Analysis of the use of Classical Regression and RegARIMA-based models in the short-term forecasting of Irish macroeconomic time-series in official statistics

Patrick Foley, Tim Linehan and Fiona O'Callaghan

Central Statistics Office, Ireland

Abstract

In a climate where there is an increasing demand for early estimates of economic indicators, this paper explores forecasting methods for producing short-term estimates of a number of Irish macroeconomic time series. Temporal disaggregation methods are explored to obtain monthly explanatory variables. Forecasting methods using classical regression models are used for exploratory data analysis. Univariate forecasting methods are explored in detail to predict GDP and its components. Several different methods are compared. The suitability of the Eurostat DG ECFIN Economic Sentiment Indicator and its component series for the regression model are further examined. Nowcasting GDP using dynamic factor models is also explored with several CSO short term economic series used to estimate their usefulness in nowcasting Irish GDP. This project used R packages including RJDEMETRA+ and TEMPDISAGG packages, as well as JDEMETRA+, Win X-13 and SAS Time Series Forecasting System.

Keywords: Univariate Forecasting, Nowcasting, Temporal Disaggregation

1. Introduction

The Central Statistics Office (CSO), Ireland is committed to producing and publishing high quality statistics. A key component of high-quality statistics is the production of timely statistical outputs. However, in producing rapid statistical estimates there is often a conflict with other dimensions of quality. Short-term forecasting methods, including nowcasting, offer potential approaches for producing early reliable estimates of official statistics. Eurostat's Handbook on Rapid Estimates[1] provides a detailed and comprehensive outline of practical and suitable compilation methods for rapid estimates. The Federal Reserve Bank of Atlanta's GDPNow[2] system propose methods for producing short term estimates of key economic indicators, such as Gross Domestic Product (GDP). In the Irish context, recent works by Casey[3] and Conefrey and Walsh[4] and D'Agostino, McQuinn and O'Brien[5] use a nowcasting framework to provide early estimates of GDP.

[1] Eurostat (2017), Handbook on Rapid Estimates.
[2] Higgins, P. "GDPNow: A Model for GDP "Nowcasting"", Working Paper 2014-07, Working Paper Series, Federal Reserve Bank of Atlanta.
[3] Casey E, "Nowcasting to Predict Data Revisions", Working Paper No. 7, Irish Fiscal Advisory Council

In this paper, various methods for short-term forecasts of Ireland's GDP are considered, namely (a) classical regression approach and (b) the univariate forecasting of sub-component time-series. Univariate forecasting was performed using RJDEMETRA+, WinX-13 and the SAS Time Series Forecasting System. Hold-out sampling was used to assess the performance of each of these methods. This paper also includes a demonstration of the use of temporal disaggregation methods using the R TEMPDISAGG package, in order to obtain monthly data for use in the regression models. In addition to this, the possibility of nowcasting using Dynamic Factor methods is also considered.

2. Datasets examined in linear regression model

In the choice of datasets used, the CSO was guided by the GDPNow table of predictor variables. Certain potential explanatory series were not included in this process because data was not available for the entire time span.

Indicator series for classical regression

To enable the classical regression approach, various indicator series that have been identified as potential indicators for GDP were produced. Table 1 lists these series by name and function. Indicator data from 2000 to 2018 was used. EU DGECFIN economic sentiment indicator (ESI) and confidence data was also used.

[4] Conefrey T. and Walsh G., "A Monthly Indicator of Economic Activity for Ireland", *Economic Letter*, Vol. 2018 No. 14, Central Bank of Ireland

[5] D'Agostino,A, K. McQuinn and D.O'Brien (2012), "Nowcasting Irish GDP." OECD Journal: Journal of Business Cycle Measurement and Analysis, 2012, Issue 2, 21-31.

Table 2.1 Indicator series for classical regression.

Name	Function	Origin
IE_INDU_CONF_IND	Industrial Confidence Indicator - Ireland	DGECFIN
IE_SERV_CONF_IND	Services Confidence Indicator - Ireland	DGECFIN
IE_CONS_CONF_IND	Construction Confidence Indicator - Ireland	DGECFIN
IE_RETA_CONF_IND	Retail Confidence Indicator - Ireland	DGECFIN
IE_BUIL_CONF_IND	Building Confidence Indicator - Ireland	DGECFIN
IE_ESI_CONF_IND	Economic Sentiment Indicator - Ireland	DGECFIN
All_Vehicles_Licenced	Vehicle Licences - Ireland	CSO
New_Vehicles_Licenced	New Vehicles Licenced - Ireland	CSO
CPI_B2006_All	Consumer Price Index (base 2006) - Overall	CSO
CPI_B2006_Food	Consumer Price Index (base 2006) - Food	CSO
CPI_B2006_AlcTob	Consumer Price Index (base 2006) - Alcohol/Tobacco	CSO
CPI_B2006_Cloth	Consumer Price Index (base 2006) - Clothing	CSO
CPI_B2006_House	Consumer Price Index (base 2006) - House	CSO
CPI_B2006_Furni	Consumer Price Index (base 2006) - Furniture	CSO
CPI_B2006_Health	Consumer Price Index (base 2006) - Health	CSO
CPI_B2006_Transport	Consumer Price Index (base 2006) - Transport	CSO
CPI_B2006_Comm	Consumer Price Index (base 2006) - Communications	CSO
CPI_B2006_Recreat	Consumer Price Index (base 2006) - Recreation	CSO
CPI_B2006_Educ	Consumer Price Index (base 2006) - Education	CSO
CPI_B2006_Rest	Consumer Price Index (base 2006) - Restaurants	CSO
CPI_B2006_Misc	Consumer Price Index (base 2006) - Miscellaneous	CSO
IND_PROD_VT_IND_10_33_MANUF_B2010	Production Indices - Manufacturing	CSO
IND_PROD_VT_IND_10_33_MINE_B2010	Production Indices - Mining	CSO
SA_IND_PROD_VT_IND_10_33_MANUF_B2010	Seasonally Adj. Production Indices - Manufacturing	CSO
SA_IND_PROD_VT_IND_10_33_MINE_B2010	Seasonally Adj. Production Indices - Mining	CSO
LR_SA_15_74_ALL	Live register - seasonally adjusted - aged 15-74	CSO
Population	Population of Ireland Estimates	CSO

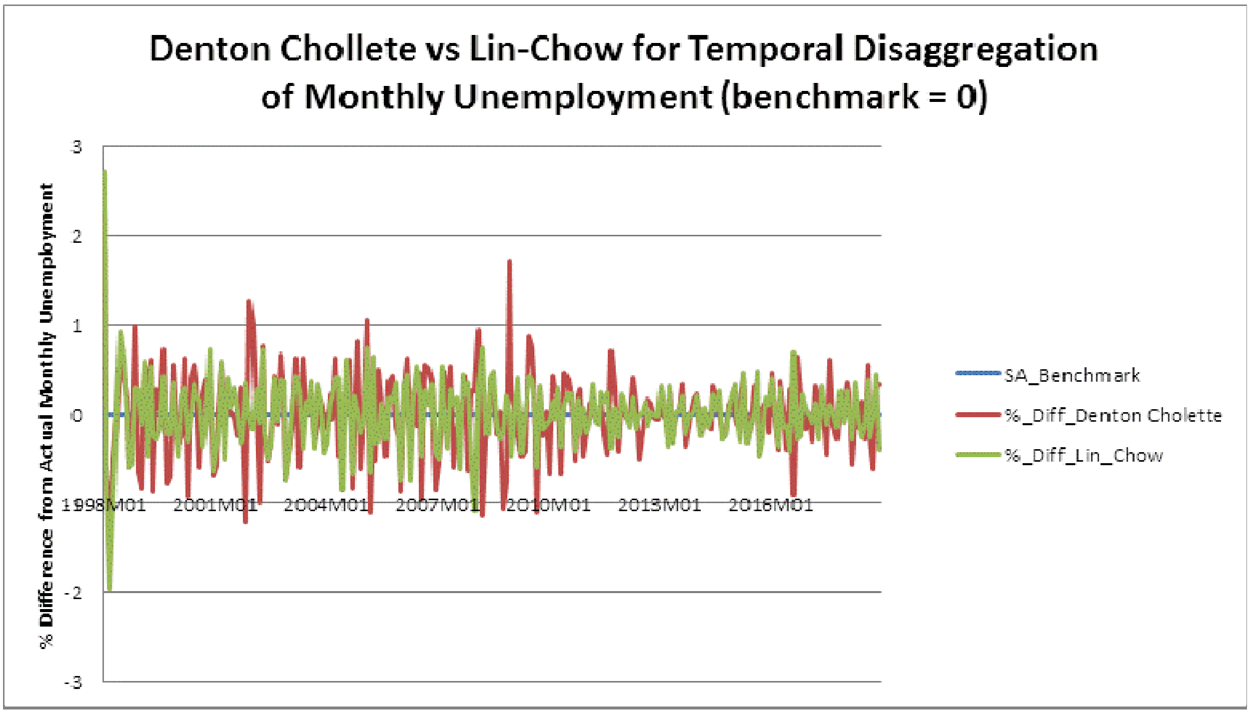
CSO time series data for Irish GDP at current Market Price data from 1995 to Quarter 1 2019 (both the original and the seasonally adjusted versions) were used. Monthly data was temporally disaggregated from the quarterly dataset using the R TEMPDISAGG package using the Denton-Cholette Method[6]. This method was suggested by the Eurostat Temporal Disaggregation guidelines[7].

As part of the initial data process, temporal disaggregation was applied to the quarterly unemployment data to obtain monthly estimates using the Denton-Cholette and Lin Chow methods. The resulting disaggregated data was then compared with the recorded

[6] Dagum, E.B. and Cholette P.A., Benchmarking, “Temporal Disaggregation, and Reconciliation Methods for Time Series”, Springer, 2006.
[7] “ESS guidelines on temporal disaggregation, benchmarking and reconciliation”, Eurostat 2018.

unemployment data and the following graph was generated (Fig. 1). When compared with the actual monthly unemployment data, the vast majority of estimates are within +/-1% of the actual monthly unemployment figure. The Lin-Chow method appears to produce closer estimates. This was a validation exercise in temporal disaggregation. As more reliable live register data was available on a monthly basis it was used in the regression model instead.

Fig 2.1. Temporal Disaggregation of Unemployment data



3. Exploratory data analysis

This step centred around using classical regression models to identify variables with a potential relationship with GDP. The use of linear regression models for building a GDP forecasting model was not seriously considered due to the numerous issues with such a linear model. A successful linear regression model requires linear relationships; multivariate normality; little or no multi-collinearity and no auto-correlation. None of these conditions is traditionally associated with simple GDP linear regression models. However, a classical regression approach can be useful in identifying potential relationships between indicator variables and GDP.

For the linear regression software, SAS was chosen. Models were generated both for GDP and seasonally adjusted GDP. A forward selection model was used, and the following results were obtained for the seasonally adjusted GDP model. (Table 3.1).

Table3.1. Summary of regression model for seasonally adjusted GDP.

Analysis of Variance						
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	Final R-Square
Model	22	2384752974	108397862	926.85	<.0001	0.9912
Error	181	21168465	116953			
Corrected Total	203	2405921439				

When considering the contributions of statistically significant factors to the models, the most significant elements can be seen in Table 3.2.

The next step was to consider the possible predictive value of the classical forecasting approach, notwithstanding the previous caveats. Observed and predicted GDP values were compared in order to establish if the model has predictive value. Fig. 3.1 shows the results of this process for the seasonally adjusted GDP model.

The differences are (relatively) patternless. However, in most recent quarters, the difference between the observed and (model-) predicted GDP values tended to be negative.

The exploratory data process had the following conclusions:

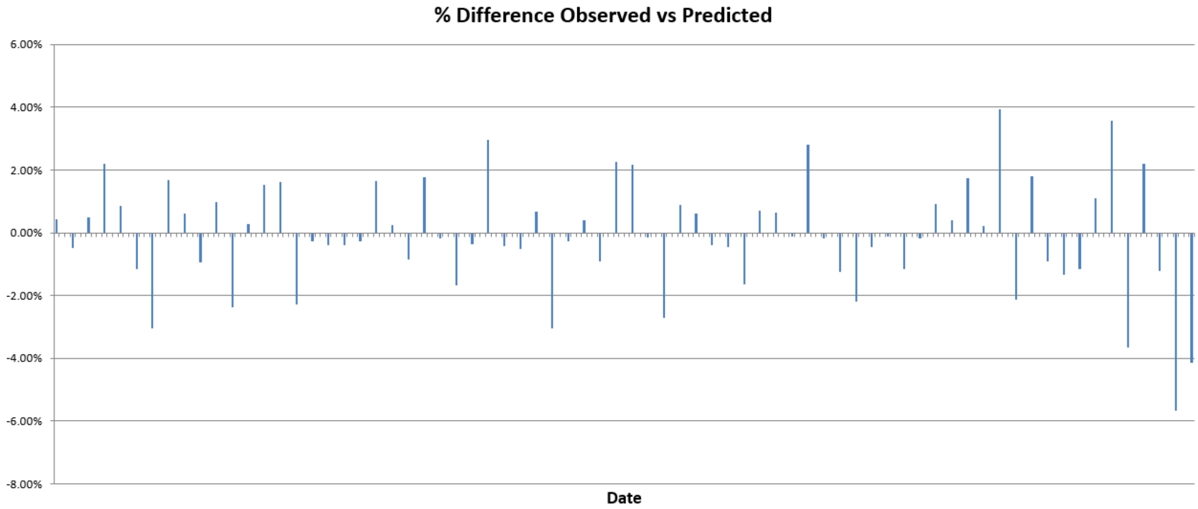
- Several variables have a high predictive role for GDP and could be useful for the nowcasting process.
- Three of the six DG ECFIN sentiment indices analysed were significant in the regression model (building, construction and retail).
- Linear regression models are, by themselves, not really suitable for forecasting. However, the models provide a useful means for identifying variables for the nowcasting process.
- While the absolute errors were generally small, the model did not forecast the direction of changes well.

The next step was to commence univariate sub-component forecasting using SAS, RJDEMETRA+ and Win X-13. The methods used would be univariate, including regARIMA based approaches.

Table 3.2. Summary of forward step-wise regression process for SA_GDP

Significant variable	Label	Origin
All_Vehicles_Licenced	Vehicle Licences - Ireland	CSO
CPI_B2006_AlcTob	Consumer Price Index (base 2006) - Alcohol/Tobacco	CSO
CPI_B2006_All	Consumer Price Index (base 2006) - Overall	CSO
CPI_B2006_Cloth	Consumer Price Index (base 2006) - Clothing	CSO
CPI_B2006_Comm	Consumer Price Index (base 2006) - Communications	CSO
CPI_B2006_Educ	Consumer Price Index (base 2006) - Education	CSO
CPI_B2006_Food	Consumer Price Index (base 2006) - Food	CSO
CPI_B2006_Furni	Consumer Price Index (base 2006) - Furniture	CSO
CPI_B2006_Health	Consumer Price Index (base 2006) - Health	CSO
CPI_B2006_House	Consumer Price Index (base 2006) - House	CSO
CPI_B2006_Misc	Consumer Price Index (base 2006) - Miscellaneous	CSO
CPI_B2006_Rest	Consumer Price Index (base 2006) - Restaurants	CSO
CPI_B2006_Transport	Consumer Price Index (base 2006) - Transport	CSO
IE_BUIL_CONF_IND	Building Confidence Indicator - Ireland	DGECFIN
IE_CONS_CONF_IND	Consumer Confidence Indicator - Ireland	DGECFIN
IE_RETA_CONF_IND	Retail Confidence Indicator - Ireland	DGECFIN
IND_PROD_VT_IND_10_33_MANUF_B201	Production Indices - Manufacturing	CSO
IND_PROD_VT_IND_10_33_MINE_B2010	Seasonally Adj. Production Indices - Mining	CSO
LR_SA_15_74_ALL	Live register - seasonally adjusted - aged 15-74	CSO
New_Vehicles_Licenced	New Vehicles Licenced - Ireland	CSO
SA_IND_PROD_VT_IND_10_33_MANUF_B	Seasonally Adj. Production Indices - Manufacturing	CSO
SA_IND_PROD_VT_IND_10_33_MINE_B2	Seasonally Adj. Production Indices - Mining	CSO

Fig 3.1. Percentage differences between observed and predicted for Seasonal GDP



4. Univariate Forecasting

The CSO's National Accounts Divisions issued the authors with the sub-component time series for GDP from 1995 to Q1 2019. This consisted of 140 GDP sub-components. These series could then each be "naively" forecasted using regARIMA or other univariate time-series methods and then be used to produce GDP estimates, in a manner similar to standard National Accounts preparation.

Univariate sub-component forecasting was carried out using three different approaches; the SAS Time Series Forecasting System, Win X-13 and RJDemetra+. An analysis was carried out using hold-out samples to establish which of these approaches performed the best.

4.1 Time Series Forecasting System - SAS

The Time Series Forecasting System (TSFS) available in SAS/ETS is a point-and-click, menu-based system that allows you to forecast values of time series variables by extrapolating trends and patterns in the past values of the series. It provides automatic model fitting and forecasting as well as interactive model development. Because of the large number of series being analysed here, the automatic model selection feature was used to choose the best-fitting model for each time series. The TSFS includes a wide range of forecasting models

- exponential smoothing - (simple, double, linear, damped-trend, seasonal)
- Winters smoothing, additive and multiplicative
- Box-Jenkins ARIMA models, including seasonal ARIMA models
- predictor variables

- simple regressors
- seasonal dummy variable regressors
- intervention (dummy) variables to model exceptional events, level shifts, or trend shifts
- adjustment variables to adjust the forecasts by fixed amounts at each period
- transfer functions or dynamic regression: use transformations, lags, or time series filters to model the impact of predictor variables
- time trend models – (linear, quadratic, cubic, logistic, logarithmic, exponential, hyperbolic, power function, exp(A+B/time))
- data transformations – (logarithmic, logistic, square root, Box-Cox)
- combining or average the predictions of other forecasting models
- external (judgemental) forecasts
- customised models

The automatic model selection checks each time series against each of the above models and selects the best model based on the criterion specified by the user. The default selection criterion is the root mean square error but other options include the mean square error, the mean absolute error, the mean absolute percent error and the r-square. The default of the root mean square error was used for this exercise.

Forecasts were produced for all subcomponent series for the periods Q1 2018 to Q1 2019. Table 4.1 provides a summary of the models that were automatically selected for the period Q1 2019. Note that an ARIMA model was not the automatically selected model in the majority of cases. In total, only 11.4% of the series were modelled as ARIMA series.

Table 4.1 Models selected by SAS TSFS, Q1 2019

Model	Percent
Winters Method -- Additive	19.2%
Log Winters Method -- Additive	15.0%
Linear (Holt) Exponential Smoothing	10.8%
Linear Trend with Autoregressive Errors	9.6%
Log Damped Trend Exponential Smoothing	9.0%
Winters Method -- Multiplicative	7.2%
Damped Trend Exponential Smoothing	6.6%
Simple Exponential Smoothing	3.6%
ARIMA(2,1,2)(0,1,1)s NOINT	3.0%
Log ARIMA(2,1,2)(0,1,1)s NOINT	2.4%
Log Linear (Holt) Exponential Smoothing	2.4%
Log Linear Trend with Autoregressive Errors	2.4%
Log Airline Model	1.8%
Log Winters Method -- Multiplicative	1.8%
Log Seasonal Exponential Smoothing	1.2%
ARIMA(0,1,1)(1,0,0)s NOINT	0.6%
ARIMA(0,1,2)(0,1,1)s NOINT	0.6%
ARIMA(2,0,0)(1,0,0)s	0.6%
Log ARIMA(0,1,2)(0,1,1)s NOINT	0.6%
Log ARIMA(0,2,2)(0,1,1)s NOINT	0.6%
Log ARIMA(2,1,0)(0,1,1)s NOINT	0.6%
Log Linear Trend	0.6%

4.2 Win X-13

Win X-13 is a Windows interface for the X-13ARIMA-SEATS program. It was developed by the US Census Bureau.

Specification files (program code specifying the forecasting models) were developed for each time series using Win X-13 and each of the sub-component time series were modelled separately. An add-on tool called "X-13 Sam" enabled the specification files to be modified in a batch form.

Most of the model specifications were automatically identified. The Transform function in the specification file determines the functional form (additive or multiplicative) of the model. The regression procedure was coded so that the models were tested to determine if regression parameters such as an Easter effect, trading effects, additive outliers, level shifts or temporary changes needed to be modelled. If they were discovered, they were automatically included in the model so that the appropriate ARIMA model could then be established.

Using the “AUTOMDL” procedure, each time the specification file is run, updated seasonal ARIMA models are automatically generated and forecasts are generated for one period ahead for each sub-component. The specification files were run in batch form using a metadata file.

The forecasts are minimum mean squared error (MMSE) linear predictions of future values based on the present and past values assuming that the true model is used—which means we assume the regARIMA model form is correct, that the correct regression variables have been included, that no additive outliers or level shifts will occur in the forecast period, that the specified ARIMA orders are correct, and that the parameter values used (typically estimated parameters) are equal to the true values. These are standard assumptions, though obviously unrealistic in practical applications. What is more realistically hoped is that the regARIMA model will be a close enough approximation to the true, unknown model for the results to be approximately valid.

4.3 RJDEMETRA+

The RJDEMETRA+ package was also used for univariate forecasting. This is, in effect, a R implementation of the JDEMETRA+ package. The advantage of the package is that, since it is an R script, the management of a large number of sub-component time series is relatively simple. In this case, each of the sub-component time series had an automatic X-13-ARIMA model selection process applied to it and forecasts were generated. These forecasts were then appended to the original time series. Since both RJDEMETRA+ and Win X-13 are X-13-ARIMA based, it is expected that, in most cases, both systems will produce very similar estimates (as is indeed the case). Differences are largely due to different decisions made for individually constructed forecasting models – as identified as necessary from diagnostics in the X-13-ARIMA model building process.

4.4 Univariate Forecasting - Results

Each of the three tools described above was used to generate forecasts for all sub-component series for five quarters (Q1 2018 to Q1 2019) and these forecasts were compared with the actual values. The metric used to evaluate the forecasts was the mean absolute percentage error (MAPE) between the forecast and the actual value across the five quarters.

In total, 140 sub-component series were analysed as part of this project. For each time series, the absolute percentage error was calculated for each quarter and this was then averaged across the five quarters to give the MAPE. The summary statistics in Table 4.2

present the average MAPE, the minimum MAPE and the maximum MAPE across the 140 sub-component series.

Table 4.2

	SAS - TSFS	Win X-13	RJDemetra+
Average	7.30%	8.33%	8.60%
Min	0.00%	0.17%	0.20%
Max	34.76%	98.62%	80.00%

The SAS TSTS outperforms both Win X-13 and RJDemetra+ with an average MAPE of 7.3%. It also had the smallest minimum and maximum values. Out of the 140 series analysed, the TSFS had the smallest MAPE for 67 series, Win X-13 had the smallest MAPE for 41 series and RJDemetra+ had the smallest MAPE for 32 series.

Of the 140 series analysed, approximately 20 relate to capital formation items. Capital formation includes expenditure on building and construction work, machinery and equipment, research, and produced and imported capital goods and services. These types of items tend to be volatile in nature and often do not follow any clear patterns. When these capital formation series are excluded from the forecasting analysis, the performance of the three methods changes considerably, (see Table 4.3).

Table 4.3

	SAS - TSFS	Win X-13	RJDemetra+
Average	5.73%	4.73%	5.60%
Min	0.40%	0.17%	0.20%
Max	26.85%	19.59%	26.03%

The TSFS no longer performs the best, it now has the largest mean and the largest minimum and maximum.

Table 4.4 shows the results for the capital formation variables only.

Table 4.4

	SAS - TSFS	Win X-13	RJDemetra+
Average	16.20%	28.73%	25.58%
Min	0.00%	5.12%	5.12%
Max	34.76%	98.62%	80.00%

These results indicate that for the capital formation variables the TSFS outperforms the other two methods. If we look at the models that the TSFS selects for the capital formation series in Table 4.5, we get:

Table 4.5

Model	Percent
Winters Method -- Additive	37.5%
Damped Trend Exponential Smoothing	16.7%
Linear (Holt) Exponential Smoothing	12.5%
Log Winters Method -- Additive	12.5%
Log Damped Trend Exponential Smoothing	4.2%
Log Linear (Holt) Exponential Smoothing	4.2%
Log Linear Trend with Autoregressive Errors	4.2%
Simple Exponential Smoothing	4.2%
Winters Method -- Multiplicative	4.2%

The approach taken in this univariate analysis was to forecast the subcomponents of GDP independently. However, aggregates of GDP were also forecasted for the same period. Interestingly, for total 2018 GDP, the direct forecast value of GDP was extremely close to the current actual value and outperformed the indirect estimate based on the forecasted sub-components.

4.5 Conclusions

As expected, the results for Win X-13 and RJDemetra+ are very similar and generally outperform the SAS TSFS. However, for the volatile capital formation variables, the models automatically selected by SAS TSFS performed best over the time-period analysed in this exercise.

These univariate methods are not expected to provide the ultimate solution in terms of early estimates for GDP as they are likely to miss turning points and other important dynamics. However, these estimates can be used as benchmarks against which other methods can be compared.

5. Nowcasting

The Methodology Division of the CSO is currently extending the univariate time series forecasting of National Accounts aggregates and sub-aggregates to investigate the potential for nowcasting to provide reliable early estimates of national accounts data. Nowcasting is frequently defined as the prediction of the present, the very near future and the very recent past. We are currently developing dynamic factor models using the JDemetra+ plug-in developed at the National Bank of Belgium. The plug-in helps to operationalize the process of nowcasting and it can be used to specify and estimate dynamic factor models. The approach allows for the use of an unbalanced dataset, therefore the most recent (within quarter) information for each of the predictive variables can be included in the model.

The current preliminary raw dataset comprises annual, quarterly and monthly publicly available data series. In general, the data is seasonally adjusted and is presented in the form of year on year percentage changes. The data spans from January 2000, where available, to the latest available figures. The data includes both hard and soft data. The hard data is primarily CSO data from National Accounts, Prices, Industry, Labour and Earnings, Demography, Retail Trade, Services and Transport. Several series not used elsewhere are included in the models in order to determine their predictive powers. These series include quarterly capital acquisitions and sales data, quarterly stocks data and monthly services data. See the Table 5.1 below.

Table 5.1 List of capital acquisition, capital sales, stocks and services data used in the nowcasting models.

		Freq	SA	Span
Industry	Capital Acquisitions in Industry (Euro Million) Industries (05 to 39)	Quarterly	Yes	2008 - Latest
	Capital Acquisitions in Industry (Euro Million) Chemicals and pharmaceuticals (20, 21)	Quarterly	Yes	2008 - Latest
	Capital Acquisitions in Industry (Euro Million) Computer, electronic and optical products; other manufacturing (26, 32)	Quarterly	Yes	2008 - Latest
	Capital Sales in Industry (Euro Million) Industries (05 to 39)	Quarterly	Yes	2008 - Latest
	Capital Sales in Industry (Euro Million) Chemicals and pharmaceuticals (20, 21)	Quarterly	Yes	2008 - Latest
	Capital Sales in Industry (Euro Million) Computer, electronic and optical products; other manufacturing (26, 32)	Quarterly	Yes	2008 - Latest
	Change in Value of All Stocks held in Industry (Euro Million) Industries (05 to 39)	Quarterly	Yes	2000 - Latest
	Change in Value of Stocks of Materials and Fuels held in Industry (Euro Million) Industries (05 to 39)	Quarterly	Yes	2000 - Latest
	Change in Value of Stocks of Finished Goods held in Industry (Euro Million) Industries (05 to 39)	Quarterly	Yes	2000 - Latest
	Change in Value of Stocks of Work in Progress Goods held in Industry (Euro Million) Industries (05 to 39)	Quarterly	Yes	2000 - Latest
	Change in Value of Stocks of Goods Purchased Resale held in Industry (Euro Million) Industries (05 to 39)	Quarterly	Yes	2000 - Latest
Services	Wholesale trade (46) Turnover	Monthly	Yes	2009 - Latest
	Transportation and storage (H) Turnover	Monthly	Yes	2009 - Latest
	Accommodation and food service activities (I) Turnover	Monthly	Yes	2009 - Latest
	Information and communication (J) Turnover	Monthly	Yes	2009 - Latest
	Professional, scientific and technical activities (M) Turnover	Monthly	Yes	2009 - Latest
	Administrative and support service activities (N) Turnover	Monthly	Yes	2009 - Latest

GDP figures for the US, the UK and the EU 27 (excluding UK) are included. Soft data in the form of the ESI for Ireland, the UK and the EU 27 (excluding UK) are also included. See Appendix A for the full list of data used.

The state-space representation of the dynamic factor model is described by two equations: a measurement equation

$$y_t = Zf_t + \xi_t \ with \, \xi_t \sim N(0, R)$$

And a transition equation,

$$f_t = A_1 f_{t-1} + \cdots + A_p f_{t-p} + u_t \ with \, u_t \sim N(0, Q)$$

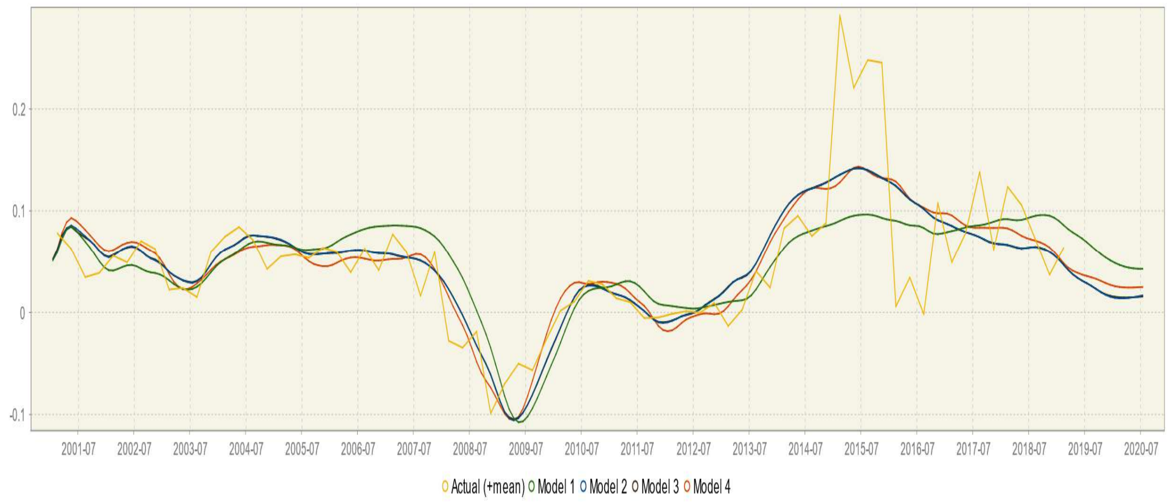
Where y_t is a N x 1 matrix of observed noisy data and f_t is an r x 1 matrix of unobserved factors, Z is a N x r matrix of factor loadings, and A_1, …, A_p are r x r matrices of parameters that determine the factor dynamics. The measurement equation links the N observables to the r underlying factors. The factors follow a VAR of order p. This representation is valid when monthly, quarterly and annual variables are combined. It is assumed that the quarterly and annual values are observed every three and 12 months, respectively. In our dataset we have variables observed at monthly, quarterly and annual frequencies.

Currently we are investigating the performance of several different models in terms of the specification of r and p to determine which models perform the best. Table 5.2 outlines the models explored. Including additional lags in Model 3 made no difference to the performance of Model 2. Figure 5.1 outlines the signals derived for each of the models used. Early investigations, based on the loading factors, suggest that the inclusion of the sentiment indices and the services indices have a positive impact on the analysis.

Table 5.2 Dynamic Factor Models initially explored.

	Number of Factor (r)	Number of Lags (p)
Model 1	2	2
Model 2	3	2
Model 3	3	3
Model 4	4	2

Figure 5.1. Signal vs Data GDP at Market Prices YOY change



Further exploration work will include an investigation into the impact of the real-time data flow on the forecasts. The use of dynamic factor models with data comprising more detailed sub-aggregates of GDP will also be explored. It is also envisaged that dynamic factor models will be explored within the CSO in the context of CPI flash estimates and monthly unemployment estimates.

Appendix

Preliminary Nowcasting Variables

		Annual Percentage Changes	Freq	SA	Span
1		Personal Expenditure on Consumer Goods and Services	Quarterly	Yes	2000 - Latest
2		Net Expenditure by Central and Local Govt. on Current Goods and Services	Quarterly	Yes	2000 - Latest
3		Gross Domestic Fixed Capital Formation	Quarterly	Yes	2000 - Latest
4		Gross domestic fixed capital formation - Machinery and Equipment	Quarterly	Yes	2000 - Latest
5		Gross domestic fixed capital formation - Intangible Assets	Quarterly	Yes	2000 - Latest
6		Gross domestic fixed capital formation - All Building &Construction (including Dwellings, Improvements, Other B&C and Transfer Costs)	Quarterly	Yes	2000 - Latest
7		Value of Physical Changes in Stocks	Quarterly	Yes	2000 - Latest
8		Exports of Goods and Services (excluding Factor Income Flows)	Quarterly	Yes	2000 - Latest
9		Imports of Goods and Services (excluding Factor Income Flows)	Quarterly	Yes	2000 - Latest
10	National Accounts	Gross Domestic Product at Market Prices	Quarterly	Yes	2000 - Latest
11		Net Factor Income from the Rest of the World	Quarterly	Yes	2000 - Latest
12		Gross National Product at Market Prices	Quarterly	Yes	2000 - Latest
13		Export of Goods (excluding Factor Income Flows)	Quarterly	Yes	2000 - Latest
14		Import of Goods (excluding Factor Income Flows)	Quarterly	Yes	2000 - Latest
15		Export of Services (excluding Factor Income Flows)	Quarterly	Yes	2000 - Latest
16		Import of Services (excluding Factor Income Flows)	Quarterly	Yes	2000 - Latest
17		Final Domestic Demand	Quarterly	Yes	2000 - Latest
18		Total Domestic Demand	Quarterly	Yes	2000 - Latest
19		Current account (UN)	Quarterly	No	2000 - Latest
20		Capital account (UN)	Quarterly	No	2000 - Latest
21		Male 15 - 24 years	Annual	No	2000 - Latest
22		Female 15 - 24 years	Annual	No	2000 - Latest
23		Male 25 - 44 years	Annual	No	2000 - Latest
24	Population	Female 25 - 44 years	Annual	No	2000 - Latest
25		Male 45 - 64 years	Annual	No	2000 - Latest
26		Female 45 - 64 years	Annual	No	2000 - Latest
27		Immigrants	Annual	No	2000 - Latest
28		Emigrants	Annual	No	2000 - Latest
29		In employment total (UN)	Quarterly	No	2000 - Latest
30		Hours worked Under 50	Quarterly	Yes	2008 - Latest
31	Earnings and Employment	Hours worked 50 - 250	Quarterly	Yes	2008 - Latest
32		Hours worked Greater than 250	Quarterly	Yes	2008 - Latest
33		Live register SA	Monthly	Yes	2000 - Latest
34		Single house Prices	Quarterly	Yes	2011 - Latest
35	Prices	Apartment Prices	Quarterly	Yes	2011 - Latest
36		CPI	Monthly	No	2000 - Latest
37	Trade	Trade Volume Index for Imports (Seasonally Adjusted) (Base 2010=100)	Monthly	Yes	2000 - Latest
38		Trade Volume Index for Exports (Seasonally Adjusted) (Base 2010=100)	Monthly	Yes	2000 - Latest
39		Capital Acquisitions in Industry (Euro Million) Industries (05 to 39)	Quarterly	Yes	2008 - Latest
40		Capital Acquisitions in Industry (Euro Million) Chemicals and pharmaceuticals (20, 21)	Quarterly	Yes	2008 - Latest
41		Capital Acquisitions in Industry (Euro Million) Computer, electronic and optical products; other manufacturing (26, 32)	Quarterly	Yes	2008 - Latest
42		Capital Sales in Industry (Euro Million) Industries (05 to 39)	Quarterly	Yes	2008 - Latest
43		Capital Sales in Industry (Euro Million) Chemicals and pharmaceuticals (20, 21)	Quarterly	Yes	2008 - Latest
44		Capital Sales in Industry (Euro Million) Computer, electronic and optical products; other manufacturing (26, 32)	Quarterly	Yes	2008 - Latest
45		Change in Value of All Stocks held in Industry (Euro Million) Industries (05 to 39)	Quarterly	Yes	2000 - Latest
46	Industry	Change in Value of Stocks of Materials and Fuels held in Industry (Euro Million) Industries (05 to 39)	Quarterly	Yes	2000 - Latest
47		Change in Value of Stocks of Finished Goods held in Industry (Euro Million) Industries (05 to 39)	Quarterly	Yes	2000 - Latest
48		Change in Value of Stocks of Work in Progress Goods held in Industry (Euro Million) Industries (05 to 39)	Quarterly	Yes	2000 - Latest
49		Change in Value of Stocks of Goods Purchased Resale held in Industry (Euro Million) Industries (05 to 39)	Quarterly	Yes	2000 - Latest
50		All - building and construction	Quarterly	Yes	2000 - Latest
51		National - all residential properties completions	Monthly	Yes	2005 - Latest
52		IPI (SA) Industries (05 to 35)	Monthly	Yes	2000 - Latest
53		IPI Traditional sector (05 to 17,181,19,22 to 25,28 to 31,321 to 324,329,33,35)	Monthly	Yes	2000 - Latest
54		IRI Modern sector (182,20,21,26,27,325)	Monthly	Yes	2000 - Latest
55		Retail volume Motor trades and automotive fuel (45,4730)	Monthly	Yes	2000 - Latest
56	Retail	Retail sales of food (4711,4721 to 4729) volume	Monthly	Yes	2000 - Latest
57		Retail sales of non food products, excluding motor trades, automotive fuel and bars - Volume	Monthly	Yes	2000 - Latest
58		Wholesale trade (46) Turnover	Monthly	Yes	2009 - Latest
59		Transportation and storage (H) Turnover	Monthly	Yes	2009 - Latest
60	Services	Accommodation and food service activities (I) Turnover	Monthly	Yes	2009 - Latest
61		Information and communication (J) Turnover	Monthly	Yes	2009 - Latest
62		Professional, scientific and technical activities (M) Turnover	Monthly	Yes	2009 - Latest
63		Administrative and support service activities (N) Turnover	Monthly	Yes	2009 - Latest
64		New Private Cars	Monthly	Yes	2000 - Latest
65	Transport	New Goods Vehicles	Monthly	Yes	2000 - Latest
66		Secondhand Private Cars	Monthly	Yes	2000 - Latest
67		Secondhand Goods Vehicles	Monthly	Yes	2000 - Latest
68		BUIL.EU_2019.TOT.COF.BS.M	Monthly	Yes	2000 - Latest
69		BUIL.IE.TOT.COF.BS.M	Monthly	Yes	2000 - Latest
70		BUIL.UK.TOT.COF.BS.M	Monthly	Yes	2000 - Latest
71		CONS.EU_2019.TOT.COF.BS.M	Monthly	Yes	2000 - Latest
72		CONS.IE.TOT.COF.BS.M	Monthly	Yes	2000 - Latest
73		CONS.UK.TOT.COF.BS.M	Monthly	Yes	2000 - Latest
74		INDU.EU_2019.TOT.COF.BS.M	Monthly	Yes	2000 - Latest
75	ESI	INDU.IE.TOT.COF.BS.M	Monthly	Yes	2000 - Latest
76		INDU.UK.TOT.COF.BS.M	Monthly	Yes	2000 - Latest
77		RETA.EU_2019.TOT.COF.BS.M	Monthly	Yes	2000 - Latest
78		RETA.IE.TOT.COF.BS.M	Monthly	Yes	2000 - Latest
79		RETA.UK.TOT.COF.BS.M	Monthly	Yes	2000 - Latest
80		SERV.EU_2019.TOT.COF.BS.M	Monthly	Yes	2000 - Latest
81		SERV.IE.TOT.COF.BS.M	Monthly	Yes	2000 - Latest
82		SERV.UK.TOT.COF.BS.M	Monthly	Yes	2000 - Latest
83		GDP SA (MP) European Union - 27 countries (from 2019)	Quarterly	Yes	2000 - Latest
84	International	GDP SA (MP) United Kingdom	Quarterly	Yes	2000 - Latest
85		DP SA (MP) United States	Quarterly	Yes	2000 - Latest