Forecasting New Zealand's labour market with multiple datasets

Peter Gardiner[a], Kirdan Lees[b]

[a] *New Zealand Treasury*
[b] *Reserve Bank of New Zealand, 2 the Terrace, Wellington 6011, New Zealand ph +64 4 4713666*

Abstract

JEL classification: C52; C53; E17; J21

** *Work in Progress* **

Individual sources of information on New Zealand's labour force can be volatile and subject to different short-run interpretations. To forecast New Zealand's unemployment rate we use a dynamic factor model to extract information from a wide range of data including key surveys, wage data, productivity data, demographics and the real economy, rather than focussing on a small number of data series. Moreover, we utilise the state-space representation of the dynamic factor model that allows the model to be estimated on mixed frequencies of data and at any point in time such that the forecasts can be updated before or after major data releases or immediately prior to policy or budgetary decisions. In addition to this flexibility, we find the model has similar if not better forecasting performance relative to simple time series benchmarks and forecasts competitive with the Reserve Bank of New Zealand, particularly over the medium term. Finally, we augment the procedure by constructing forecasts of the unemployment rate indirectly, from forecasts of the number of unemployed and New Zealand's labour force which appears to help improve the dynamic factor model forecasts in the near term.

Email addresses: `pete.gardiner@treasury.govt.nz` (Peter Gardiner), `kirdan.lees@rbnz.govt.nz` (Kirdan Lees)

1. Introduction

New Zealand's labour market data has proved particularly volatile in recent times, perhaps as a result of the impact of the Global Financial Crisis but perhaps also as the result of the treatment of seasonality and underlying structural changes in New Zealand's demographics and industrial profiles. Of course, these factors are interrelated and a framework that distills information from a wide range of data should help in building a forecast of the state of the labour market.

This paper uses a Dynamic Factor Model (DFM) to synthesize data from several sources to provide both a forecasts for unemployment in addition to a 'nowcast' that provides insight into the current state of the New Zealand labour market. Moreover we use the framework to examine whether constructing an unemployment forecast via participation and employment is superior to a direct forecast of unemployment.

The DFM approach has been used in several papers to produce forecasts for GDP growth using large datasets (see for example Baffigi et al. (2004) and Giannone et al. (2008))) and more recently real-time smoothed growth nowcasts for most of the countries studied in the IMF World Economic Outlook (see Matheson (2011)) and Martinsen et al. (2011) report forecasts of unemployment using Dynamic Factor Models for Norway.

We use several data sources. Key sources of data include the Household Labour Force Survey, the Quarterly Employment Survey, in addition to key wage measures such as the Labour Conditions Index, measures of labour tightness in the Quarterly Survey of Business Opinion and a range of demographic data. Moreover, we use a range of hard and survey data sources on the state of the New Zealand business cycle.

We find that the DFM produces a relatively smooth unemployment forecast that performs well against simple benchmarks. It appears that forecasting unemployment directly rather than indirectly via participation and employment growth is marginally superior, at least according to the data sample we examine. In addition, we explore whether constructing forecasts of the unemployment rate indirectly from forecasts of the size of the labour force and the number of unemployed.

Finally, we examine the behaviour of our unemployment forecasts over 2010 to understand how the forecasts update with the arrival of new data series.

Section 2 presents the methodology of our dynamic factor model while section 3 discusses our data sources, including the timing of the arrival of data. Section 4 presents results of the forecasting comparisons. Concluding comments are made in section 5.

2. Methodology

2.1. Dynamic factor model

Dynamic factor models decompose the target series (here we focus on a univariate case, that could be real GDP growth for example) into a common component χ_t and an idiosyncratic component ϵ_t. The common component aims to encapsulate the variation between the target series and a wide range of economic indicators.

$$y_t = \mu + \chi_t + \epsilon_t \tag{1}$$

where μ is a constant, $\epsilon_t \sim N(0, \psi)$ and the common component $\chi_t = \Lambda F_t$ where $F_t = (F_{1t}, ..., F_{rt})'$ and $\Lambda = (\lambda_1, ..., \lambda_r)$ such that the target series is related to the indicators via a linear combination of a small number of r factors. The dynamics of the factors are given by the vector autoregressive process:

$$F_t = \sum_{i=1}^{p} \beta_i F_{t-i} + B\nu_t \tag{2}$$

where $\nu_t \sim N(0, I_q)$, the β_i matrices are $r \times r$ matrices, p is the lag length of the process and the matrix B, is a $p \times q$ matrix with q the number of underlying common shocks that drive the target series. Note that equations 1 and 2 give a state-space representation of the Dynamic Factor model that can be used for Kalman filtering.

2.2. The jagged-edge problem and bridging equations

The advantage of the state space set-up outlined in equations 1 and 2 is that the common component of the target series can be estimated even when the indicators have missing values at the end of sample due to the periodic arrival of indicators. Crucially this implies that the setup can incorporate all available information in a timely manner. For example, assume that the forecaster's task is to forecast the target series y_t based on the information set $\Omega^n_{t,j}$, where n is the number of indicators available and j represents the vintage of data within the lowest frequency of data, that is:

$$\hat{y}_{t,h,j} = Proj(y_t \mid \Omega^n_{t,j}) \tag{3}$$

where j represents the jth from $j = 0, 1..., J$ data release for the target variable. In practice we think of updating the estimate for the target variable at the monthly frequency on a daily basis, in which case j runs from 1 to 31. This implies that we can also compute the "news" component of a specific data release as:

$$NEWS(Omega^n_{t,j}) = Proj(y_t \mid \Omega^n_{t,j}) - Proj(y_t \mid \Omega^n_{t,j-1}) \tag{4}$$

Note that under our setup, data arrival reflected in the jth vintage influences the projection via both the information content and the updating of parameter estimates. This enables the model to be run at any point at time while maximising the information content of available data. In principle, this helps the model improve on forecasts that rely heavily on outturns of the forecast variable in question. Moreover, this allows the modeller to produce forecasts updated prior to particular data releases or immediately prior to key policy, budgetary or financial market decisions.

2.3. Determining the number of factors to employ

A key decision is the number of factors to employ. Bai & Ng (2002) suggest using a two-step procedure to determine the number of factors while others (see for example Giannone et al., 2008; Forni et al., 2001, 2005) suggest increasing the number of factors r until a fixed proportion of the variation in the target variable is explained (often 80 percent). However, Matheson (2011) notes variation in the explanatory of factor models across countries and adopts a rule where factors are added until the marginal increase in R-squared from regressing the target series on the common factors is less than 2.5 percent and we adopt that rule here.

3. Data and models

3.1. Selecting indicator data

One decision in using indicator series to forecast a target series is deciding which indicator series to use. We are guided by both Stock & Watson (2002) and Boivin & Ng (2006) who suggest not using all available indicators and using judgment to ensure key categories of data with specific macroeconomic properties are populated. But since we want to use forecast the unemployment

Table 1: Key Labour Market Data features

	Quarterly Employment Survey
Measures	Quarterly changes in average hourly/weekly earnings, paid hours and jobs filled.
Target population	Firms with any Full-time employees (FTEs).
Stratification	Industry and fte size based, and no sample rotation
Exclusions	Agric., agric. services, fishing, int'l sea transport, res property owners, pvt h/holds employing staff, businesses with $< 30k$ of sales
Stratification	Area based, Primary Sampling Units over selected for Maori
Target R/rate	85 percent

	Household Labour Force Survey
Measures	labour market attachment, primarily employment, unemployment and NILF at national level
Target population	civilian non-institutionalised usually resident Nzers 15+ yrs
Stratification	Household survey, 15,000 h/holds or about 30,000 individuals interviewed each quarter
Exclusions	Non-private dwellings, hospital patients, inmates, armed forces, retirement home residents
Stratification	area based, Primary Sampling Units over selected for Maori
Target R/rate	90 percent

rate, we ensure a wide range of labour market and demographic data is brought to bear on the forecasting exercise. Here, we outline key features of sources of labour market information.

There is little doubt that the New Zealand labour market has been variable during the past couple of years. This variability has three potential sources: real world, sampling and non-sampling. This suggests using a variety of source data from which to construct forecasts. We can test whether recent labour market volatility is real world by looking at extent to which estimates of the same or similar variables are related. There are three primary sources of labour market statistics in New Zealand: the Household Labour Force Survey (HLFS), Quarterly Employment Survey (QES) and the Linked Employer-Employee Database (LEED). While LEED provide deep insights into the networks that underpin New Zealand's labour market, the delay with which they are available preclude their usefulness for forecasting purposes.

3.2. Key Labour Market Data

With a high degree of correlation in measures of employment from different sources, we can have some confidence in the robustness of estimates from these statistics. Therefore, it is important when considering the state of the New Zealand labour market to try and look through the quarterly variability to attain a medium term picture. The appendix to the working version of the paper contains a complete listing of the indicators employed in our indicator.

3.3. The candidate models

To test the out-of-sample forecast performance of the dynamic factor model, we construct forecasts from two simple statistical benchmarks and the forecasts from the Reserve Bank of New Zealand *Monetary Policy Statements*. The first benchmark model is an autoregressive (AR) that contains up to four lags determined by the Bayesian information criteria. The second benchmark model consists of a simple Bayesian VAR (BVAR) model with four lags that includes output, unemployment, interest rates, inflation the exchange rate with a time series prior that drives the parameter values on differences towards zero. Both time series benchmarks are relatively standard. In addition to the time series models, we compare the forecasts from the Dynamic Factor Model with the forecasts for unemployment contained within the Reserve Bank of New Zealand's *Monetary Policy Statements*. This forms a hard test for the Dynamic Factor Model since the Reserve Bank has access to a wide range of models and can add judgment based on a nuanced understanding of labour market data.

Finally, changes in the unemployment rate can be driven by large changes in the participation rate (see Stephens (2007) for the case of New Zealand). Changes in labour market conditions, but also low frequency demographic trends can alter the participation rate can drive changes in the unemployment rate. Here we test whether of unemployment forecasts can be improved by constructing unemployment forecasts from factor model forecasts of the size of the labour force minus factor model forecasts of the number of employed worked.

We estimate all models over the period starting from 1992Q1 avoiding much of the volatility in the macroeconomic data in the late 1980s during the period of labour and good market reforms. We retain the final six years 2005Q1 - 2010Q4 to test the performance of each model relative to the root mean squared errors delivered by the naive random walk model that simply takes the last observable unemployment rate as the forecast.

Table 2: Out-of-sample forecasting performance: RMSE relative to random walk model

h-step ahead	RW	AR	BVAR	GLUM	RBNZ	AR-P	BVAR-P	GLUM-P
1	1.00	**0.61**†	**0.63**†	0.73	0.89	**0.59**	**0.65**†	**0.54**†
2	1.00	0.77	0.74	**0.51**	0.63	0.76	0.76	**0.41**†
3	1.00	0.82	0.81	**0.50**†	**0.57**	0.81	0.83	**0.43**†
4	1.00	0.92	0.94	**0.56**	0.61	0.92	0.95	**0.53**
5	1.00	0.89	0.91	0.64	**0.62**†	0.90	0.93	0.64
6	1.00	0.95	1.00	**0.67**	**0.65**†	0.96	0.98	**0.67**
7	1.00	0.97	1.05	0.70	**0.67**†	0.98	1.02	0.72
8	1.00	1.01	1.13	0.76	**0.69**†	1.02	1.09	0.78

Note the following mnemonics: random walk model (RW), autoregressive model (AR), Bayesian VAR model (BVAR), Gardiner-Lees unemployment model (GLUM), Reserve Bank of New Zealand unemployment forecasts (RBNZ), autoregressive models with participation (AR-P), Bayesian VAR model with participation(BVAR-P), and Gardiner-Lees model with participation (GLUM-P).
Bold font indicates significance at the 10% level and the † symbol indicates significance at the 5% level on one-sided tests of forecast performance better than the random walk model.

4. Results

4.1. Out-of-sample forecasting

We calculate the bias and root mean-squared errors for each model and calculate t-stats based on the Diebold-Mariano test for forecast significance in small samples The table below shows the results of the out-or-sample forecasting exercise.

Table shows that all the models add value relative to a naive random walk forecast. The AR and BVAR forecasts before well in the near-term, returning lower RMSEs that the Reserve Bank of New Zealand for the next quarter. However, the Reserve Bank's longer term forecasts perform well, significantly outperforming the random walk model based on the Diebold-Mariano test statistics. The GLUM factor model also performs very well over the medium term and also outperforms the random-walk model.

The final three columns of the table show the performance of the models that compute forecasts of the unemployment rate by constructing the unemployment rate from separate forecasts of labour force participation and the number of unemployed. For the AR and BVAR models, similarly forecast performance is delivered with strong near term performance relative to the counterparts in the first columns of the table that operate directly on the unemployment rate forecast. However, for the GLUM model, including participation by constructing dynamics forecasts of the labour force and then unemployment appears to have increased the accuracy of the factor model forecasts.

5. Conclusion

Labour market data in New Zealand has been particularly volatile. We explore a Dynamic Factor model setup that allows for data of mixed frequency and moreover, missing observations at the end of history that allow the model to be produce forecasts using all information available at any date. We show that a Dynamic Factor model that has competitive forecasting ability relative to simple statistical benchmarks and forecasting from the Reserve Bank of New Zealand's *Monetary Policy Statements*. In addition, constructing dynamic factor model unemployment forecasts indirectly from forecasts of the size of the labour force and the number of unemployed appears to help forecast the unemployment rate, particularly in the near term. Future work could use the Dynamic Factor model setup to examine the impact of the arrival of key data on the forecasts of the unemployment rate.

6. Appendix

Following Matheson (2011), we apply the following scheme to tidy the dataset prior to estimation of the dynamic factor model:

1. Missing values within the sample are linearly interpolated.
2. The seasonal series are adjusted using X11.
3. Quarterly and annual series are interpolated to the monthly frequency using linear interpolation; the daily and weekly series are converted into monthly averages.
4. Log quarterly differences are taken of the non-stationary series, except those that are measured in percentages or can take negative values, in which case quarterly differences are taken. The remaining series are left as levels.
5. The series that only change 10 percent of the time are discarded.
6. The series with less than 3 years worth of data are discarded.
7. The series not released in the past year are discarded (to avoid discontinued data).
8. Outliers are removed, where observations greater/less than 6 times the interquintile range are replaced with the next highest/lowest admissible value.
9. Missing observations at the beginning of the sample are backdated using the DFM, with the number factors set to explain 60 percent of the variation in the data.

Mnemonic	Brief Description
AUSELUR@ANZ	Australia Total Labor Force Unemployment Rate (SA, %)
AUSNGDPC@ANZ	Australia Gross Domestic Product (SA, Mil.Chn.Q308-Q209.A$)
AUSWWLT@ANZ	Australia Avg Weekly Earnings FT Adult Total Earnings (SA, A$)
NZSRMFC@ANZ	New Zealand Roy Morgan Survey Future Conditions Index (SA, Index)
NZNPH@ANZ	New Zealand Housing Price Index All Households (NSA, Q4-03=1000)
NZNPOW@ANZ	New Zealand ANZ World Commodity Price Index (NSA,Jan-86=100)
NZNXTWIN@ANZ	New Zealand Exchange Rate Nominal Trade Weighted Index (1970-2008=100)
NZEFCP@ANZ	New Zealand Credit Aggregates Private Sector Credit (Mil.NZ$)
NZNFA@ANZ	New Zealand Loans to Agriculture Sector (EOP, NSA, Mil.NZ$)
NZNPOP@ANZ	New Zealand Population (Thous)
NZNPM@ANZ	New Zealand Net Migration (Persons)
NZNPD@ANZ	New Zealand External Migration Departures (Persons)
NZNPA@ANZ	New Zealand External Migration Arrivals (Persons)
NZSHW@ANZ	New Zealand Work Put in Place All Buildings (SA, Mil.NZ$)
NZSTRS@ANZ	New Zealand Retail Sales (SA, Mil.NZ$)
NZNFKG@ANZ	New Zealand Gross Index NZSX All Indexes (Jun-30-86=1000)
NZNVOUR@ANZ	New Zealand Business Outlook Survey Unemployment Rate Retail (%)
NZNVOUM@ANZ	New Zealand Business Outlook Survey Unemployment Rate Manufacturing (%)
NZNVOUA@ANZ	New Zealand Business Outlook Survey Unemployment Rate Agriculture (%)
NZNVOUC@ANZ	New Zealand Business Outlook Survey Unemployment Rate Construction (%)
NZNVOUS@ANZ	New Zealand Business Outlook Survey Unemployment Rate Services (%)
NZNRB3@ANZ	New Zealand Bank Bill Yield 90 days (% per annum)
CEXP@USECON	University of Michigan Consumer Expectations (NSA, Q1-66=100)
SP500@USECON	Stock Price Index Standard and Poor's 500 Composite (1941-43=10)

ARSR1Z@act	Nom. Retail Sales - Core industries total s.a.
ARASRZ@act	Nom. Retail Sales - All industries total s.a.
ARRDURZ@act	NOM. RETAIL SALES - DURABLES - S.A.
ARRNDURZ@act	NOM. RETAIL SALES - NON-DURABLES - S.A.
ARRSERVZ@act	NOM. RETAIL SALES - SERVICES - S.A.
ACTRZ@act	NEW VEHICLES - NEW CARS and STATION WAGONS - INC CARS PREV REG OVE
AHDAYSALZ@act	REINZ - Median days to sell s.a.
AHSALEDZ@act	REINZ - No of Dwelling Sales - s.a.
ABFALIST@act	Barfoot and Thompson - Available listings - Auckland
ABFHSALE@act	Barfoot and Thompson - No of House sales - Auckland
ADVPH@act	Consents - DWELLINGS - HOUSES and FLATS - NEW - VALUE - TOTAL NZ
ADNNAPZ@act	Consents - DWELLINGS - NON-APARTMENT DWELLING UNITS - NUMBER
ABVPTZ@act	Consents - OTHER BUILDING - NEW ALT - VALUE -TOTAL NZ - SA
ABNPTZ@act	Consents - OTHER BUILDING - TOTAL NEW, ALT - NUMBER - NZ - SA
AWTSTKMZ@act	Wholesale Trade, Stocks of Materials, Total s.a.
AWTSTKFZ@act	Wholesale Trade, Stocks of Finished Goods, Total s.a.
AWTOI@act	Wholesale Trade, real operating income
EBEACNZ@exp	QSBO, ECONOMY-WIDE, NEXT 3 MONTHS, Average Costs s.a.
EBEACZ@exp	QSBO, ECONOMY-WIDE, PAST 3 MONTHS, Average Costs s.a.
EBEASPZ@exp	QSBO, ECONOMY-WIDE, PAST 3 MONTHS, Av. sell pr.
EBECU@exp	QSBO, ECONOMY-WIDE, Capacity Utilisation
EBEDTANZ@exp	QSBO - ECONOMY-WIDE - NEXT 3 MONTHS - Domestic Trading Activity
EBEFLUZ@exp	QSBO - ECONOMY-WIDE - Find labour unskilled (s.a.)
EBEGBOZ@exp	QSBO - ECONOMY-WIDE - General bus situation (s.a.)
EBEIBZ@exp	QSBO - ECONOMY-WIDE - New investblgds (s.a.)
EBEIPMZ@exp	QSBO - ECONOMY-WIDE - New invest P and M

EBELTN@exp	QSBO - ECONOMY-WIDE - NEXT 3 MONTHS - Lab. Turnover
EBENENZ@exp	QSBO - ECONOMY-WIDE - NEXT 3 MONTHS - No.s emp (s.a.)
EBEOWNZ@exp	QSBO - ECONOMY-WIDE - NEXT 3 MONTHS - Overtime Wkd
EBEPRFNZ@exp	QSBO - ECONOMY-WIDE - NEXT 3 MONTHS - Profitability (s.a.)
ENAOT@exp	NATIONAL BANK - ACTIVITY OUTLOOK - Next 12 Months - Total
ENBCT@exp	NATIONAL BANK - BUSINESS CONFIDENCE - Next 12 Months - Total (
ENCR@exp	NATIONAL BANK - Residential Construction Intentions - Next 12 months
ENCUT@exp	NATIONAL BANK - CAPACITY UTILISATION - Total (All Sector)
ENE@exp	NATIONAL BANK - EMPLOYMENT INTENTIONS - Next 12 Months - Total
ENIET@exp	NATIONAL BANK - INFLATION EXPECTATIONS - Next 12 Months - Total
ENIIT@exp	NATIONAL BANK - INVESTMENT INTENTIONS - Next 12 Months - Total
ENP@exp	NATIONAL BANK - PROFIT EXPECTATIONS - Next 12 Months - Total
ENPI@exp	NATIONAL BANK - PRICING INTENTIONS - Next 3 Months - Total
ENUR@exp	NATIONAL BANK - UNEMPLOYMENT RATE - Next 12 Months - Total

7. References

Baffigi, A., Golinelli, R., & Parigi, G. (2004). Bridge models to forecast the euro area gdp. *International Journal of Forecasting*, *20*(3), 447–460.

Bai, J. & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, *70*(1), 191–221.

Boivin, J. & Ng, S. (2006). Are more data always better for factor analysis. *Journal of Econometrics*, *132*, 169–94.

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2001). The generalized dynamic factor model: identification and estimation. *Review of Economics and Statistics*, *82*, 540–554.

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association*, *100*(471), 830–40.

Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, *55*(4), 665–76.

Martinsen, K., Ravazzolo, F., & Wulfsberg, F. (2011). Forecasting macroeconomic variables using disaggregate survey data. *Norges Bank Working Paper Series*, (2011-04).

Matheson, T. (2011). Indicators for tracking growth in real-time. *IMF working paper*.

Stephens, D. (2007). Labour market participation in new zealand. *Westpac Occasional Paper*.

Stock, J. H. & Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, *20*(2), 147–62.