Quantitative Methods in Finance

COVID-19 Population Gap

Charaf Zguiouar
Francisco Javier Alvarez Aparicio
Lucas Rouleau
Ludovic Paul
Maysaa Rais
Natalia Cárdenas Frías

November 12, 2023

Abstract

The primary objective of this study is to evaluate the discernible impact of the COVID-19 pandemic on global and national population growth trends, specifically focusing on France, Japan, Sweden, the USA, and Colombia. We aim to quantify the population gap attributed to the pandemic by analyzing historical population dynamics before December 2019. To facilitate accurate future trend predictions, the study will employ an array of statistical models to compare their prediction, including AR(1), ARIMA, SARIMA, hybrid ARIMA-ANN, and hybrid SARIMA-LSTM for Time Series Forecasting. The hybrid models seem particularly promising in this endeavor.

Introduction

As the COVID-19 pandemic unfolded, it brought deep changes across the globe that are yet to be studied and fully understood. Among those assessing the impact of the health crisis on demography can be interesting and non-trivial. Indeed, the official death toll associated with the disease is just shy of 7 million worldwide as of November 2023 [WHO, 2023]. The hypothesis that this constitutes by itself an important boost to mortality leading to lower population level with respect to 'potential' seems natural but it is too simplistic as the pandemic and the public policies installed to affront it generated broader changes. Top of mind comes the fact that lockdown policies could contribute to modifying mortality rates from other sources. For instance, since broad parts of societies where this policy took place didn't leave their homes, it is plausible that deaths due to violence or accidents didn't occur. However, it is also plausible to see hikes in the deaths from diseases that were not screened and treated on time (e.g. cancer) as people avoided getting into hospital settings. In the same spirit, the effect of the crisis on natality and migration is non-obvious as fertility and migrating decisions are affected by uncertainty.

Leveraging the fact that demography has long been studied using time series analysis [Saboia, 1974, being a seminal work in the subject] we decided to run a variety of univariate models to have a first approximation of the realized effect of the pandemic on population level in different geographies. We use monthly data from the first two decades of the century to get forecasts of the population level of five countries and the world. We proxy the counterfactual, natural population levels by these forecasts and attribute the difference in the realized population figures to the COVID-19 effect. This is far from a rigorous, causal analysis of the question but can help us get a first idea of the direction of the effect (if any) of a major public health crisis on demographic variables that can have crucial effects on societies in the long term (think for instance on growth and productivity questions). It is also to be noted that since we are working only with population data we are not unbundling the contributions of the three determinants of population changes (mortality, fertility, and migration). It is also to be noted that by focusing on higher frequency data from a reduced number of years we are ensuring to have enough data points to be able to get decent estimates but we avoid caching secular changes in the demographic dynamics.

1 Literature Review

As recommended, we display the results of our literature review in the following tabular.

Study	Journal CNRS rating and citations	Econometric technique used	Aim of study	Data Set	Advantages of methodology	Limitations methodology	Results if relevant
Forecasting Mortality: A Parameterized Time Series Approach - McNown and Rogers [1989]	Demography, rang 1 CNRS. 196 citations (Google Scholar)	ARIMA, Parameterized model schedules and TS methods	Mortality dynamics, Population forecast	Forecast mortality rate	Data from US Social Security Administration	TS: flexible modeling of trend dynamics to accommodate LT growth rates. ARIMA: good approximation of the observed series and result in good forecasts compared to older methods (e.g. large simultaneous equation models) + allows to have some theoretical foundation behind the closed-form expressions of the model.	Focus on mortality series but use simple TS methodology we might want to explore. They couldn't integrate mortality probability's non-linearities in their study so the forecast ER aren't correct.
European demographic forecasts have not become more accurate over the past 25 years Keilman [2008]	Population and development review, rang 2 CNRS. 89 citations (Google Scholar)	Regression model with the error in the total fertility rate, life expectancy, and scaled net migration as the dependent variables.	Flexible platform for forecasting populations, Forecast age patterns of mortality and fertility.	National statistical agencies data (1950-2004)	Widely used by national statistical agencies. Transparent and easy to understand. Allows for the decomposition of population change into its 3 components: fertility, mortality and migration.	Assumes past trends will continue into the future. Relies on assumptions about future fertility, mortality and migration patterns which are subject to uncertainty.	Accuracy of demographic forecast has not improved in the last 25y. Forecasts tend to be too optimistic in particular for fertility and migration. Factors that influence the accuracy of the demographic projections include forecast horizon, forecast variant and the level of uncertainty of the underlying data.
Probabilistic Approaches to Population Forecasting. Lee [1998]	Population and development review, rang 2 CNRS 282 citations (Google Scholar)	Presents methods used in probabilistic approaches to population forecasting.	Presents methods used in the literature regarding population forecasting, addressing common uncertainty issues that arise.	Since it is a revision of the methods, he addresses the main problems one can encounter with the data selection process and the limited historical data quality availability.	Mention of various methodologies, ex-post evaluation, aggregate time series, expert judgment (takes into account opinions to generate random scenarios)	Existing challenges for capturing and representing uncertainties. Limited historical data quality, autocorrelation assumptions, and complexity of model specification.	Traditional methods rely mostly on scenario-based methods that lead to inconsistencies and misrepresentation.
An Econometric Model for Forecasting Regional Population Growth Plaut [1981]	International Regional Science Review rang 2 CNRS 71 citations (Google Scholar)	Proposes six models: econometric, constant growth rate, ARIMA (0, 2, 1), ARIMA (1, 1, 0), Glickman model.	Propose different models to forecast regional population growth for Texas, USA.	Data from Texas, USA, from 1960 to 1978, for the econometric models the author has data on wages, employment, and income per capita	With the use of the different methodologies, the author mentions that having more models is better for the accuracy of the prediction of the future population in a region.	Even if the econometric model seems to be useful for the case of Texas, further analysis would be needed to assess its replication in other parts of the US (external validity).	Note: Since the author is focusing in the prediction of population growth for a State of the United States of America which is heavily impacted by migration, it is understandable the different models he proposed, taking into account also labor immigrants which is something we are not doing.
Modeling and Forecasting Populations by Time Series: The Swedish Case Saboia [1974]	Demography, rang 1 CNRS 89 citations (Google Scholar)	Employs autoregressive (AR) and moving average (MA) models and Box and Jenkins' time series analysis to capture population dynamics.	Forecast Swedish population using AR and MA models, Compare results with other forecasting methods.	Mid-year population of Sweden at five-year intervals from 1780 to 1970, their source is Keyfitz and Flieger (1971) and the Demographic Yearbook (United Nations, 1971)	Time series forecasts can be given by normal distributions and confidence intervals can be easily obtained	A limitation that is encountered in the time series analysis techniques is the estimation of the parameters, at least 50 observations of the time series are needed.	More research is needed to understand the meaning of the parameters of a time series population model.

2 Empirical Strategy

2.1 Data Collection

We collected monthly demographic data for five countries: France, Japan, the USA, Sweden, and Colombia, as well as for the world. The data set covers several years before March 2020, and it extends until August or September 2023, allowing us to train our model with sufficient data points even if the period study is less than two decades.

To collect monthly data for the population in order to take into account the seasonality, we went to official governmental statistics websites for France, the USA, and Sweden (see References). For Japan, we found the data on the website ≪Statistic Dashboard≫ which gives data provided by the national government and private companies. Although we had difficulties finding the last country, we finally managed to find monthly data for a 5th country, Colombia, as well as for the world population, thanks to Factset (a data provider), in .xlsx format.

After the files had been collected (4 csv and 2 xlsx), we proceeded to the cleaning. We noticed that the date format from the csv was widely different from one file to another, making the cleaning and formatting tenuous and time-consuming via Python. We preferred to standardize the date format throughout all files manually. We took this occasion to convert all csv files to an xlsx format and remove useless columns to make the data easier to import and manipulate later on. In the end, there was only the date and population for each country.

After importing the files in Python, we performed some operations to clean the data (transform into a data frame, transpose some columns, delete unnecessary rows, and convert the population into millions for all countries...). Also, because some countries have data for different dates, we have sliced the data to have common dates for all countries. These data take into account population changes from 2001-01-01 to 2023-06-01.

2.2 Methodology

To conduct a thorough analysis of population forecasting, various models such as AR, MA, ARIMA, SARIMA, and Hybrid ARIMA (combining ARIMA and ANN models) are used. This study assesses the effectiveness of these models in various demographic scenarios, both globally and at a national level.

In 1974, Saboia [1974] harnessed the power of AutoRegressive (AR) and Moving Average (MA) models, enhancing their predictive capabilities with Box-Jenkins' time series analysis techniques, to effectively forecast population trends in Sweden. McNown and Rogers [1989] provided compelling evidence for the effectiveness of AutoRegressive Integrated Moving Average (ARIMA) models in mortality forecasting, leveraging parameterized time series approaches to yield accurate results.

Zhang [2003] work inspired the creation of the Hybrid ARIMA and ANN model, aiming to enhance accuracy by capturing both linear (via ARIMA) and non-linear (via ANN) components of time series data.

2.3 Comparative Analysis

Regarding the **Hybrid ARIMA** model's performance, it shows promising low mean squared error values, notably zero for Japan, the USA, Colombia, Sweden, and the world. However, this model displays fluctuating forecasts, encompassing a blend of both underestimation and overestimation trends. For instance, Japan's predictions gradually declined from 4.839 to 4.831, while France depicted a broader range from 4.178 to 4.117. Similarly, the USA's predictions ranged between 5.804 and 5.821. During the test phase, Japan's errors fluctuated between approximately -0.006 and -0.003, while France showed a range of around 0.07 to 0.08, and Colombia ranged between 0.011 and 0.013.

Comparatively, the **SARIMA-LSTM** model revealed a rather consistent but varied performance when forecasting COVID-19 cases. For Japan, Colombia, and Sweden, the model performs reasonably well with low Mean Absolute Percentage Error values: 15.24%, 13.35%, and 7.98% respectively, the predictions closely align with the actual data for these countries. The Mean Squared Error, Mean Absolute Error, and Root Mean Squared Error are reported as zero for all countries which could indicate a possible overfitting to the training data.

The AR(1) model exhibits moderate predictive performance in forecasting population gaps across various countries. While displaying moderate mean squared error (MSE) values, it maintains some bias in its estimates for countries like France, Colombia, and Sweden, ranging from -0.017 to 0.007. However, its mean absolute percentage error (MAPE) values fall between 0.026 and 0.441, demonstrating significant variation in accuracy across different nations. Notably, the RMSE values span between 0.0004 and 0.0202, indicating a relatively diverse margin of error in forecasting.

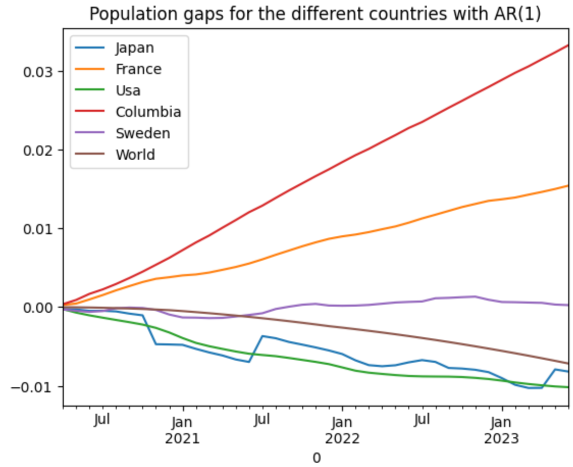
Comparatively, the ARIMA model presents more varied results with lower errors in MSE and RMSE across countries like Japan, the USA, and Sweden. It showcases a wider variation in bias across regions, with notable underestimations in Colombia and overestimations in France. The MAPE values are notably lower than the AR(1) model, ranging between 0.027 and 0.42, indicating relatively improved accuracy in its predictions.

The SARIMA model shows a mixed performance, presenting higher MSE, RMSE, and MAPE values across countries like Japan and Colombia. However, it displays consistency in bias, ranging from 0.001 to 0.014, with varying underestimation or overestimation tendencies in different regions. This model shows more uniformity in the errors across different countries, indicating a balanced yet slightly less accurate predictive performance compared to the ARIMA model, especially in countries like Japan and Colombia.

3 Results

3.1 Basic AR(1) model

This is the most simple type of time series model and relies on the possibility of forecasting the population at time t with the realization of the population and $t-1$. In this case, we automatically control for seasonality and differentiate the population series d times until they are stationary (per the results of an ADF test) and check the number of relevant lags to consider capturing in the model. For all of the geographies g we find that the



Summary

Figure 1: Auto-regressive models, order 1
Note that the measure in the y-axis is the log of the population level for each geography in millions. Since the log function is a monotonic and positive transformation, the intuitive interpretation of the graphs to see the population gap remains unchanged.

previous month's population is sufficient to perform estimations of the form:

$$\forall g,\ \Delta^{d_g} logPop_{g,t} = \psi_g \Delta^{d-1_g} logPop_{g,t-1} + \varepsilon_{g,t},\ \varepsilon_g \sim WN$$

Importantly, this estimation implies that we consider the underlying mechanisms driving the demographic dynamics of a given country to be unchanged between 2001 and 2019 which can be plausible as they are very slow-moving variables. We also avoid any questioning on the quality and precision of the data we are using. Finally, we are not considering any interdependencies between all the countries (in particular, this wouldn't allow any migration mechanism): the estimation for each is independent of the others. After integrating back the series, we get the following forecast outcome that we can compare with the data points we collected for the 2020-2023 period (Figure 1). In our opinion, the most interesting result from this first approximation is that the effect of the pandemic on population seems to be heterogeneous across geographies: while the effect is negative and rather large in Japan, France and Colombia seem to have a larger population than expected after COVID-19. The result in Sweden and across the world seems to be rather neutral. Regarding performance, note that this model performs best on the world's data set using RMSE and MAPE criteria, and for Colombia under a lowest bias criterion.

3.2 ARIMA model

Mathematically, an ARIMA model can be represented as:

$$(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)(1 - B)^d Y_t = c + (1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q)\epsilon_t$$

where Y_t denotes the observed values, p represents the autoregressive order, d represents the differencing order, q represents the moving average order, ϕ and θ are parameters to be estimated, B is the backshift operator, ϵ_t represents the error term, and 'c' is a constant term.

The ARIMA (AutoRegressive Integrated Moving Average) models were employed to assess the population gap during the COVID-19 pandemic, focusing on the difference between predicted and observed values. These models are represented as ARIMA(p, d, q), where p denotes the autoregressive component, d signifies differencing and q represents the moving average component. The best-fit models for different countries had varying orders (in particular, seasonality was not achieved at the same d for all countries). For instance, the model for 'Japan' was determined as ARIMA(2, 2, 25), while 'France' had ARIMA(2, 0, 36), 'USA' with ARIMA(2, 2, 39), 'Colombia' with ARIMA(2, 1, 35), 'Sweden' with ARIMA(2, 2, 40), and 'World' with ARIMA(2, 3, 38). The d parameter is particularly essential as it represents the differencing order required to stationarize the series, which is fundamental in assessing the population gap over time.

The ARIMA models, with orders (p, d, q) determined for each country, produced predictions for the population gap during the COVID-19 pandemic. Key performance figures show that 'Sweden' had the most accurate predictions with an MSE of 6.65e-07 and a MAE of 0.000649, indicating a minimal gap between predicted and observed values. On the other hand, 'France' had a higher MSE of 3.75e-04 and MAE of 0.015733, suggesting a less accurate prediction.

3.3 SARIMA

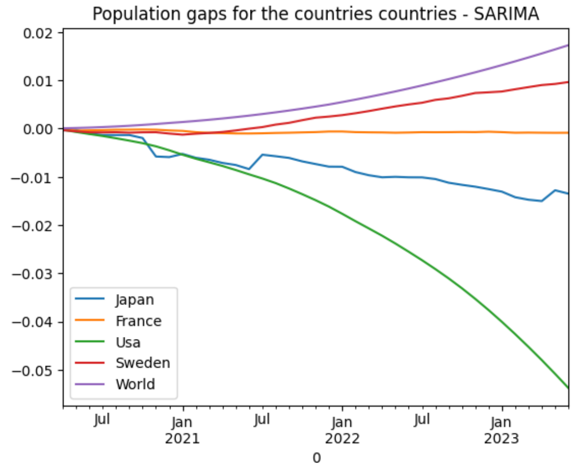
We also implemented this variation of an ARIMA model, automatizing the choice of the best parameters with Python, to better capture seasonal changes in the population dynamics. The results are presented graphically in Figure 2. While we still have heterogeneous effects that might change signs, the results are not the same as in previous models, highlighting that our results might not be robust.

3.4 Hybrid Models

The hybrid model features an Artificial Neural Network (ANN) capturing the non-linear component by modeling the residuals (e_t). These residuals, e_t, are fed into the ANN to generate the non-linear component in the form of a function, f, defined by the neural network, and accounting for the random error, ε_t, at each time t. The combined forecast of this hybrid model is derived as the sum of the ARIMA model forecast $(\hat{L}_t)$ and the ANN model forecast for the residuals $(\hat{N}_t)$. The equation for the hybrid model's combined forecast is:

$$\text{Hybrid Model Forecast} = \hat{L}_t + \hat{N}_t$$

The hybrid ARIMA model shows fluctuating forecast ranges for countries like Japan, France, and the USA. This model's predictions vary between underestimation, ranging from -0.0002 to 0.004, and overestimation. For instance, it consistently underestimates figures by 0.0006 to 0.004 across different time points. Conversely, the SARIMA-LSTM model consistently exhibits a downward trajectory in its forecasts, maintaining a steady under-



Summary

Figure 2: SARIMA models and gaps
Note that the measure in the y-axis is the log of the population level for each geography in millions. Since the log function is a monotonic and positive transformation, the intuitive interpretation of the graphs to see the population gap remains unchanged.

estimation trend, averaging between 0.0006 to 0.004 compared to actual figures. The SARIMA-LSTM's uniform underestimation leads to a consistent downward slope in the forecasted population gaps, whereas the hybrid ARIMA model demonstrates a wider and less predictable forecast range due to its varying overestimation and underestimation patterns. Specifically, the SARIMA-LSTM consistently underestimates figures by 0.0006 to 0.004, while the Hybrid ARIMA model displays a wider range from -0.0002 to 0.004, indicating a less stable prediction pattern.

Looking at the training errors of the models, the SARIMA model's population gap predictions consistently show a trend of underestimation, with figures ranging from 0.0006 to 0.004 below the actual population gap. Meanwhile, the Hybrid ARIMA model presents a broader range of errors, fluctuating between -0.0002 to 0.004, showcasing a mix of overestimation and underestimation. The SARIMA-LSTM's consistent underestimation leads to a steady decline in the forecasted population gaps, while the Hybrid ARIMA model's wider error range suggests a less predictable and more volatile pattern, with a mix of overestimation and underestimation.

The SARIMA-LSTM's uniform underestimation trend results in a gradual downward slope in the forecasted population gaps. Conversely, the Hybrid ARIMA model delivers a more volatile forecast, displaying a wider variance and indicating an unpredictable mix of overestimation and underestimation. Specifically, the SARIMA-LSTM consistently underestimates figures by 0.0006 to 0.004, while the Hybrid ARIMA model's wider range from -0.0002 to 0.004 suggests a less stable prediction pattern.

3.5 Model perfomance

In order to find the best-performing model, using Tables 1, 2, and 3 present the best-performing models for each country based on the values of root mean squared error (RMSE), mean squared error (MSE), and bias,

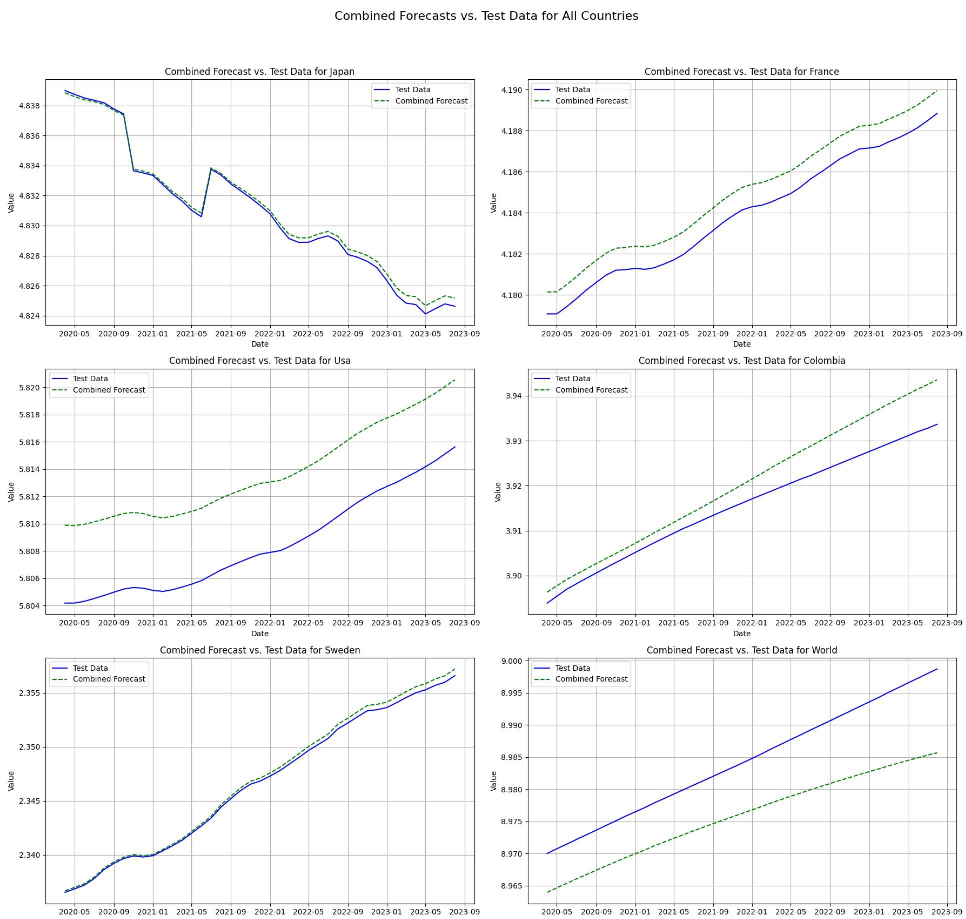


Figure 3: SARIMA-LSTM model forecasts

respectively. The selected models and their corresponding metrics offer valuable insights into the performance of each model for the specific context of each country.

Country	Best Model	Lowest RMSE
Colombia	SARIMA-LSTM	0.005631
France	SARIMA-LSTM	0.001097
Japan	SARIMA-LSTM	0.000303
Sweden	SARIMA-LSTM	0.000352
Usa	SARIMA-LSTM	0.005243
World	AR(1)	0.000246

Table 1: Best model and its corresponding lowest RMSE for each country.

Country	Best Model	Lowest Bias
Colombia	AR(1)	-0.017314
France	Hybrid ANN	-0.035729
Japan	Hybrid ANN	-0.002040
Sweden	Hybrid ANN	-0.003207
Usa	Hybrid ANN	-0.006157
World	SARIMA-LSTM	-0.008461

Table 2: Best model and its corresponding lowest Bias for each country.

Country	Best Model	Lowest MAPE
Colombia	SARIMA-LSTM	0.125410
France	SARIMA-LSTM	0.026207
Japan	SARIMA-LSTM	0.005413
Sweden	SARIMA-LSTM	0.013149
Usa	SARIMA-LSTM	0.090173
World	AR(1)	0.002284

Table 3: Best model and its corresponding lowest MAPE for each country.

4 Discussion and Conclusion

In our view, this exercise can only be viewed as a very preliminary approximation to the question of the pandemic's effect on demographic data, not only because of the very narrow geographic scope of our analysis but because it is too broad. Indeed, we do not examine the components of demographic dynamics that might as well have opposite effects for different countries and require more variables to properly account for the effect of the pandemic. Namely, using macroeconomic and political data can help us understand migration flows that can heavily affect the demographic dynamics of a given country. In a similar fashion, desegregation between genders and age groups can also give us a better understanding of the differential effects of the crisis (e.g. if drops in population are mainly due to higher risk for the elderly) and its implications for the demographic future of the countries. Also, the data treatment left us dubious of its quality: we need to note that most the population data relies on estimations made between censuses that happen with several years of difference and can be in fine misleading and carry noise to the analysis[1]. It is possible that we must wait several years to have the definitive data for the early 2020's.

However, the comparison of several univariate models has the advantage that it can give us a good idea if there is any effect to uncover and in which direction, and if there seems to be any robust relationship. While our results seem to not be very robust for many of the countries we analyze it gives us some hints on the challenges of this endeavor. In particular, note that the sign of the population gap switches between models for most of the countries but for Japan, signaling that the effects might be clearer in this economy that is already aging very fast.

The SARIMA-LSTM model appears as a strong contender in predicting population gaps across diverse geographies, showcasing promising precision and low error metrics such as Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). While it displays remarkable accuracy in countries like France, Japan, Sweden, and the USA, the consistently low error rates also raise a concern about potential over fitting to the training data, which might limit its applicability to new, unseen data.

[1] See for example this recent case in Colombia here. With the 2018 census, the Government realized that their estimation of the total population used was 9% representing almost 5 million people higher than in reality.

Team Contributions

Writing, Literature Review, Model Selection, Model Calibration, Population Gap Analysis, Data Collection

Team Member Name	**Contribution**	**Level of Contribution**
Charaf Zguiouar	Building and calibrating the models, implementing the econometric models and the hybrid ones.	High+
Francisco Javier Alvarez	Writing, Literature Review, Model Calibration, Population Gap Analysis	Medium
Lucas Rouleau	Writing, Data collection, Data cleaning	Medium
Ludovic Paul	Writing, Data Collection, Data Cleaning and Analysis, Structure of project	Medium
Maysaa Rais	Writing, Literature Review, Model Calibration, Population Gap Analysis	High
Natalia Cardenas Frías	Writing, Literature Review, Population Gap Analysis, Model testing	High

Table 4: Team Contributions

References

WHO Coronavirus (COVID-19) Dashboard. https://covid19.who.int, 2023.

Nico Keilman. European demographic forecasts have not become more accurate over the past 25 years. *Population and development review*, 34(1):137–153, 2008.

Ronald D Lee. Probabilistic approaches to population forecasting. *Population and Development Review*, 24:156–190, 1998.

Robert McNown and Andrei Rogers. Forecasting mortality: A parameterized time series approach. *Demography*, 26(4):645–660, 1989.

Thomas R Plaut. An econometric model for forecasting regional population growth. *International Regional Science Review*, 6(1):53–70, 1981.

Joao LM Saboia. Modeling and forecasting populations by time series: The swedish case. *Demography*, 11(3): 483–492, 1974.

G Peter Zhang. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50: 159–175, 2003.

Sources for Data collection: Population France monthly (1975-2023): https://www.insee.fr/fr/statistiques/serie/000436387

Population USA monthly (1959-2023): https://fred.stlouisfed.org/series/POPTHM

Population Japan monthly (1975-2023): https://dashboard.e-stat.go.jp/en/timeSeriesResult?indicatorCode=020101000000001

Population Sweden monthly (2000-2023): https://www.statistikdatabasen.scb.se/pxweb/en/ssd/

Colombia world : Factset

Appendix

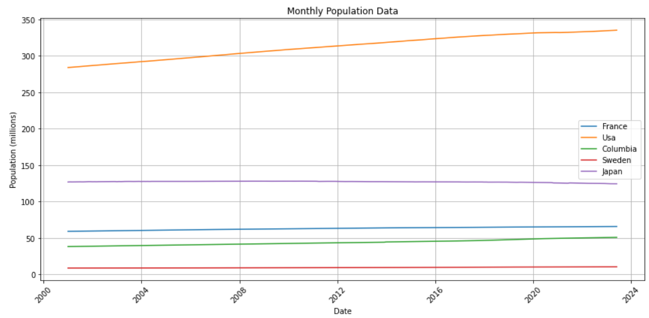


Figure 4: Evolution of the population of the 5 countries between 2001-01-01 and 2023-06-01

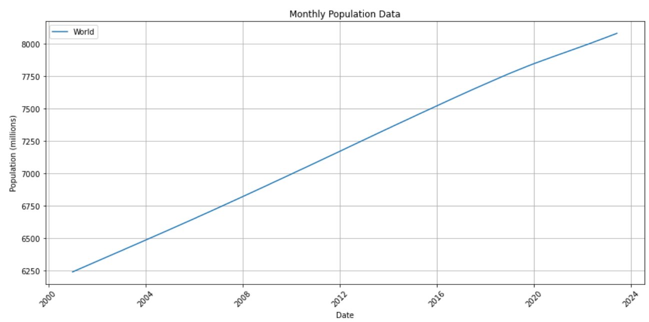


Figure 5: Evolution of the world population

Table 5: Descriptive statistics

	Japan	France	USA	Colombia	Sweden	World
Count	272.0	272.0	272.0	272.0	272.0	272.0
Mean	127.15	63.16	312.92	44.14	9.61	7188.29
Std	0.91	1.94	15.66	3.66	0.54	553.25
Min	124.48	59.27	283.96	38.38	8.88	6237.51
25%	126.87	61.65	299.38	41.10	9.10	6702.95
50%	127.40	63.44	314.27	43.68	9.50	7194.91
75%	127.79	64.84	327.88	46.76	10.12	7685.47
Max	128.10	65.95	335.50	51.09	10.55	8092.63

Table 6: Descriptive statistics of the monthly growth rate of the population

	Japan	France	USA	Colombia	Sweden	World
Count	271.0	271.0	271.0	271.0	271.0	271.0
Mean	-6.95E-05	0.000394	0.000616	0.001057	0.000636	0.000961
Std	0.000534	0.000166	0.000200	0.000545	0.000324	0.000126
Min	-0.00378	-4.58E-05	-0.000166	-8.62E-06	-9.15E-05	0.000691
25%	-0.000323	0.000290	0.000520	0.000853	0.000424	0.000896
50%	-5.47E-05	0.000398	0.000665	0.000933	0.000595	0.001029
75%	0.000227	0.000515	0.000753	0.001050	0.000841	0.001055
Max	0.00317	0.000741	0.000916	0.008390	0.001791	0.001086

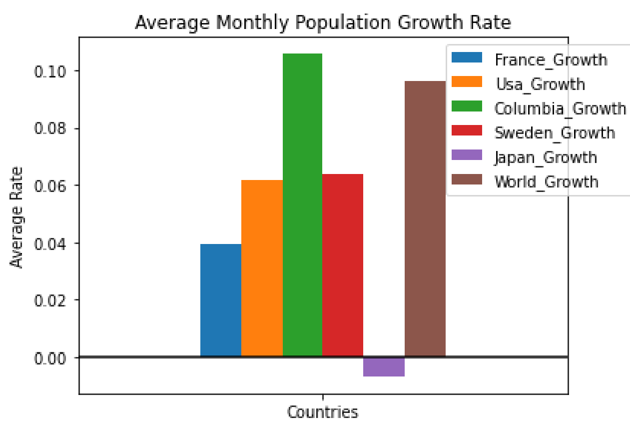


Figure 6: Average monthly population growth rate

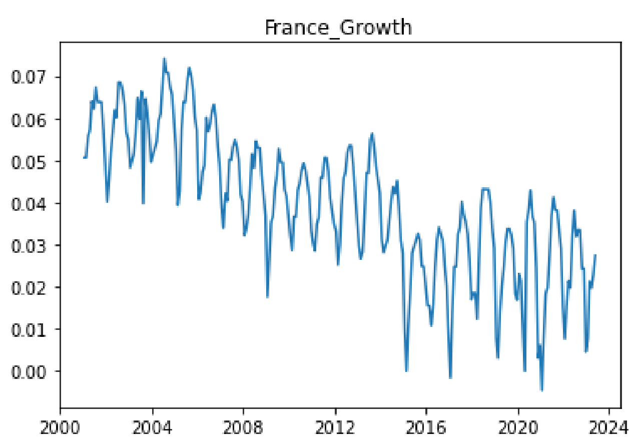


Figure 7: France's population growth rate over time

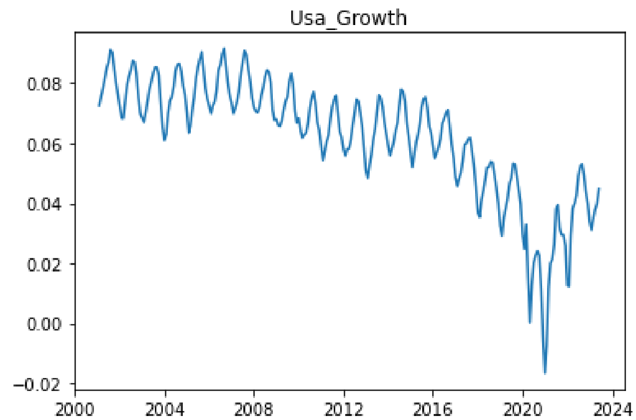


Figure 8: USA's population growth rate over time

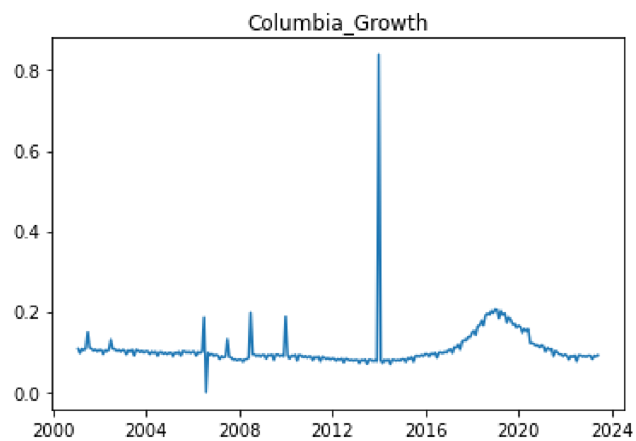


Figure 9: Columbia's population growth rate over time

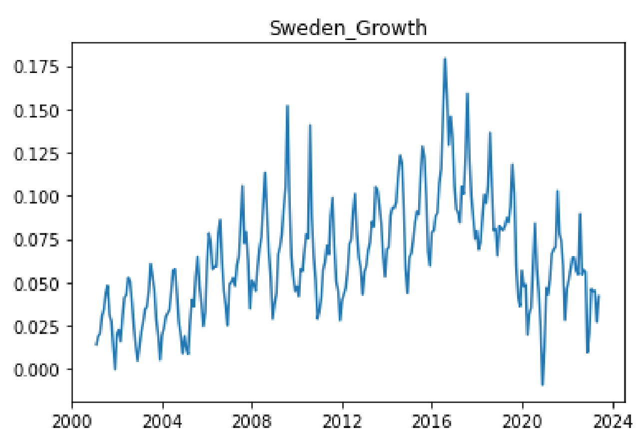


Figure 10: Sweden's population growth rate over time

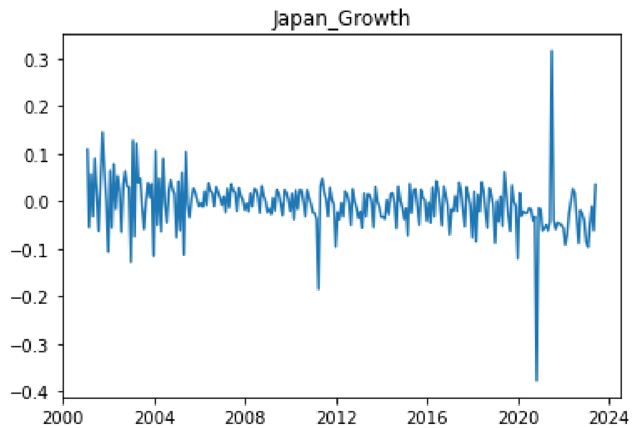


Figure 11: Japan's population growth rate over time

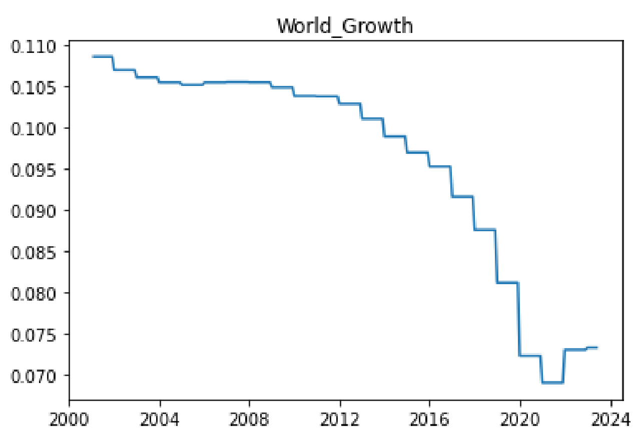


Figure 12: World's population growth rate over time

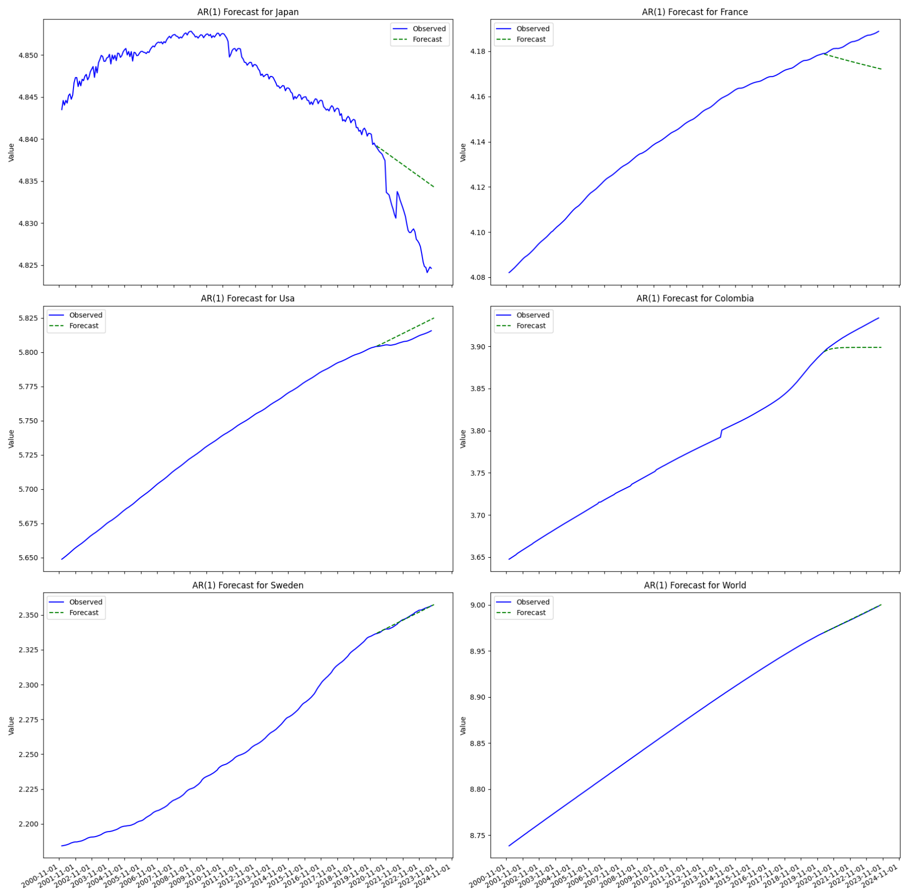


Figure 13: AR(1) Model Forecasts

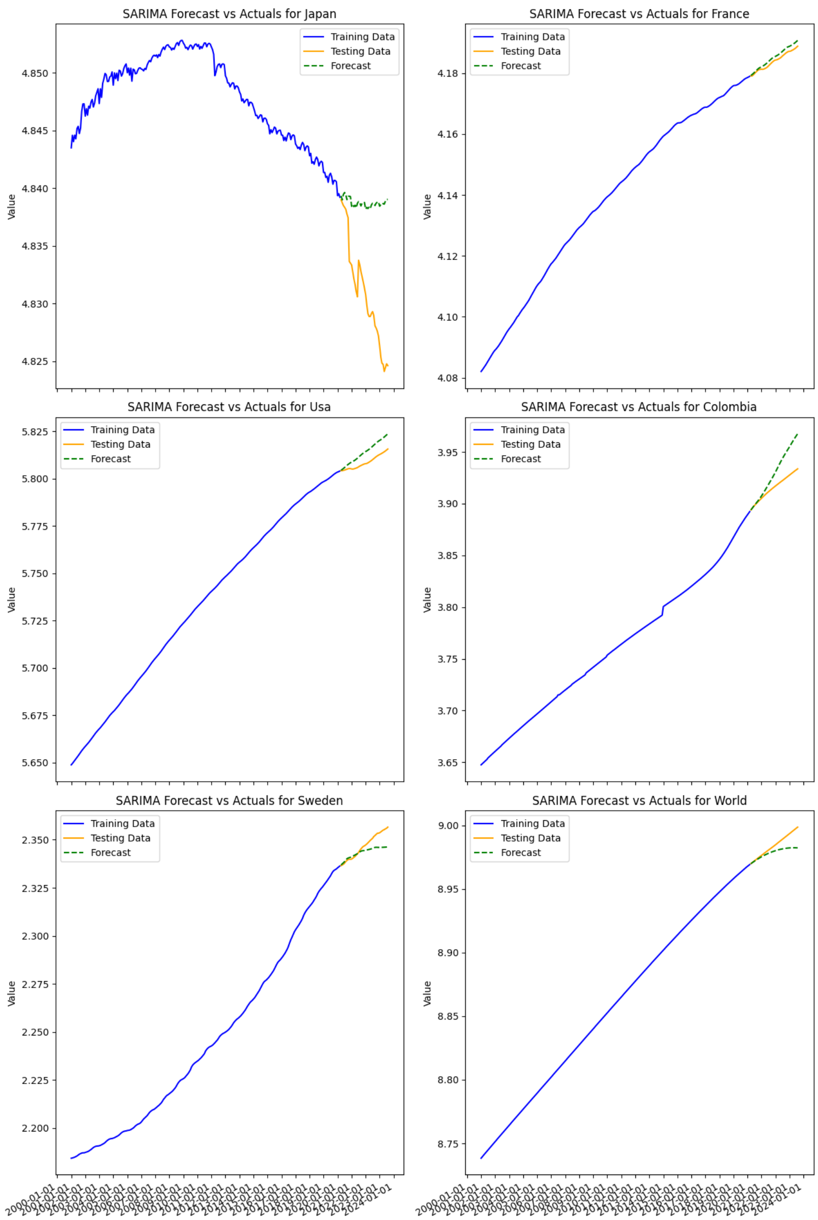


Figure 14: SARIMA Model Forecasts

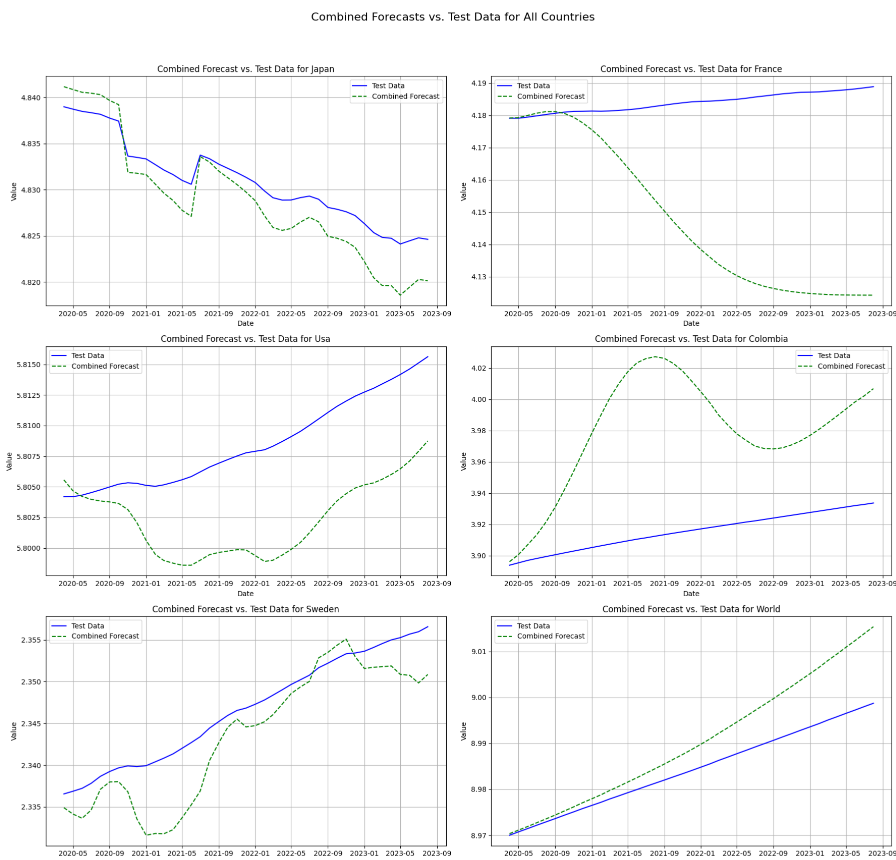


Figure 15: Hybrid ARIMA Model Forecasts

Table 7: Mean Squared Error (MSE) for different models

Country	Model				
	AR(1)	ARIMA	Hybrid ANN	SARIMA	SARIMA-LSTM
Colombia	4.096837e-04	2.031006e-04	0.005064	0.000329	3.170805e-05
France	9.278912e-05	3.746904e-04	0.001912	0.000002	1.202473e-06
Japan	4.475346e-05	2.405598e-05	0.000009	0.000079	9.175799e-08
Sweden	5.561803e-07	6.650375e-07	0.000019	0.000025	1.241455e-07
Usa	3.761557e-05	4.413014e-05	0.000047	0.000034	2.748419e-05
World	6.031690e-08	5.756466e-05	0.000061	0.000057	7.608202e-05

Table 8: Bias for different models

Country	Model				
	AR(1)	ARIMA	Hybrid ANN	SARIMA	SARIMA-LSTM
Colombia	-0.017314	-0.012667	0.064160	0.014094	0.004918
France	-0.008374	0.015733	-0.035729	0.001259	0.001096
Japan	0.005947	0.004343	-0.002040	0.007847	0.000223
Sweden	-0.000025	0.000581	-0.003207	-0.003181	0.000309
Usa	0.005594	0.006085	-0.006157	0.005375	0.005238
World	0.000205	-0.005721	0.006005	-0.005676	-0.008461

Table 9: Mean Absolute Percentage Error (MAPE) for different models

Country	Model				
	AR(1)	ARIMA	Hybrid ANN	SARIMA	SARIMA-LSTM
Colombia	0.441430	0.323454	1.638205	0.359231	0.125410
France	0.200061	0.376077	0.857100	0.030073	0.026207
Japan	0.123156	0.089950	0.056691	0.162508	0.005413
Sweden	0.026116	0.027662	0.148795	0.158866	0.013149
Usa	0.096287	0.104740	0.107544	0.092515	0.090173
World	0.002284	0.063629	0.066784	0.063131	0.094158

Table 10: Mean Absolute Error (MAE) for different models

Country	Model				
	AR(1)	ARIMA	Hybrid ANN	SARIMA	SARIMA-LSTM
Colombia	0.017314	0.012667	0.064160	0.014098	0.004918
France	0.008374	0.015733	0.035877	0.001259	0.001096
Japan	0.005947	0.004343	0.002738	0.007847	0.000261
Sweden	0.000613	0.000649	0.003490	0.003736	0.000309
Usa	0.005594	0.006085	0.006248	0.005375	0.005238
World	0.000205	0.005721	0.006005	0.005676	0.008461