,' data is only available for 1947.

of convergence regression of population growth on a polynomial of population density. The second approach is, following Desmet and Rappoport (2017) and Desmet and Fafchamps (2005), a non-linear Kernel regression.

Regarding the first approach, we will estimate the following equation:

$$\Delta ln(POP)_{it} = \alpha_0 + \sum_{k=1}^{K} \alpha_k [\ln{(DENS_{it})}]^k + \sum_{j=1}^{J} \beta_j x_{ijt} + e_{it} \quad (1)$$

In this context, the dependent variable is the population growth rate between periods t+1 and t. The primary independent variable is the logarithm of density in the initial period t. Here, 'k' represents the order of the polynomial, while 'x' denotes the control variables.

In the specification of equation (1), the values of the alfas characterize the type of (eventually non-linear) relationship between population growth and initial density. As we will see in next section, a third-order polynomial provides enough flexibility to unveil several interesting features of the dynamics of congestion and agglomeration effects. What we are testing in the first equation is a form of conditional convergence, similar to economic growth studies, while allowing for non-linearities assuming a particular functional form.

In the second approach, we will estimate the following nonparametric regression

$$\Delta ln(POP)_{it} = G(\ln{(DENS_{it})}, X_{it}) + e_{it} \quad (2)$$

Where the function G(.) is possibly non-linear. Kernel regressions yield continuous nonlinear approximations of growth versus initial density (Desmet and Rappaport 2017). These estimations provide an alternative more flexible way to capturing non-linearities in the sense that nonparametric regression is agnostic about the functional form between the outcome and the covariates and is therefore not subject to misspecification error.

In this way, based on these nonparametric kernel estimations, we compute a function that describes the mean of the outcome, which is population growth, at different values of the initial population density. In Section 4, we present these results by plotting the average annual population growth as a function of the initial log density. In this context, a positive slope would indicate the presence of agglomeration economies, a negative slope would indicate the dominance of diminishing marginal returns or congestion costs over the benefits of agglomeration, and a zero slope would somewhat support Gibrat's law (the orthogonality of growth and initial levels).

3.2. Data

Argentina is a federal state comprising 24 first-level administrative divisions: 23 provinces and the Autonomous City of Buenos Aires. These provinces, in turn, are subdivided into second-level administrative units known as departments, or *partidos* as in the case of Buenos Aires. Our focus is on examining the relationship between initial population density and subsequent growth at the lowest possible level of geographical disaggregation, which corresponds to the departments.

The challenge when dealing with small geographic units over long time periods is that their boundaries, shapes, and quantity usually change, particularly tending to increase over time. Between 1895 and 2010, there were numerous modifications in departments boundaries. In 1895, the number of departments was 394, while by 1960 it increased to 487. This change is not a result of territorial expansion but rather stems from modifications in sizes and limits within the existing AU.

In order to build a set of geographically consistent administrative units there are (at least) two strategies: (1) the approach proposed by Talassino et al. (2022) and (2) the method introduced by Desmet and Rapapport (2017). The first approach involves constructing a classification of geographical land areas that remains time-invariant throughout the entire analyzed period; it means that the set of administrative unites (and observations in the data set) is unique for the whole period under analysis. The second approach generates geographically consistent areas only over successive twenty-year intervals meaning that there is a separate dataset for each of the ten twenty-year periods they study, accounting for changes in the number of administrative units over time. However, the issue with the second strategy is that more densely populated units are more likely to be split.[7]

Following the classification proposed by Talassino et al. (2022), we reorganize the geographic and administrative divisions of each census year between 1895 and 2010 into 360 standardized administrative units (hereafter AU).

There is some concern about how the definition of observable units can lead to different estimation outcomes. Briant et al. (2010) address this concern and refer to it as the modifiable area unit problem (MAUP). These authors argue that when spatial units remain small, altering their size only slightly changes economic geography estimates, and altering their shape matters even less. They assert that both of these distortions are secondary compared to specification issues. Since we are using the smallest possible definition of administrative units, which are departments in this case, this problem should not be a cause of concern.

Using total population data from the censuses conducted in 1895, 1914, 1947, 1960, 1970, 1980, 1991, 2001, and 2010, which were provided for departments based on the valid administrative organization of each respective period, we aggregated these figures to match our 360 AUs for each census year.

In addition to population data, for each quasi-administrative unit, we possess information on area in square kilometers (which facilitates the construction of the *density* variable), as well as distances to the nearest port[8] (*dist_port*) in kilometers. We also control for whether the AU serves as the capital of its province and whether it belongs to the Gran Buenos Aires (*gba*) area[9]. A set of binary variables has been added to the dataset to account for the effects of the

[7] Given that regions with higher population density tend to be more often split, if regions with agglomeration economies experience higher growth, the estimation of the correlation between population growth and initial population will be downwardly biased.

[8] These distances represent the measurements from the centroid of the AU to the nearest port. We take into account the four primary ports of Argentina, situated in Buenos Aires, Rosario, La Plata, and Bahía Blanca.

[9] Following the classification proposed by INDEC (2003), the inclusion in the Great Buenos Aires (GBA) area is considered in a broader sense. This means that we encompass the 30 AUs or *partidos* that are part of the GBA Agglomerate, as opposed to the 24 that fit the definition of GBA alone (INDEC 2003).

AU being located within a specific region[10]. In fact, we treat the capital of the Republic (*Capital Federal*) as a region itself and use it as our base group in the regression of equation (1) presented in Section 4.

We recognize that the demographic growth of a specific AU can be influenced by the dynamism of neighboring units. To address this effect, we utilize population density as an indicator of dynamism and consider two key factors: geographic proximity and adjacency. In the first case, we compute the average population density of the departments whose centroids fall within a 10-kilometer radius of the AU's centroid (*avproxden*); in the second case, we calculate the average population density of the departments that share a border with the AU (*avadjden*).

Below, we present some descriptive statistics for a subset of the variables included in our dataset.

Variable	Mean	Min	Max	Standard Deviation
population	64,215.8	44	2,982,580	186465.7
density	169.25	0.0004803	14532.3	969.247
dist_port	502.7	18.03421	1806.05	344.3391
avproxden	171.71	0	3641.42	458.4099
avadjden	180.44	0.0096812	4849.64	509.4378
pop growth	0.014072	-0.1485	0.226642	0.0205644

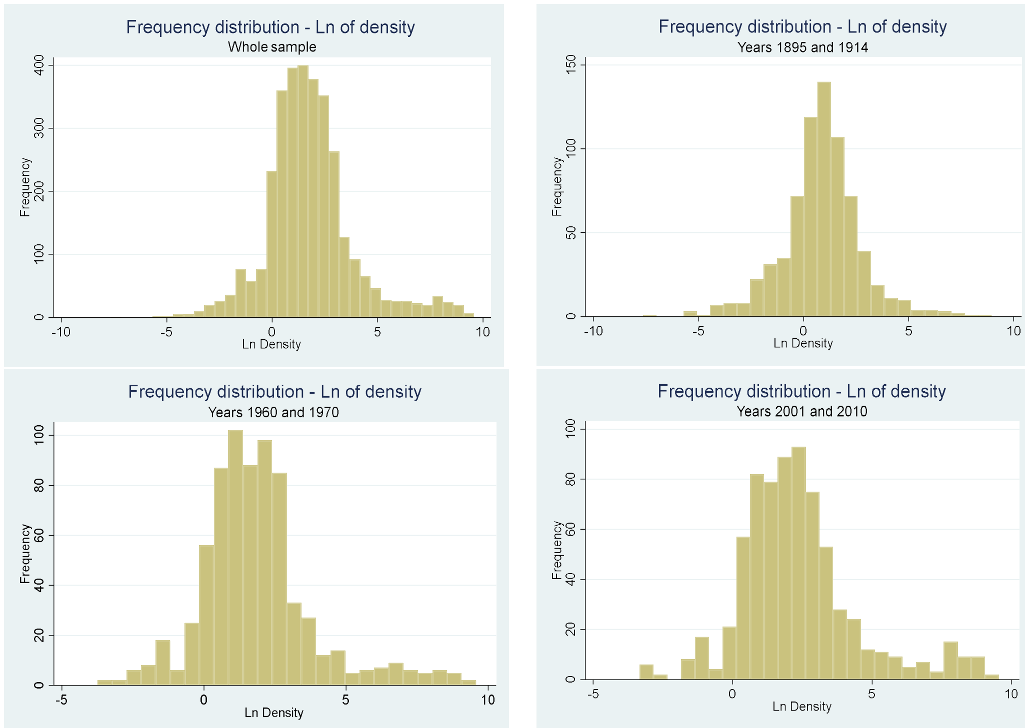
There are only twelve departments with a density of less than 0.02 inhabitants per square kilometer in the whole sample; eleven of them are in 1895 and eight of them are in Patagonia. The department of Antofagasta de la Sierra in Catamarca is in that set in 1895 and in 1914; the other two departments in this set are in Chaco and La Pampa in 1895, both in areas of new settlement. As expected, the departments with the highest density, in addition to Capital Federal, are neighbors of Capital Federal (Barracas al Sud, San Martín and San Isidro) and in the period after 1970.

The distribution of the natural logarithm of the densities (our independent variable in the regressions) is presented in Figure 1. It tends to shift to the right with time with a mean of 0.56 in 1895 and 2.39 in 2010. This shift is much more acute in the right tale of the distribution with the minimum value increasing from -7.64 in 1895 to -3.3 in 2010; in the right tale of the distribution the change is smaller with the maximum moving from 8.08 in 1895 to 9.55 in 2010. This suggests a hypothesis that will be analyzed in further detail in the next section that once the density is high enough, population growth tend to stall or decelerate.

---

[10] The composition of the regions was determined as follows, listing the provinces and their corresponding regions in parentheses: Capital Federal (CF); Buenos Aires, Córdoba, Santa Fé, and La Pampa (BACS); Mendoza, San Juan, San Luis, and La Rioja (CUY); Entre Ríos, Corrientes, and Misiones (LIT); Chaco and Formosa (NEA); Jujuy, Salta, Tucumán, Santiago del Estero, and Catamarca (NOA); Neuquén, Río Negro, Chubut, Santa Cruz, and Tierra del Fuego (PAT).

Figure 1



4. Results

The estimations of equation (1) are conducted for the entire dataset using pooled OLS, and the results are presented in Table 1. For these estimations we are setting the value of *k* in 3, this means we are using a polynomial of order three[11].

The first column corresponds to the estimation of equation (1) without control variables. This is equivalent to testing for unconditional convergence in terms of population growth concerning initial density levels. As we can observe, the parameters are highly significant and confirm a non-linear relationship between growth and initial log of density. The sign and significance of these parameters remain robust even with the inclusion of control variables (columns 2 and 3).

Columns 2 and 3 differ in that the latter includes control variables that identify whether the AU belongs to a specific region. The AU being the capital of its province has a positive effect on population growth, as seen in columns 2 and 3, with the magnitude of this effect closely aligning with the average growth of the AU's populations over the entire period (see Table 1). Additionally, belonging to the GBA agglomerate has a positive, significant, and even bigger effect on population growth. The *gba* parameter almost doubles the previous mentioned average population growth. One can interpret this effect as a sort of spillover from the *Capital Federal* to its neighboring areas. Interestingly, when exploring the influence of neighboring units on a specific AU's population growth, through the variable *avproxden,*

[11] Using a polynomial of order four, the parameter of density to the power of four is not statistically significant.

we observe the reverse effect[12]. The parameter is negative and highly significant; this implies that an AU with a relatively high-density neighbor would have smaller population growth rates, as this neighbor acts as an attracting force. Nevertheless, this effect is relatively small.

Another significant variable, although with a very small positive effect, is the distance to the nearest port (*dist_port*). This variable could account for two opposing effects: on the one hand, a positive relationship between distance to the nearest port (located in the middle of the country) and population growth, due to the fact that an AU located more far away from the center (as the ones belonging to Patagonia or NEA regions, for example) could be considered as land abundant, so we have a convergence effect and higher population growth. On the other hand, a negative relationship could emerge because the AUs located near the ports are more likely to present agglomeration effects.

As evident from column 3, the binary variables included for the different regions are not statistically significant, except for Patagonia. For this region, the parameter is positive and significant, with a magnitude very similar to the average growth. This suggests that if an AU belongs to Patagonia, it exhibits a higher rate of population growth. This can be interpreted as a convergence effect within the context of land abundance mentioned in the previous paragraph.

Table 2: Pooled OLS estimates

VARIABLES	(1) Pop. growth	(2) Pop. growth	(3) Pop. growth
laglndens	-0.002388***	-0.001954***	-0.001050***
	(0.000276)	(0.000280)	(0.000296)
laglndens2	0.001931***	0.001701***	0.001552***
	(0.000086)	(0.000084)	(0.000085)
laglndens3	-0.000187***	-0.000210***	-0.000198***
	(0.000011)	(0.000010)	(0.000011)
capital		0.013199***	0.009197***
		(0.001541)	(0.001580)
gba		0.029506***	0.028253***
		(0.002005)	(0.001983)
dist_port		0.000004***	0.000007***
		(0.000001)	(0.000002)
avproxden		-0.000002***	-0.000002***
		(0.000001)	(0.000001)
BACS			0.000083
			(0.007124)
CUY			-0.001303
			(0.007193)
LIT			-0.000113
			(0.007264)
NEA			0.010936
			(0.007726)

[12] In section 3 we mentioned another variable to control for neighboring effects that is *avadjden*. However, we don't include it in the regressions due to the fact that is highly correlated with *avproxden*.

NOA			-0.004999
			(0.007266)
PAT			0.012892*
			(0.007325)
Constant	0.010512***	0.008101***	0.006534
	(0.000530)	(0.000848)	(0.007071)
Observations	2,880	2,880	2,880
R-squared	0.151937	0.224603	0.255314

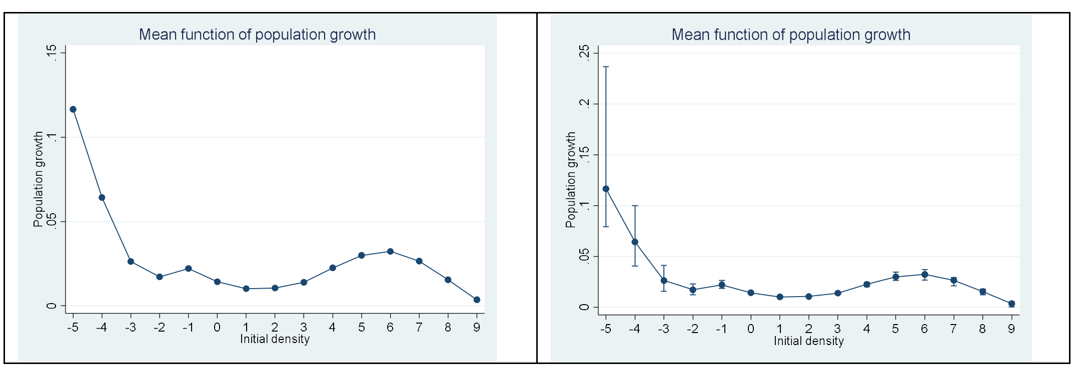
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The value of the parameters associated with initial log of density imply that, conditional of the other variables, population growth has a kind of overturned S-shape with a downward slope when values of the log of initial density are smaller than 0.36, upward slope when they are between 0.36 and 4.86 and a downward slope again for values larger than 4.86. This implies that in the left part of the distribution of densities, the larger is this variable, the smaller is the subsequent population growth; in other words, for the 637 observations with very low density (with the log of density smaller than 0.36), the model predicts convergence; 254 of them (40%) are either from 1895 or 1914 and 114 are from Patagonia (70% of the AUs from Patagonia are in this group).

For departments in the middle range of log of density (between 0.36 and 4.86) agglomeration effects dominate. There are 2370 departments in this range and 1078 are from the region Pampa Húmeda. Finally, the 233 departments with an initial log of density above 4.86 start to experience congestion and their population growth is inversely related to initial density. The region of Pampa Húmeda is clearly overrepresented in this group (145 belong to this region) and NEA and Patagonia do not have any department in this group.

Next, we move to present results corresponding to the estimation of the nonparametric kernel regressions.[13] Figure 2 shows the predicted population growth rates (the outcome) for a vector of values of a specific covariate (the natural logarithm of the density). In the right part of the Figure, we have the confidence bands of the estimation with bootstrap standard errors.

Figure 2



[13] The non-parametric regression was run using the Stata software and the command *npregress* with option *kernel*. The regression output is presented in the appendix.

The results from Kernel regressions are consistent with our results using a polynomial approach: the correlation between population growth and initial density is characterized by three regimes: when the initial log of density is smaller than 1, although with an interruption between -2 and -1, there is convergence in population density in the context of land abundance and decreasing marginal returns to land.

When the initial log of density is between 1 and 6, there is divergence with dominance of agglomeration economies. For departments with log of density above 6, congestion generates convergence with lower growth rates for the most densely populated AUs.[14]

According to these estimations, when the log of population density is less or equal to -3, typical of sparsely populated departments in Patagonia at the beginning of the 20th century or the department Antofagasta de la Sierra in the province of Catamarca, the population growth is between 3% and 12%. But, when the initial log of density is around 1 (typical of AUs of Buenos Aires that do not belong to GBA on the entire period, for example[15]) or around 8 (typical of some departments from the provinces of Buenos Aires and Mendoza in the last quarter of 20th century and beginning of the 21st), population growth is around 1.3%.

5. Conclusions

The current large actual disparities in concentration of population and economic activity in Argentina are the result of a complex combination of historical processes and economic dynamics. The analysis of the distribution of population growth across small administrative units and across time between the end of the 19th century and the beginning of the 21st century show that three main forces help to explain the pattern of demographic concentrations. For low density and land abundant areas the dominant process is one of convergence: the lower the initial density, the higher the population growth; this happens more frequently in the initial periods of our study (at the end of the 19th century and first decades of the 20th century) and in the regions that were only lately incorporated to the national economy. This is consistent with the results obtained by Desmet and Rapapport (2017) for the US in the 19th century.

On the other hand, for the AUs within the intermediate range of density levels there is divergence with higher growth for departments with higher levels of initial density suggesting that agglomeration effects dominate. These results are consistent with findings from Desmet and Rapapport (2017) beginning in the middle of the 20th in the US and from Beltrán Tapia et al. (2018) for Spain between 1910 and 1970.

We also identify a range of AUs that present convergence in population density because the congestion forces dominate the potential agglomeration effects. This is consistent with the findings of Beltrán Tapia et al. (2018) that suggest that in Spain, agglomeration effects weaken because of the lower intensity of the reallocation from agriculture and the increase in congestion costs. In this sense, our results suggest that the process in Argentina is a combination of what happened in the land abundant US in the 19th century and the process

[14] Although the inflection points in the nonparametric regressions are not identical from the ones obtained using the polynomial estimates, they are in fact similar: the local minimum using Kernel is 1 while using the polynomial is 0.36. In the case of the local maximum, using Kernel is 6 while using the polynomial is 4.86.
[15] Nevertheless, many AU from different provinces presents values of log of density around 1.

of reduction of rural-urban migration and declination of traditional industries in the last decades of the 20th century Spain.

Our results also suggest that in very large urban areas like the Great Buenos Aires the consequences of congestion implied a higher average population growth related to the demographic spillover from the extremely high-density Capital Federal. However, in general, the existence of nearby AUs with high density implied a lower population growth, probably because of agglomeration economies generating the "absorption" of population from low-density districts to high-density ones.

6. References

Aráoz and Nicolini (2020). "Regional Growth and the Persistence of Regional Income Inequality in Argentina in the First Half of the Twentieth Century". *Journal of Latin America Studies* 52, 293-316.

Aráoz, Nicolini and Talassino (2020). "Growth and convergence among Argentine provinces since 1895" in Badía-Miró, Tirado and Willebald (eds.): *Time and Space: Latin American Regional Development in Historical Perspective*, Palgrave editors. 65-96.

Assadourian, Carlos Sempat (1982). *El sistema de la economía colonial: mercado interno, regiones y espacio económico*. Instituto de Estudios Peruanos, Lima.

Briant et al. (2010). Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations? *Journal of Urban Economics* 67, 287-302.

Ciccone and Hall (1996). Productivity and the density of economic activity. American Economic Review 86, 54/70.

Combes and Gobillon (2015). The Empirics of Agglomeration Economies. *Handbook of Regional and Urban Economics*, Volume 5A, Chapter 5, 247-348.

Cortés Conde, Roberto (1979). El progreso económico Argentino 1880-1914. Editorial Sudamericana.

Desmet and Rappaport (2017). The settlement of the United States 1800-2000. *Journal of Urban Economics* 98, 50-68.

Glaesser (2010). Introduction. In Glaeser, E. (ed.). *Agglomeration economies*. University of Chicago Press.

INDEC (2003). ¿Qué es el Gran Buenos Aires? Publicaciones del INDEC.

Beltrán Tapia, Francisco J. & Díez-Minguela, Alfonso & Martinez-Galarraga, Julio, 2018. "Tracing the Evolution of Agglomeration Economies: Spain, 1860-1991," The Journal of Economic History 78, 81-117.

Fajgelbaum and Redding (2022) Trade, Structural Transformation, and Development: Evidence from Argentina 1869–1914. *Journal of Political Economy* 130, 1249-1317.

Lattes, A. (1974a). "El crecimiento de la población y sus componentes demográficos entre 1870 y 1970". In Recchini de Lattes, Z and Lattes, A. (comp.). *La población de Argentina*, CICRED Series, 29-66.

Lattes, A. (1974b). "Redistribución espacial y migraciones". In Recchini de Lattes, Z and Lattes, A. (comp.). *La población de Argentina*, CICRED Series, 95-113.

Maloney, William and Valencia Caicedo, Felipe (2016). The persistence of (subnational) fortune. The Economic Journal 126, 598, 2363-2401.

Pantelides, Edith (2009). "Completing the fertility transition: the case of Argentina". In United Nations Population Division. *Completing the Fertility Transition. Population Bulletin of the United Nations*, Special Issue, Nos. 48/49 2002, New York.

Pantelides, Edith and Moreno, Martín (cords.) (2009). Situación de la población en Argentina Buenos Aires. Programa Naciones Unidas para el Desarrollo – PNUD – UNFPA.

Rofman, Rafael et al. (2022). Odisea demográfica. Tendencias demográficas en Argentina: insumos claves para el diseño del bienestar social. Buenos Aires: CIPPEC, UNICEF y UNFPA.

Talassino, Aráoz and Nicolini (2022). "Regional pattern of economic activity in Argentina between the 19th and the 20th centuries". Paper presented in Figuerola Workshop, Universidad Carlos III de Madrid. January 2023.

Talassino, Mauricio y Herrera Gómez, Marcos (2021). "Impacto de los efectos espaciales en la convergencia regional. Análisis departamental para la Argentina". Documento de Trabajo de la Red Nacional de Investigadores en Economía (RedNIE), Nº 42..

United Nations (1992). "Estudios de casos de políticas de la población. Argentina". Políticas de la población, Documento 26. Nueva York.

Walter, Richard J. (1993). *Politics and Urban Growth in Buenos Aires, 1910–1942*. Cambridge University Press.

7. Appendix

A. Results from the Kernel regression

```
Local-linear regression                    Number of obs      =        2,879
Kernel   : epanechnikov                    E(Kernel obs)      =        1,211
Bandwidth: cross validation                R-squared          =       0.1885
------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Percentile
      growth |   Estimate   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Mean         |
      growth |   .0143327   .0004169    34.38   0.000     .0136075    .0150048
-------------+----------------------------------------------------------------
Effect       |
   laglndens |  -.0004696   .0003153    -1.49   0.136    -.001105    .0001329
------------------------------------------------------------------------------
Note: Effect estimates are averages of derivatives.
```

B. Contributions of different regions and different years to the three groups of AUs

Ln Density smaller than 0.36										
					year					
region	1895	1914	1947	1960	1970	1980	1991	2001	2010	TOTAL

Capital	0	0	0	0	0	0	0	0	0	0
Pampa	45	19	11	11	15	16	16	16	14	163
Cuyo	32	32	19	17	17	16	15	12	13	173
Litoral	10	5	1	0	0	0	0	0	0	16
NEA	5	4	1	0	0	0	0	0	0	10
NOA	37	29	16	17	16	15	13	9	9	161
Patag	18	18	16	16	15	11	8	7	5	114
TOTAL	147	107	64	61	63	58	52	44	41	637

Ln Density between 0.36 and 4.86										
					year					
region	1895	1914	1947	1960	1970	1980	1991	2001	2010	TOTAL
Capital	0	0	0	0	0	0	0	0	0	0
Pampa	108	128	130	126	121	118	116	115	116	1078
Cuyo	21	19	31	33	33	33	33	35	34	272
Litoral	38	43	46	47	47	46	46	46	46	405
NEA	0	1	4	5	5	5	5	5	5	35
NOA	41	49	62	61	61	61	63	67	67	532
Patag	0	0	2	2	3	7	10	11	13	48
TOTAL	208	240	275	274	270	270	273	279	281	2370

Ln Density larger than 4.86.										
					year					
region	1895	1914	1947	1960	1970	1980	1991	2001	2010	TOTAL
Capital	1	1	1	1	1	1	1	1	1	9
Pampa	1	7	13	17	18	20	22	23	24	145
Cuyo	2	4	5	5	5	6	7	8	8	50
Litoral	0	0	1	1	1	2	2	2	2	11
NEA	0	0	0	0	0	0	0	0	0	18
NOA	1	1	1	1	2	3	3	3	3	0
Patag	0	0	0	0	0	0	0	0	0	233
TOTAL	5	13	21	25	27	32	35	37	38	233