Building Predictive Models to Forecast Population Trends in Ireland for 2017 - 2021

MSc Research Project
Data Analytics

Tom Fitzgerald
x16103271

School of Computing
National College of Ireland

Supervisor: Vikas Sahni

National College of Ireland
Project Submission Sheet – 2017/2018
School of Computing

Student Name:	Tom Fitzgerald
Student ID:	x16103271
Programme:	Data Analytics
Year:	2018
Module:	MSc Research Project
Lecturer:	Vikas Sahni
Submission Due Date:	13/08/2018
Project Title:	Building Predictive Models to Forecast Population Trends in Ireland for 2017 - 2021
Word Count:	5,070

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature:	
Date:	15th September 2018

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
3. Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Building Predictive Models to Forecast Population Trends in Ireland for 2017 - 2021

Tom Fitzgerald
x16103271
MSc Research Project in Data Analytics

13th August 2018

Abstract

This research explores how to improve the accuracy of population forecasts in Ireland. Population projections underpin strategic national policy across key areas including infrastructure, health, education and housing. Census data shows that net migration is a major driver of Ireland's population trends. It has also proven to be highly volatile and difficult to predict. This research identifies and gathers pertinent official economic, financial, migration and demographic data, and using a Principal Component Analysis, establishes fifteen key determinants for net migration. These are used as inputs for an ARIMA forecasting model, which when tested against actual historic data, outperforms previous official predictions. This model is then extended to forecast net migration for the five-year period 2017 to 2021.

1 Introduction

Population projections underpin key strategic policy objectives for Ireland's future allocation of resources. Ireland's high level strategic national planning framework, Project Ireland 2040, sets out to shape future growth and development to the year 2040 based on official population projections, Dept. of Housing Planning & Local Government (2018). Likewise, the National Development Plan 2018 - 2027, Government of Ireland (2018) sets out the investment priorities of € 116 billion over the next decade based on these population projections. This highlights the necessity for accuracy in forecasting future populations projections.

The background and literature review section of this paper examine recent population trends within Ireland. This research identifies net migration as a major component behind recent population change in Ireland and how a failure to accurately predict net migration trends has undermined previous population projections, Timoney (2012). An exploration of contemporary international research in this area informs the identification of some of the key drivers behind net migration.

The methodology section outlines the approach taken to examine and forecast net migration in Ireland. A CRISP-DM methodolgy is employed. Official data within an Irish context is gathered for internationally established drivers of net migration.

Furthermore, this research identifies and collates data on pertinent new candidate predictor variables from official datasets across a range of economic, social and migratory factors. All variables identified are analysed using a Principal Component Analysis (PCA) factor reduction technique, which then informs a range of time series models.

The implementation section is in two stages.

Firstly, the PCA is performed upon the full set of predictor variables examined, which determines fifteen key variables.

These fifteen variable identified in the PCA are then used as input predictor variables for the target variable of net migration across a range of time series models.

In the evaluation section, the results of these models are compared against actual historic data for the period 2014 to 2016. This is done across a range of established statistical criteria for evaluating population and time series forecasts. This process clearly demonstrates an ARIMA (0,1,0) model as the best performing model. It is then extended to forecast future net migration within Ireland for the five-year period 2017 to 2021.

The conclusion discusses how this research addresses its aim of providing a more accurate means of predicting future population trends for net migration. It examines the inherent strengths and weaknesses of the approach used. Future areas and applications that may be explored are discussed including the possibilities and benefits to be gleamed from more granular examinations of net migration trends.

This paper and its findings were presented by the author at the European Network of Housing Research Conference 2018, in Uppsala, Sweden, as part of the Housing Market Dynamics workshop ,European Network for Housing Research (2018). The paper will be published on the conference website.

1.1 Background

Net migration has played a significant role in the annual population change within Ireland for the period from 1995 to 2016, as shown in Figure 1. This role was particularly acute during Irelands economic boom, when at its peak in 2007, it was responsible for nearly three quarters of the annual population change, Gilmartin (2012).

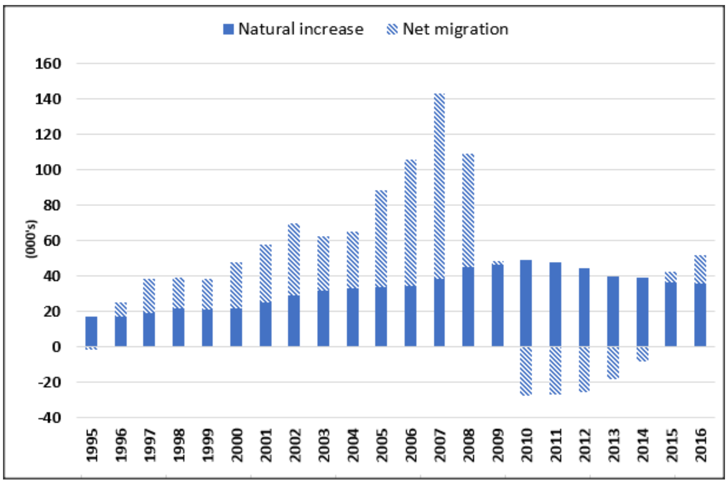

Figure 1: Annual Population Change by Component 1995 - 2016

Furthermore, not only is net migration a major component behind population change within Ireland, but it is acknowledged as the most volatile aspect, Shaw (2007). Unlike the components of natural population increase; fertility and mortality, it is less linked to population trends, and more aligned to economic factors. Whereas historical data on births and deaths are based on official registration systems, migration data entails a degree of estimation and hence have a greater inherent degree of error. In fact, the consistent historical poor estimation of population growth within Ireland has been due to a failure to predict the net migration component, Dignan (2010).

Table 1, shows how the actual net migration figures compare to the Irish Central Statistics Office (CSO) M1 scenario projections made in 2002, 2008 and 2013.

Year	Actual Net Migration (000's)	2002 Based M1 (000's)	2002 Based Difference	2008 Based M1 (000's)	2008 Based Difference	2013 Based M1 (000's)	2013 Based Difference
2002	41.3	30	-27%				
2003	30.7	30	-2%				
2004	32	30	-6%				
2005	55.1	30	-46%				
2006	71.8	30	-58%	60	-16%		
2007	104.8	30	-71%	60	-43%		
2008	64.3	30	-53%	60	-7%		
2009	1.6	30	1775%	60	3650%		
2010	-27.5	30	-209%	60	-318%		
2011	-27.4	30	-209%	50	-282%	-19.1	-30%
2012	-25.7	30	-217%	50	-295%	-19.1	-26%
2013	-18.7	30	-260%	50	-367%	-19.1	2%
2014	-8.5	30	-453%	50	-688%	-19.1	125%
2015	5.9	30	408%	50	747%	-19.1	-424%
2016	16.2	20	23%	40	147%	-19.1	-218%
Total	**315.9**	**440**	**39%**	**590**	**87%**	**-114.6**	**-136%**

Table 1: Actual Versus Official Predicted Net Migration, 2002 - 2016

The composition of net migration has also seen significant change, with migrants from the ten EU accession states in 2004 being the main driver of immigration during the economic boom up to 2007, Barrell et al. (2007). This trend has subsequently subsided. The volatile nature of net migration in Ireland over the period 1995 to 2016 is demonstrated in Figure 2 below.

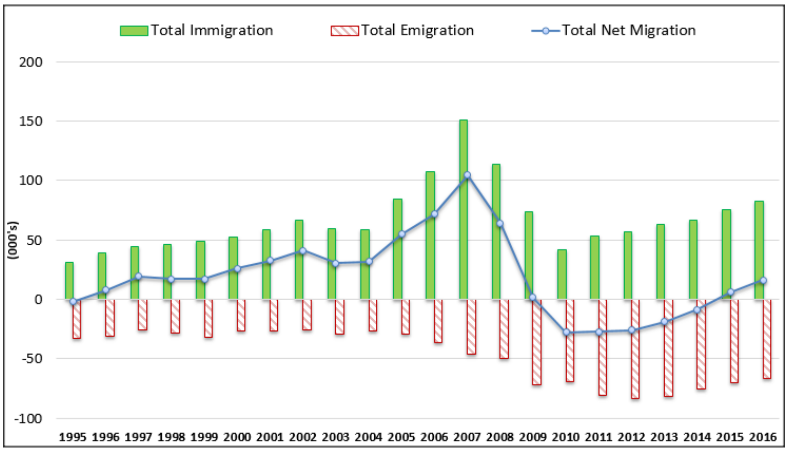

Figure 2: Immigration, Emigration and Net Migration in Ireland, 1995 - 2016

The time period of the graph contains the twin challenges in predicting net migration; shifting economic circumstances and changing trends, Dignan (2010).

It is therefore imperative for future planning to try and improve our ability to predict future net migration, and minimise the associated uncertainty. To do so requires an understanding of the main drivers behind it. This is a universal issue, and as such there is a significant body of research work conducted in this area, which is discussed in the next section, the literature review.

2 Review of Literature

Demographic projection models are a very important tool for informing policymakers plans in key areas, including education, infrastructure, health care and housing, Dignan (2010). As such, a great deal of research exists in this area, seeking to continuously improve the prediction efforts of such models. The United Nations is at the forefront of this research, and since 2014, they publish population projections for all nations to 2100 using the cohort component projection model, United Nations, Department of Economic and Social Affairs (2017). This model is also that which the CSO employ for their projections. Both Alkema et al. (2015) and Dignan (2010) examine the areas of uncertainty across the key components of the model.

The cohort component demographic projection model can be summarised as follows:

Projected Population = Current Population + births - deaths + net migration

Within Ireland, Census Data provides accurate data for the base population figure. In his study, Dignan (2010) examines population projection models produced by both UK and Irish government agencies, emphasising the uncertainty due to previous migration projections. In a study of fifty years of official UK population predictions, Shaw (2007) also examined this area, and notes that shorter term predictions are far less susceptible to uncertainty, particularly in relation to migration. This research has therefore limited its scope to net migration over a five year period, at which point new Census data will be available to corroborate its findings. In his analysis, Dignan (2010) identifies two main sources of uncertainty from past Irish models, changing trends and changing economic circumstances.

The use of multiple scenarios by the CSO acknowledges the inherent uncertainty involved in any attempt at population prediction. The CSO's latest population predictions for 2017 to 2051 (CSO, 2018) account for three future net migration scenarios; M1, M2, and M3. M1, the high net migration projection is for net inward migration of thirty thousand per annum over the period, with M2 at twenty thousand, and M3 at ten thousand, Central Statistics Office (2018).

In their study on migration within the EU, including Ireland, Moral-pajares and Jiménez-jiménez (2014) examine the economic pull factors which help determine the choice of destination for migrants. The pull factors examined are; GDP per head of population and the unemployment rate. A high GDP is shown to act as the greatest pull factor. Data for Ireland over the period, which includes the economic recession, shows the highest fluctuation in net migration of any of the fifteen countries studied, highlighting how exposed the Irish economy is to international economic turbulence. This fact is corroborated by Duffy et al. (2014) in their study which highlights the huge part migration has played as a component in populations trends within Ireland. This is particularly noticeable during periods of economic volatility, where migration has been the main factor driving population change within Ireland.

Serious consideration in this research is given to economic pull factors which are accepted among researchers as having a major role in migration trends, including Arnott and Chaves (2012). Indeed, Thomas Ng et al. (2008) attempt to predict population trends in Hong Kong based solely on ten economic indicators. In a small market driven economy like Ireland, recent history has shown that economic factors do play a major part is determining net migration.

3 Methodology

The aim of the project was to improve population forecasts. This is to be achieved by building a model to provide a more accurate means of predicting future net migration in Ireland for the five-year period 2017 to 2021, than current official estimates.

The CRISP-DM methodology was used as the overarching framework for research. This methodology is accepted as international best practice in conducting data mining and modelling studies, Wirth (2000). The process diagram for this approach is shown below in Figure 3.

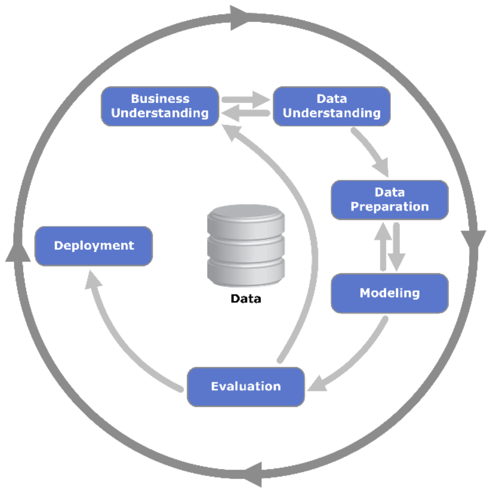

Figure 3: CRISP-DM Process Diagram

This methodology involves six stages which are iterative, thus enabling a continuous refinement of the model. The stages involved are: business understanding, data understanding, data preparation, modelling, evaluation and deployment.

The literature review forms the basis for the business understanding and initial data understanding stages, enabling insights into the drivers behind net migration in Ireland and elsewhere. This stage helped identify key datasets required for the project.

A further aspect of the data understanding stage and a key consideration of the study was the availability of historic data within Ireland. In attempting to identify the key indicator variables for net migration in Ireland, official data was gathered for a wide range of potential candidate variables, covering migratory, social and economic factors for the period 1995 to 2016.

This study collects data on key variables identified in previous international research within an Irish context. Migrant data sources, such as those identified in the U.K. by Lymperopoulou (2018) were sought for Ireland. In addition, an extensive search of official datasets allows for the inclusion and examination of new variables not previously considered but shown here to be beneficial.

New datasets examined as part of this research include data published by the CSO on the numbers employed across a range of diverse sectors within Ireland. This study examines data pertaining to the construction, agricultural, retail, manufacturing and service sectors. Such sectors were identified by Fitzgerald (2014), as of particular economic relevance during the period this research examines.

Key economic variables identified by previous international research are also examined including the unemployment rate, GDP, GNP and stock and bond market performance indicators. For the first time, a new Irish specific measure of national economic activity is considered; modified Gross National Income, which attempts to control for globalisation effects that are disproportionately impacting the measurement of the size of the Irish economy, Central Statistics Office (2017b).

Other new datasets incorporated into this study include official figures from the Department of Justice on the number of asylum applicants per year, and CSO data on the number of trips made by people arriving to and departing from Ireland to visit relatives. The latter proving to be indicative of net migration trends.

All datasets used were identified and gathered from official national and international organisations, including the CSO, the Irish Department of Justice, Eurostat, the OECD and the World Bank. Each dataset was extracted as a .csv file, and collated together in Excel.

An initial exploratory analysis of some of the variables identified in international research shows close relationships existing within an Irish context. This is shown in Figures 4 and 5 which explore the correlation between the unemployment rate and net migration within Ireland, and the employment rate of foreign nationals in key sectors and net migration.

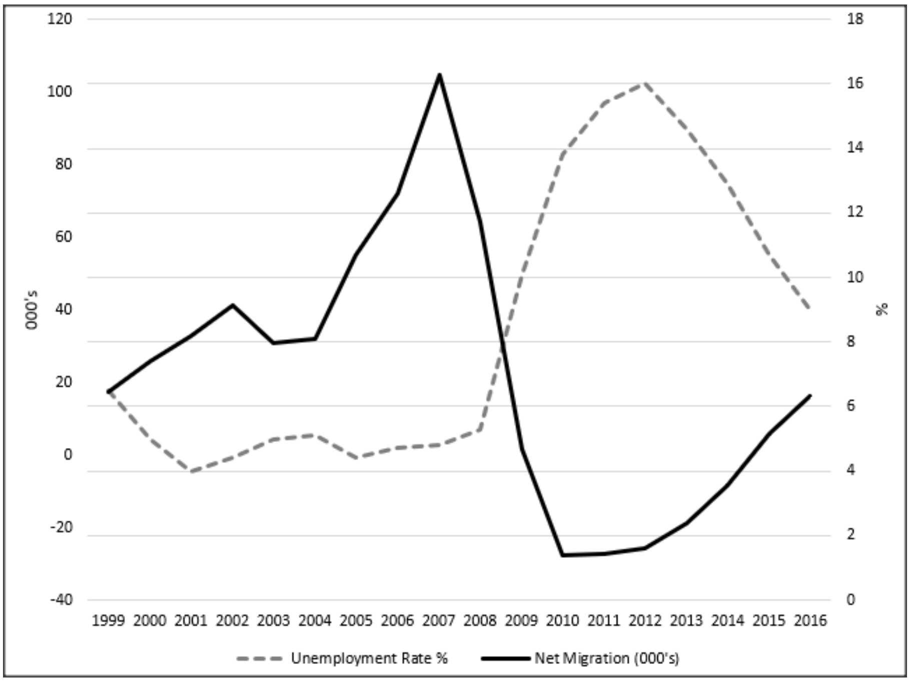

Figure 4: Net Migrations and Unemployment Rate, Ireland 1999 - 2016

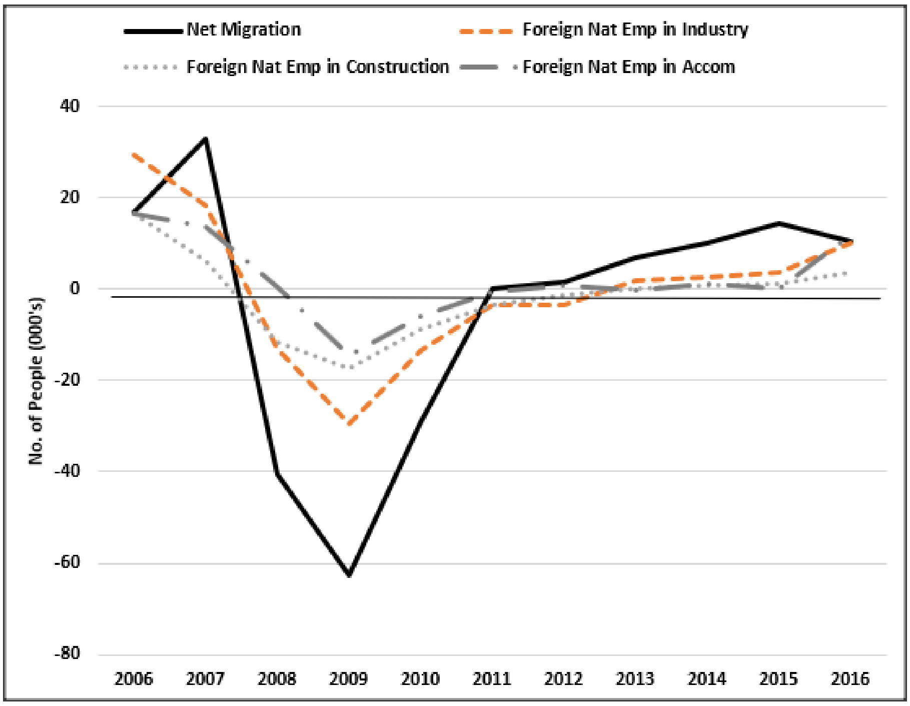

Figure 5: Yearly change in Net Migration and Foreign Nationals employed in key sectors

The modelling stage consisted of the combination of first undertaking a PCA on the potential predictor variables identified.

In total over thirty potential indicator variables were examined. Such a high number of variables can prove prohibitive in conducting any analysis and may undermine the successful application of data modelling, Donoho (2000). A recognised method employed to handle such a scenario is a PCA, which has the objective of reducing a data set to a more manageable size while retaining as much of the original information as possible, Constantin (2014).

The variables identified by the PCA, were used as the input predictor variable for the target variable net migration across a range of regression and time series models, including the ARIMA (Auto-Regressive Integrated Moving Average) model.

The ARIMA model, also known as the Box Jenkins Model, has been shown to be beneficial for population forecasting as they provide prediction intervals to accompany their point forecasts, Tayman et al. (2007).

An ARIMA model is classified as an "ARIMA (p,d,q)" model, where the three parameters are:

p: the number of autoregressive terms,

d: the number of nonseasonal differences needed for stationarity, and

q: the number of lagged forecast errors in the prediction equation.

In the implementation of this research, a PCA was used to reduce the number of variables examined, while an ARIMA (0,1,0) model was used to forecast future net migration.

4 Implementation

4.1 Principal Component Analysis

In preparing and analysing the data, and implementing the PCA factor reduction technique, IBM SPSS Statistics Version 24 was used.

The basic premise behind a PCA is the identification of clusters of variables called components that capture the most variance possible within the dataset, Wright et al. (2009).

The numeric nature of all variables under consideration meant that no variable was excluded as part of this analysis. This negated the inherent limitation of a PCA in handling non-numeric data.

To facilitate the PCA, the variables were scaled to parameters, as recommended by Wold et al. (1987). Datasets pertaining to number of persons were scaled to a base of 000's. Figures relating to national income indicator such as GNI and GDP were scaled to 000's Million Euro.

The underlying mechanics of a PCA involves a least squares method, therefore outliers can have a direct bearing on results (Wold, 1987). As such, outlier detection tests were performed across all variables. This involved an analysis of z-scores and boxplots for all variables, which are established best practice for univariate outlier detection, Seo and Gary M. Marsh (2006). A few instances of outliers were found for measures of national income. Given that one of the key aspects of this research was its use of actual official data for Ireland, these outliers were not removed.

The PCA was performed using a Varimax rotation method. This is the most popular rotation method and allows for a clear, easily interpreted solution as the components identified have a small number of large loadings, Williams (2010).

The initial step of the PCA involves identifying the number of components which need to be considered, an approach which is based on whether they have sufficiently large eigenvalues. Table 2 give the breakdown of the variance explained by the first six components. The first three components capture 88.7 per cent of the total variance, with subsequent components demonstrating increasingly diminished returns.

Component	**% of Variance**	**Cumulative %**
1	52.0	52.0
2	22.9	74.9
3	13.8	88.7
4	3.8	92.6
5	2.5	95.1
6	1.8	96.9

Table 2 : Principal Component Analysis : Total Variance Explained by Component

A widely accepted means of determining the number of components to retain, is based on the visual interpretation of plotting the eigenvalues against their ordinal numbers, known as a scree plot, Kanyongo (2005) and Costello and Osborne (2005). The number of factors to be retained is indicated by the number of eigenvalues above the point of inflection, that is the point at which the line changes directions.

Figure 6, plots the eigenvalues against their associated component, highlighting the relevance of each component. The graph shows the point of inflection at component number four. Using the established criterion outlined, the first three components require further investigation.

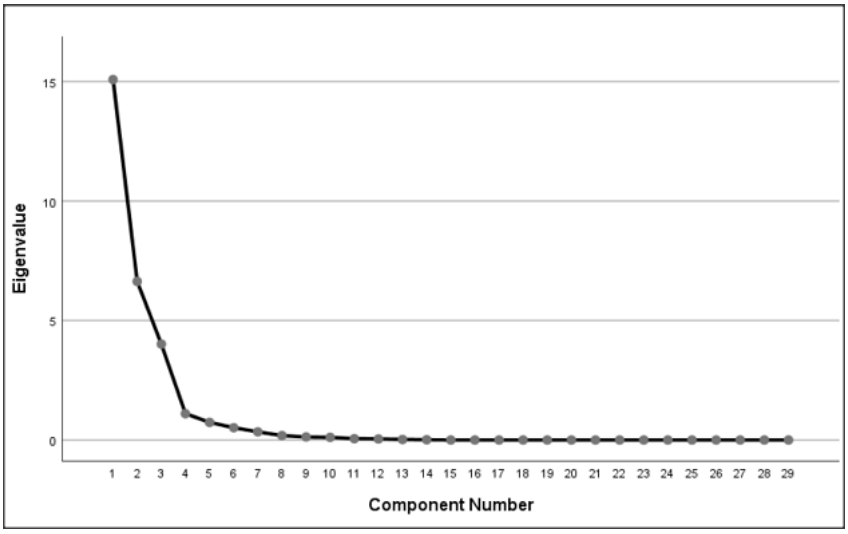

Figure 6: Principal Component Analysis : Scree Plot of Eigenvalues

The findings of the rotated component matrix where either the negative of positive correlation values were greater than .75 (highly correlated) are shown in Table 3. Component 1, which captures 52% of total variance, shows strong positive and negative correlations for a range of variables, including those identified in previous research, such as indicators relating to unemployment and the construction industry.

Component 2, which captures almost 23% of the variance, consists of a range of indicators primarily concentrated on the numbers employed in certain sectors of the economy, while Component 3, which captured 13% of total variance, consists mainly of national economic indicators, and persons employed in the accommodation and food service activities.

The national economic variables identified by the PCA in Component 3 demonstrate a high degree of multicollinearity and as such this research focuses on the new modified Gross National Income as a predictor variable. This new measurement of national income was recommended by the Economic Statistics Review Group and is designed to exclude globalisation effects that are disproportionately impacting the measurement of the size of the Irish economy, Central Statistics Office (2017b). This indicator was created due to the widespread practice of multinational companies' moving balance sheets to Ireland, causing GDP to rise by more the 26% in 2016, European Commission (2018).

It is therefore selected for modelling as it has been designed to the reduce the statistical noise on such indicators within Ireland due to the presence of large multinational corporation's profits.

Component 1 - 52% of Total Variance	
Unemployment Rate 15 - 24 yr olds	0.986
Unemployment Rate 15 - 74 yr olds	0.982
Volume of Production Index in Building and Construction	-0.955
Persons Employed in Industry	-0.934
Persons Employed in Construction	-0.892
Persons Employed in Agriculture, forestry and fishing	-0.792
Persons Employed in Human health and social work activities	0.773
Component 2 - 22.8% of Total Variance	
Persons Employed in Wholesale and retail trade	0.903
Euro Exchange Rate v Dollar	0.874
Persons Employed in Administrative and support service activities	0.872
Persons Employed in Transportation and storage	0.857
Persons Employed in Public administration and defence, compulsory social	0.849
Overseas Trips by Irish Residents to visit friends / relatives	0.764
Component 3 - 13.8% of Total Variance	
GNI	0.958
GNP	0.957
GDP	0.939
Modified GNI	0.910
Persons Employed in Accommodation and food service activities	0.759

Table 3 : Principal Component Analysis : Component Loading Breakdown with positive or negative correlation greater than .75

4.2 Time Series Modelling

SPSS Modeler was used for the time-series forecasting. SPSS Modeler allows for the building of a range of regression and time-series forecasting models, including ARIMA and Exponential Smoothing models. It produces statistical analysis for each model on which performance can be compared and evaluated.

4.2.1 Training and Testing Data

Following the PCA resulting in the reduction to fifteen key predictor variables, SPSS Modeler was used to build a range of time-series models to examine their suitability for predicting net migration in Ireland. Data was available for all fifteen variables from 2002 to 2016, a period of fifteen years.

To build and test the model a hold-out method was employed, with the data from 2002 to 2013 used as the training data to build models, which predicted net migration for the three-year period 2014 to 2016. This was an 80% / 20% split. Both the hold-out method and the split ratio are consistent with recommended practice in time series forecasting Zhang (2003). This enabled the various models performance to be tested against actual historic data.

4.2.2 Model Design

The design of the time series models within SPSS Modeler is shown in Figure 7 below. This design allowed for the data to be tested across a range of time series models. A similar design was employed for the Auto Numeric modelling.

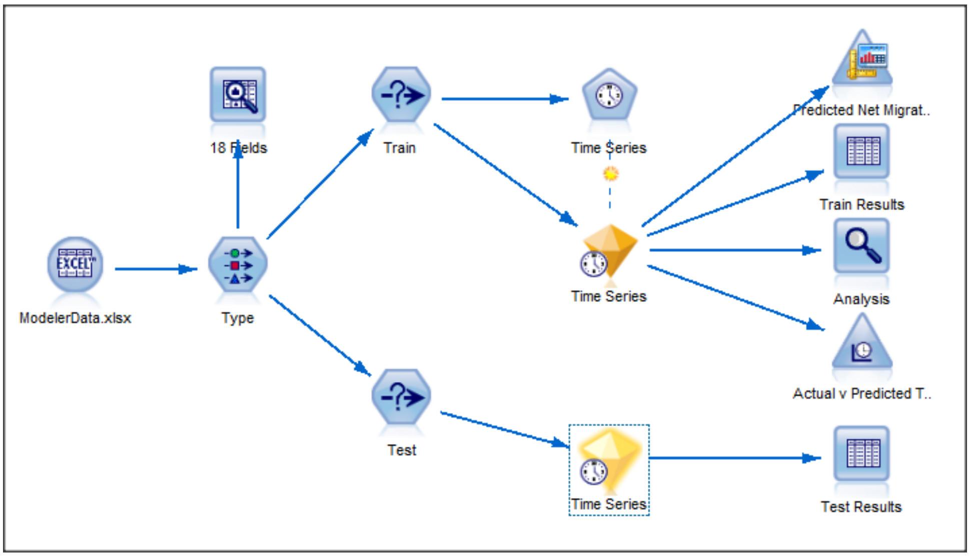

Figure 7: SPSS Modeler Design Overview

Following an evaluation of the performance of the range of models, the best performing model was then extended using the full dataset from 2002 to 2016, to forecast net migration for the five-year period 2017 to 2021.

5 Evaluation

5.1 Model Selection

SPSS Modeler's Auto-Numeric modelling node supports continuous numeric range outcomes across a range of models. A general linear regression model was determined as producing the best predictive outcome when tested against actual net migration data for 2014 to 2016.

The Expert Modeler Time Series feature within SPSS Modeler allows for a range of ARIMA and Exponential Smoothing Models to be compared. Of the models evaluated using this feature, an ARIMA model with parameters (0,1,0), proved to have the greatest accuracy against actual data for 2014 to 2016.

The results of these two models are compared in the next subsection.

5.2 Comparison

The predicted net migration forecasts for the Linear Regression Model and the ARIMA (0,1,0) models for the period 2014 to 2016 are shown in Table 4. Actual CSO M1 scenario predictions for the period, Central Statistics Office (2013) are shown alongside for comparison.

The ARIMA model is the only one which captures the move from negative to positive migration in 2015. The Linear Regression Model correctly forecasts negative net migration for 2014, but predicts an increase in negative net migration for 2015, when there was actual positive net migration. The limitations of the static CSO M1 prediction can be seen as it fails to capture the trend over the three years, one of the main challenges in predicting net migration identified by Dignan (2010).

Year	**Net Migration**	**ARIMA**	**Linear Regression**	**CSO M1**
2014	-8.5	-3.0	-19.6	-19
2015	5.9	12.8	-30.8	-19
2016	16.2	28.5	-3.1	-19

Table 4: Actual Versus Predicted Net Migration, 2014 - 2016 (000's)

A statistical evaluation of the results of the two models developed as part of this research and official forecasts was conducted.

A range of statistical measures used for evaluating time-series forecasts were employed, including the Mean Square Error (MSE), the Root Mean Square Error (RMSE) and the Mean Absolute Percentage Error (MAPE).

The MAPE is an internationally recognised measure used in the prediction accuracy of population projections. MAPE has been adjudged by Swanson et al. (2000) as being a reliable and most informative measure of forecasting error in this field.

Table 5 gives a breakdown of these findings. The ARIMA model has a lower MSE and RMSE than both the Linear Regression model and the CSO forecast. Furthermore, it has a lower MAPE score.

The ARIMA model also has a higher linear correlation than the other forecasts.

The findings clearly demonstrate the ARIMA model outperforming both the Linear Regression Model and the CSO M1 forecast against all evaluation criteria.

Measure	**ARIMA**	**Linear Regression**	**CSO M1**
Min Error	5.5	11.1	10.5
Max Error	12.3	36.7	35.2
MSE	76.5	613.4	656.4
RMSE	-8.7	24.7	25.6
MAPE	.95	2.65	1.23
Linear Correlation	.99	0.51	N/A

Table 5: Comparative Analysis of Model Performance against Historic Data, 2014 - 2016

On this basis, it was chosen as the model to use to extend the forecast of net migration to 2017 to 2021.

5.3 Forecasting Net Migration : 2017 to 2021

For this forecast, the ARIMA model was trained with the entire dataset from 2002 to 2016. The results of this forecast are shown below in Table 6. The model predicts a year on year increase in net migration over the period, culminating in annual positive net migration exceeding fifty thousand people in 2021. This exceeds the current CSO, optimistic M1 scenario, which sets inward migration at an annual rate of thirty thousand.

Year	ARIMA	CSO M1	CSO M2	CSO M3
2017	23.4	30.0	20.0	10.0
2018	30.6	30.0	20.0	10.0
2019	37.8	30.0	20.0	10.0
2020	45.0	30.0	20.0	10.0
2021	50.2	30.0	20.0	10.0

Table 6: Forecasted Net Migration for 2017 - 2021 (000's)

Figure 8, shows the graph of this forecast and it's associated upper and lower 95% Confidence Intervals plotted against the three current CSO scenarios.

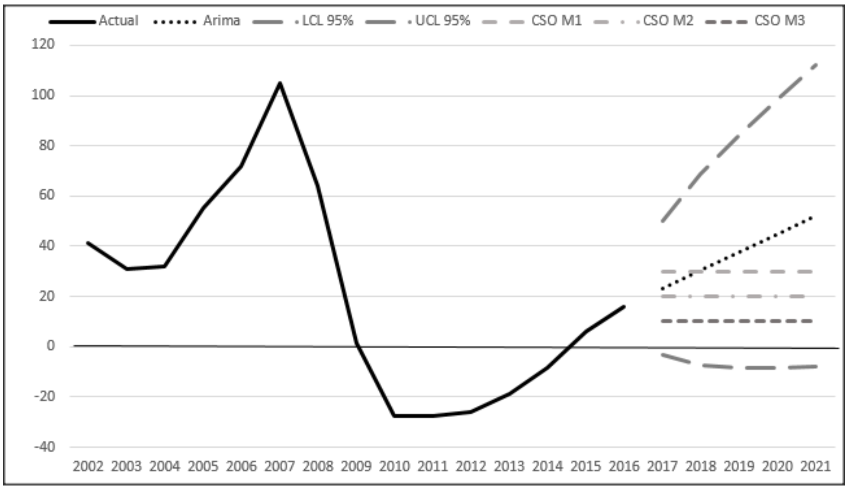

Figure 8: Forecasted Net Migration for 2017 - 2021 (000's)

6 Conclusion and Future Work

6.1 Discussion

In seeking to improve population projections, this research focuses on the need for improvement in forecasting net migration. While any attempt at net migration forecasting faces uncertainty, the methodology employed in building this model looks at improving upon current official predictions. The results of the ARIMA model for 2014 to 2016, show that over this period it did improve upon official estimates.

The use of a range of economic variables, identified in the PCA, enable it to better meet one of the key challenges facing migration forecasting, economic circumstances. The role that economic factors play in migration flows is well established. This studies identifies certain sectors of the economy as of particular importance within an Irish context, particularly the construction industry. This is unsurprising given its predominant role in the Irish economic performance of recent years, Kelly (2009).

The ARIMA model also demonstrated an ability to capture the upward net migration trend that emerged in Ireland over the period 2014 to 2016. However, in forecasting an upward trend in net migration for the period 2017 to 2021, it cannot take into account such one-off events as the economic uncertainty and potential change in migratory flows associated with BREXIT. The outcome of such events are typically beyond the scope of such models, however future studies in the area may be able to examine how these factors affect Ireland.

While the availability of a range of economic and migrationary data within Ireland has proved beneficial, finding datasets pertaining to potential social factors that can influence net migration proved difficult. Within this research, the GINI index, a measure of a nations income inequality acts as the sole social indicator within the dataset. Whereas Moral-pajares and Jiménez-jiménez (2014), used the Migrant Integration Policy Index (MIPEX), Huddleston (2015) as a social indicator, this data has not been updated since.

6.2 Future Work

The process established in this research can be modified to incorporate and examine further pertinent datasets as they are identified or become available.

Future potential data sources for social factors that affect migration include the new EU Social Scoreboard. This monitors EU member states performance across fourteen social indicators linked to the European Pillar of Social Rights, European Commission (2017). Also, the United Nations Happiness Index, has been collating country specific data since 2006 for a range of social factors, United Nations (2018). These both present excellent future opportunities to extend the scope of this research to further incorporate social factors.

The output of this research can also be used as an important consideration across other key policy areas, including housing. The different characteristic that new migrants cohorts have been shown to have in relation to housing needs, including a higher usage rate for private rented accommodation, Martin et al. (2018) may help inform future public policy in this area.

Furthermore, with the population of Ireland getting steadily older since the 1980's, and the numbers over 65 years of age increasing by a fifth between Census 2011 and Census 2016, Central Statistics Office (2017a), the requirement for replacement migration will play an increasing role in future years. This will require more long-term strategic thinking on the part of governments in managing future net migration, United Nations (2001). Increased understanding of the key drivers of net migration and age-specific migration forecasting models will help in this regard.

A key aspect of the national development plan, Project Ireland 2040 is developing a region focused strategy for national development, Dept. of Housing Planning & Local Government (2018). Given the clear role of net migration in this growth, there is future scope to extend this study to examine sub-national net migration in Ireland, similar to work done in the U.K. by Lomax et al. (2013).

7 Acknowledgements

Firstly, I wish to thank my wife Céire for her support and understanding throughout the course of my studies including this paper. This was particularly important at the time of the pregnancy and birth of our first child Sive, when she showed truly exceptional levels of understanding and support.

I also need to thank my family and my in-laws for helping out with duties on the home front when needed, affording me every opportunity to concentrate on my studies.

I am extremely grateful to my employers for offering invaluable assistance.

Finally, I wish to acknowledge the direction and guidance of my supervisor, Vikas Sahni, whose input along the way has been insightful and invaluable.

References

Alkema, L., Gerland, P., Raftery, A. and Wilmoth, J. (2015). The United Nations Probabilistic Population Projections: An Introduction to Demographic Forecasting with Uncertainty., *Foresight (Colchester, Vt.)* **2015**(37): 19–24.

Arnott, R. D. and Chaves, D. B. (2012). Demographic Changes, Financial Markets , and the Economy, **68**(1).

Barrell, R., Fitzgerald, J. and Riley, R. (2007). EU enlargement and migration : Assessing the macroeconomic impacts, *Economic and Social Research Institute* **44**.

Central Statistics Office (2002). Population and Labour Force Projections, 2006-2036, (December).

Central Statistics Office (2008). *Population and Labour Force Projections, 2011 - 2041 Central*.

Central Statistics Office (2013). Population and Labour Force Projections, 2016-2046.

Central Statistics Office (2017a). Census of Population 2016 - Profile 3 An Age Profile of Ireland.
URL: *https://www.cso.ie/en/releasesandpublications/ep/p-cp3oy/cp3/assr/*

Central Statistics Office (2017b). National Income and Expenditure Accounts 2016.

Central Statistics Office (2018). Population and Labour Force Projections 2017 - 2051.

Constantin, C. (2014). Principal Component Analysis - a Powerful Tool in Computing Marketing Information., *Bulletin of the Transilvania University of Brasov. Series V: Economic Sciences* **7**(2): 25–30.

Costello, A. B. and Osborne, J. W. (2005). Best Practices in Exploratory Factor Analysis : Four Recommendations for Getting the Most From Your Analysis, *Practical Assessment, Research & Education* **10**: 1–9.

Dept. of Housing Planning & Local Government (2018). Project Ireland 2040 - National Planning Framework.
URL: *http://npf.ie/wp-content/uploads/Project-Ireland-2040-NPF.pdf*

Dignan, T. (2010). Exploring Population Projections: Sources of Uncertainty and the User Perspective, *Journal of the Statistical and Social Inquiry Society of Ireland* **XL**(110): 33–59.

Donoho, D. L. (2000). Aide-Memoire. High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality, *American Math. Society Lecture-Math Challenges of the 21st Century* pp. 1–33.

Duffy, D., Byrne, D. and Fitzgerald, J. (2014). Alternative Scenarios for New Household Formation in Ireland, *ESRI Special Article [Online]* .

European Commission (2017). The EU and Social European Scoreboard 2017.

European Commission (2018). Country Report Ireland 2018, (1176).
URL: *https://ec.europa.eu/info/sites/info/files/2018-european-semester-country-report-ireland-en_1.pdf*

European Network for Housing Research (2018). ENHR Conference 2018, Workshop 13: Housing Market Dynamics, pp. 27–28.
URL: *http://www.enhr2018.com/digitalAssets/689/c_689982-l_1-k_rev.ws-13-housing-market-dynamics.pdf*

Fitzgerald, J. (2014). Ireland's Recovery from Crisis, *CESifo Forum* **2**: 8–13.

Gilmartin, M. (2012). The changing landscape of Irish migration, 2000-2012, *National Institute for Regional and Spatial Analysis* (69): 2000–2012.

Government of Ireland (2018). Project Ireland 2040: National Development Plan 2018-2027.

Huddleston, T. (2015). *Migrant Integration Policy Index 2015*, CIDOB and MPG.

Kanyongo, G. Y. (2005). Determining The Correct Number Of Components To Extract From A Principal Components Analysis: A Monte Carlo Study Of The Accuracy Of The Scree Plot, *Journal of Modern Applied Statistical Methods* **4**(1): 120–133.

Kelly, M. (2009). The Irish property bubble: causes and consequences, *Dublin: UCD* .

Lomax, N., Norman, P., Rees, P. and Stillwell, J. (2013). Subnational migration in the United Kingdom: Producing a consistent time series using a combination of available data and estimates, *Journal of Population Research* **30**(3): 265–288.

Lymperopoulou, K. (2018). Data Sources on International Migration in the UK, Data in Europe: Migration [webinar].

Martin, C., Hulse, K. and Pawson, H. (2018). *The changing institutions of private rental housing: an international review.*, number 292.

Moral-pajares, E. and Jiménez-jiménez, F. (2014). Migration within the EU-15: Pull factors and choice of destination, *Revista de Economía Mundial, 1* (16): 181–200.

Seo, S. and Gary M. Marsh, P. D. (2006). A review and comparison of methods for detecting outliersin univariate data sets, *Department of Biostatistics, Graduate School of Public Health* pp. 1–53.

Shaw, C. (2007). Fifty years of United Kingdom national population projections : how accurate have they been ?, *Office of National Statistics* p. 2007.

Swanson, D. A., Tayman, J. and Barr, C. F. (2000). A note on the measurement of accuracy for subnational demographic estimates, *Demography* **37**(2): 193–201.

Tayman, J., Smith, S. K. and Lin, J. (2007). Forecasts : an exploratory analysis of time series models, pp. 347–369.

Thomas Ng, S., Skitmore, M. and Wong, K. F. (2008). Using genetic algorithms and linear regression analysis for private housing demand forecast, *Building and Environment* **43**(6): 1171–1184.

Timoney, K. (2012). Revisions to Population , Migration and the Labour Force, *Quarterly Economic Commentary* (Winter): 2007–2011.

United Nations (2001). Replacement Migrations: Is it a Solution to declining and Ageing Populations?, p. 157.

United Nations (2018). World Happiness Report 2018, *World Happiness Report* pp. 146–159.

United Nations, Department of Economic and Social Affairs, P. D. (2017). World Population Prospects: The 2017 Revision, Methodology of the United Nations Population Estimates and Projections.

Williams, L. J. (2010). Principal Component Analysis, *English* **2**(4): 433–470.

Wirth, R. (2000). CRISP-DM : Towards a Standard Process Model for Data Mining, *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining* (24959): 29–39.

Wold, S., Bro, R. and Smilde, A. K. (1987). Principal component analysis, *Analytical Methods* **6**(9): 2812–2831.

Wright, J., Peng, Y., Ma, Y., Ganesh, A. and Rao, S. (2009). Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices, *Advances in Neural Information Processing Systems (NIPS)* pp. 2080—-2088.

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