CENTRE FOR APPLIED MACROECONOMIC ANALYSIS
The Australian National University

THE AUSTRALIAN NATIONAL UNIVERSITY

CAMA Working Paper Series **September, 2007**

SIX LEADING INDEXES OF NEW ZEALAND EMPLOYMENT

Edda Claus
The Australian National University
Trinity College, Dublin

Iris Claus
The Australian National University
New Zealand Inland Revenue

CAMA Working Paper 17/2007
http://cama.anu.edu.au

Six Leading Indexes of New Zealand Employment*

by

Edda Claus

Centre for Applied Macroeconomic Analysis and Institute for International Integration Studies, The Sutherland Centre, Level 6, Arts Building, Trinity College, Dublin 2, Ireland. Telephone: +353 1 896 3201. Fax: +353 1 896 3939. Email: clause@tcd.ie. Corresponding author.

Iris Claus

Centre for Applied Macroeconomic Analysis and Inland Revenue, Policy Advice Division, P.O. Box 2198, Wellington, New Zealand. Telephone: +64 4 978 6028. Fax: +64 4 978 1623. Email: iris.claus@ird.govt.nz.

*The paper builds on a project funded by the New Zealand Department of Labour. The views expressed in this paper and remaining errors are those of the authors and should not be attributed to the New Zealand Department of Labour or Inland Revenue.

Abstract

This paper constructs six leading indexes of New Zealand employment and compares their short term forecasting performance. Forecasting New Zealand employment is particularly difficult owing to the volatility of the data and the short sample size of available time series. These restrictions make leading indexes especially appealing. The paper has two aims. The first is to construct an effective forecasting tool. The second is to evaluate leading indexes constructed using different methods available in the literature. The results show that an index constructed using the traditional NBER method dominates in terms of forecasting performance. The results also suggest that increasing the dataset does not strengthen the index and that exogenously determining the weights of component series can add to forecasting performance.

Key Words: Composite index of leading indicators; Diffusion index; Kalman filter; Principal component analysis; Employment; Forecasting

1 Introduction

This paper constructs six leading indexes of New Zealand employment. It has two aims. The first is to build a forecasting tool for policy makers. The second is to assess the relative performance of leading indexes constructed using various methods available in the literature.

Forecasting New Zealand variables is particularly difficult owing to short sample sizes of available data. Moreover, time series tend to be volatile due to the small size of the market and the characteristics of the economy. New Zealand is geographically isolated, being separated from Australia by approximately 2000 kilometers. Its closest neighbors are New Caledonia, Fiji, and Tonga. New Zealand has a population of around 4.1 million people and relatively open immigration policies. It is heavily dependent on trade, particularly in agricultural products (dairy products, meat, forest products, fruit and fish), and exports roughly 30 percent of its output. This makes New Zealand vulnerable to weather and global economic conditions and international commodity prices. New Zealand's exports are relatively diffuse. In 2006 its major merchandise export partners were Australia (21 percent), the United States (13 percent), Japan (10 percent), China (5 percent), and the United Kingdom (5 percent).[1]

Data restrictions (short sample sizes and volatile time series) make the use of leading indicators particularly appealing. Short sample sizes limit the number of variables that can be included in a structural model to forecast employment. Autoregressive (AR) or autoregressive integrated moving average (ARIMA) models are an alternative forecasting tool but may be less appealing particularly in a policy making environment as these models ignore all outside information. Though leading index models are parsimonious they draw on a wide range of information.

Leading indexes have long been widely used. Examples of organizations relying on leading indexes are the US and Canadian Conference Boards, the Federal Reserve Banks of Chicago, Dallas, Kansas City, and Philadelphia, Statistics Canada, the Canadian Department of Finance, the New Zealand Department of Labour, the Australian Department of Employment and Workplace Relations, the European Central Bank, and the Organisation of Economic Co-operation and Development (OECD). The construction of leading indexes typically focuses on output and to a lesser extent on inflation while relatively little attention has been paid so far in the literature on the forecasting performance of leading indexes of employment.[2] Output and employment share many similar characteristics. Both are cyclical series and forecasting quarterly changes typically involves careful consideration of all sectors of the economy. Generally, the performance of an index is assessed compared to other types of forecasting models, such as an AR or structural model rather than other types of leading indexes.

Since the introduction of leading indexes by W. C. Mitchell and A. F. Burns during the 1930s and 1940s two main approaches have emerged in the literature. These are indexes that extract information from a large set of variables and those that rely only on a few selected series.

We construct six leading indexes of New Zealand employment ($lie1, ... \ lie6$) for the period 1990Q1 to 2005Q3 using the various methods available in the literature. $lie1$ is a simple index, solely relying on the direction of change of all series in the dataset.

*lie*2 is a classical composite index of leading indicators in the tradition of the National Bureau of Economic Research (NBER). *lie*3 and *lie*4 are hybrids. *lie*3 includes those series identified by the NBER method but uses a kalman filter to extract the common component from those series. *lie*4 extracts the common component of all series within one sector using a kalman filter and accumulates the extracted components into one overall index. *lie*5 and *lie*6 use principal component analysis on the entire dataset. The only difference between *lie*5 and *lie*6 is that the dataset of the latter also includes each series at lags 1 and 2. Though their construction method is different, *lie*1 and *lie*5/*lie*6 are called diffusion index in the literature (see, for example, Kennedy 1994, Stock & Watson 2002*b*). To avoid confusion, we refrain from using the term diffusion index.

The results show that the index constructed using the traditional NBER method dominates in terms of forecasting performance. The results also suggest that increasing the dataset does not strengthen the index and that exogenously determining the weights of component series can add to forecasting performance.

The paper is organized as follows. Section 2 outlines the data and methodologies and section 3 presents the six leading indexes of New Zealand employment. Section 4 discusses the forecasting performance of each index when included in a simple short term forecasting model. Section 5 offers some concluding remarks.

2 Data and methodologies

In broad terms, constructing a leading index focuses on finding variables that tend to be affected by the same shocks as the reference series, in this case employment, but with a lead. The approach focuses on co-movements and does not attempt to establish causal relationships between series. A leading index should be diversified and broadly cover the different sectors of an economy with a minimum of duplication.

The dataset used here includes 95 quarterly variables that can be divided into seven broad categories. These are: (i) labor market indicators, (ii) domestic activity indicators, (iii) trade indicators and commodity prices, (iv) foreign activity indicators, (v) consumer and business confidence indicators, (vi) financial variables, and (vii) monetary variables. All series are quarterly and cover the sample period 1990Q1 to 2005Q3. Appendix A lists all 95 variables.

Component series are denoted by Y_t^j. Most variables Y_t^j are transformed and standardized to ensure symmetrical treatment of positive and negative changes, or $X_t^j = 200 * \frac{Y_t^j - Y_{t-1}^j}{Y_t^j + Y_{t-1}^j}$. For series that contain zeros or negative values, or that are already in index or percentage form, the simple first difference is used ($X_t^j = Y_t^j - Y_{t-1}^j$) while for series that are a difference of two series (such as yield spreads) or a ratio the raw series is used ($X_t^j = Y_t^j$). For *lie*3 to *lie*6, the data are also demeaned and standardized to have unit variance.

We apply four methods commonly employed in the literature and construct two hybrid indexes. The first leading index of employment (*lie*1) includes all 95 variables. Following Kennedy (1994) and The Conference Board, it solely relies on the direction of change of the component series. An individual series is given a value of 1 if it increased, 0.5 if it remained unchanged, and 0 if it decreased in a particular quarter.[3] For the final index, the values of the component series are summed up at each point in time, divided by the total number of series in the dataset, and multiplied by 100.

*lie*2 is a classical composite index of leading indicators in the tradition of the NBER. In broad terms, the NBER approach aggregates a number of series into one overall index. This involves two main issues. First, which series should be included as components and second, how should the series be cumulated into an index. Typically series are included because they track employment cycles, not because they are the operational counterparts of variables in an economic theory of employment. Here the decision for inclusion or exclusion of each series is primarily based on statistical significance, forecasting performance and timeliness with which the data are released relative to the employment series. Building the composite index of leading indicators also requires a choice for the weights of the individual components. The NBER scores the component series by economic significance, statistical adequacy, historical conformity to cycles, cyclical timing record, smoothness, and promptness of publication. The components with the highest score receive the largest weight.[4] The approach used here follows Claus & Claus (2002) and is based on the concordance of component series with employment; see Harding & Pagan (2002). The concordance statistic is calculated as the number of times the component series and employment moved in the same direction or remained unchanged. As this study builds a leading index, the concordance statistic is constructed for 1 to 4 lags[5] and the maximum value of the four quarters is chosen. For all variables in the index, the maximum value is at lag 1.[6]

The final *lie*2 index includes six series, the job vacancy rate (ANZ Bank), a survey measure of labour as the limiting factor of business activity (Quarterly Survey of Business Opinion, New Zealand Institute of Economic Research), company tax (actual cash receipts, New Zealand Treasury), migration (permanent and long term arrivals, Statistics New Zealand), Southern oscillation index (National Institute of Water and Atmospheric Research), and the New Zealand stock market index (nz10cap, Datastream). All series are released before employment which makes the index particularly useful for forecasting.

Probably, the two most contentious issues of the NBER method is the use of exogenously determined weights and the selection of series for the final index. Stock & Watson (1989) propose an alternative to the NBER method and model a leading index in a latent factor framework. This approach is used in the construction of the third and forth index, $lie3$ and $lie4$. It applies the kalman filter to numerically extract the common components of various series. $lie3$ and $lie4$ are modelled as

$$X_t^j = \lambda^j(L) lie_t + u_t \tag{1}$$

$$lie_t = \theta(L) lie_t + n_t \ , \tag{2}$$

where $lie_t = lie3_t \wedge lie4_t$, $t = 1, ...T$, and $j = 1, ...N$. L is the lag operator $(L > 0)$. $\lambda^j(L)$ and $\theta(L)$ are vector lag polynomials and $|\theta_i| < 1$. u_t are the idiosyncratic factors with $E[u_t n_\tau] = 0 \ \forall t, \tau$ where τ is a time subscript and $E[u_t u_t] = 1$. n_t is a white noise error. In the final indexes, $L = 1$.

$lie3$ is the result from applying the kalman filter to the six series included in $lie2$. $lie3$ is broadly in line with Stock & Watson (1989). We extract a labour market leading index from the six series that is then included in a single equation forecasting model of employment growth. Stock & Watson (1989) extract a coincident (business cycle) index from four series. The leading index is a six months forecast of the coincident index generated by estimating a vector autoregression (VAR) model that includes the coincident index and seven observed leading variables.[7]

*lie*4 is constructed by applying the kalman filter to the variables of each of the seven sectors. The seven resulting series are then cumulated into one index by using the same concordance method as for *lie*2. It would be useful to apply the filter to all 95 variables. However, standard computing capabilities do not allow the filter to be applied to all 95 series.

*lie*5 and *lie*6 are also built in a latent factor framework. Their construction follows Stock & Watson (2002*a*, 2002*b*) and Brisson, Campbell & Galbraith (2003) and applies principal component analysis to the entire dataset. The non-linear least squares function of equation (1) is:

$$V\left(\tilde{lie}, \tilde{\lambda}\right) = (NT)^{-1} \sum_i \sum_t \left(X_t^j - \tilde{\lambda} \tilde{lie}_t\right)$$

where N is the number of variables. *lie*5 is a $(T \times k)$ matrix including the first (in terms of the largest eigenvalue) k eigenvectors of $N^{-1}(XX')$ where $X_t^j \in X$ (see Stock & Watson 2002*a*). The only difference between *lie*5 and *lie*6 is that the dataset for *lie*6 also includes each of the series at lags 1 and 2; see Brisson et al. (2003).

3 Six leading indexes of employment

The six leading indexes are estimated in GAUSS 6.0. The kalman filter is estimated using the MAXLIK procedure with the BFGS algorithm. The eigenvectors for *lie*5 and *lie*6 are from the eigrs2 procedure and $k = 1$ is chosen by minimizing the Akaike and Bayesian information criteria from regressing j eigenvectors ($j = 1$ to 10) on employment growth. All indexes, except *lie*1 are standardized so that each index has the same historical average as employment.

Table 1 summarizes the six indexes. Column (1) shows the index name, column (2) the construction method and column (3) gives the number of variables included.

Table 1: Summary table for six leading indexes

Index name (1)	Method (2)	Number of variables (3)
*lie*1	scoring component series changes	95
*lie*2	composite index	5
*lie*3	hybrid with kalman filter	5
*lie*4	hybrid with kalman filter; 2 steps	95
*lie*5	principal component analysis	95
*lie*6	principal component analysis	3·95

Figures 1a to f plot the six indexes for 1990Q1 to 2005Q3. All indexes except *lie*1 and *lie*6 are set to 100 in 1990Q1. *lie*1 is a percentage share and *lie*6 is set to 100 in 1990Q3. The shaded areas in each figure indicate employment downturns from a peak to a trough. A peak is defined as the quarter prior to at least two consecutive declines in employment and a trough is defined as the quarter prior to at least two consecutive increases in employment. The figure shows two downturns in employment over the sample period, 1990Q2 to 1991Q4 and 1996Q3 to 1998Q4. For the indexes, a signal of a cyclical downturn is defined as two consecutive declines in the index following at least two consecutive rises while a cyclical upturn is defined as two consecutive rises following at least two consecutive falls in the index. A false signal is defined as two consecutive rises (declines) in the index that are not followed by a cyclical upturn (downturn) in employment.

Figures 1a to f show that the six indexes are useful though with varying degrees leading cyclical indexes. In terms of early warnings for cyclical turning points, $lie2$ (Figure 1b) and $lie4$ (Figure 1d) outperform the others. The indexes signal the cyclical downturn in 1990 and the upturns in 1992 and 1998 with a lead of one quarter. $lie2$ signals the 1996 downturn with a lead of two quarters and $lie4$ with a lead of ten quarters. $lie2$ gives three and $lie4$ two false signals; $lie2$ in 1996, 2000, and 2003 and $lie4$ in 2000 and 2004. Barring the 1990 downturn, $lie3$ (Figure 1c) signals turning points with a lag and gives three false signals. $lie5$ (Figure 1e) and $lie6$ (Figure 1f) appear to be countercyclical. If both indexes are taken to be countercyclical $lie5$ gives advance signals of the 1990 downturn, the 1991 and 1999 upturns with a lead of one quarter, and the 1996 downturn with a lead of seven quarters. The index gives three false signals in 2000, 2003 and 2004. $lie6$'s signals are more coincident rather than leading. $lie1$ (Figure 1a) is more volatile than the other five indexes limiting its usefulness as a cyclical indicator.

Table 2 gives the correlations between the growth rates in five of the six leading indexes ($lie1$ is a share) and that of employment at lags zero to four quarters. $lie2$ exhibits the highest correlation in a single quarter, that is 0.449. Reflecting the counter-cyclicality of $lie5$ and $lie6$ the correlations are negative at every lag.

Figure 1: Six leading indexes

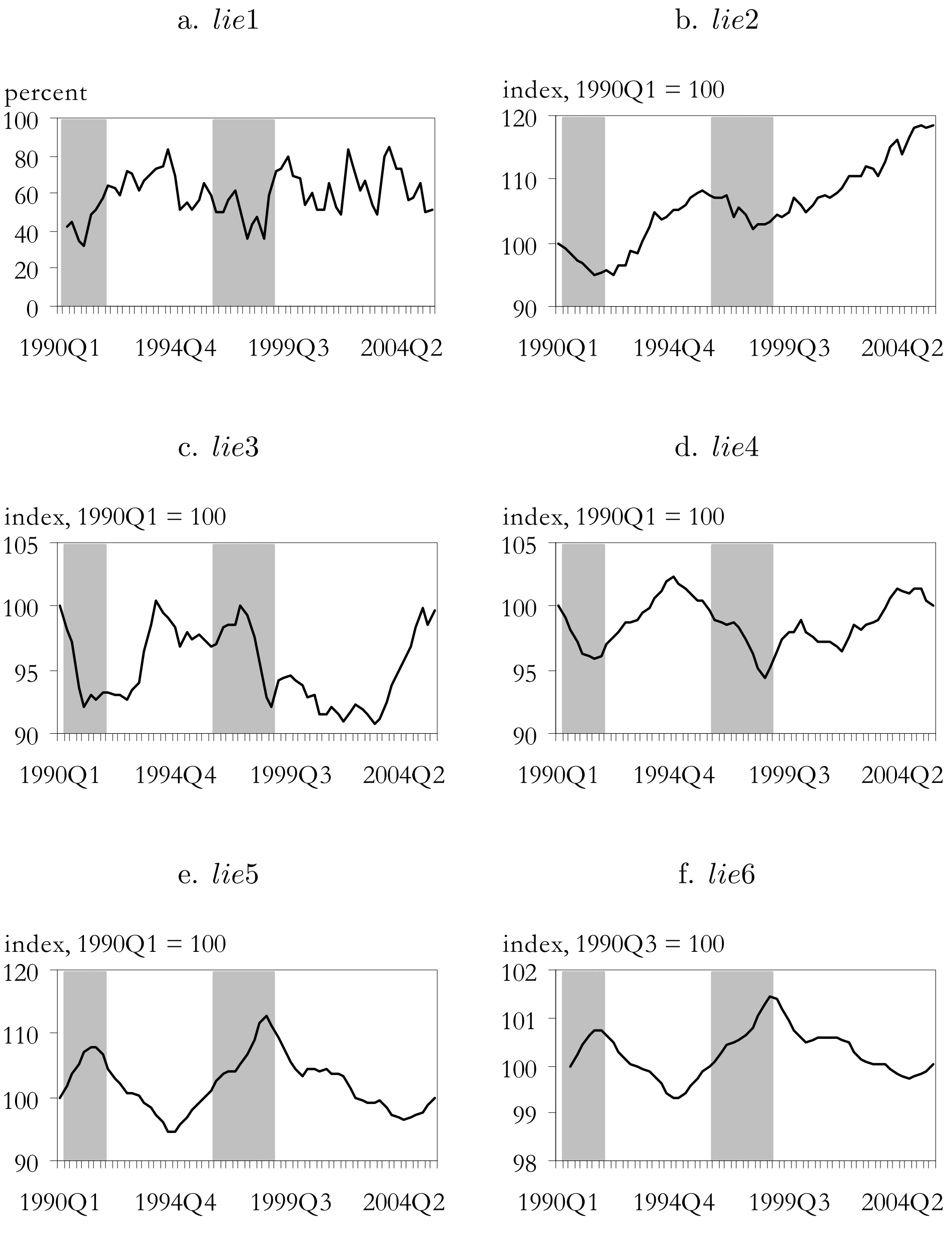


Table 2: Correlations with employment growth

Index	——— Employment growth ———				
	t	$t-1$	$t-2$	$t-3$	$t-4$
(1)	(2)	(3)	(4)	(5)	(6)
$lie1$	0.345	0.264	0.352	0.404	0.324
$lie2$	0.252	0.279	0.133	0.449	0.342
$lie3$	0.239	0.101	0.089	0.355	0.184
$lie4$	0.242	0.128	0.242	0.341	0.230
$lie5$	-0.186	-0.146	-0.212	-0.293	-0.316
$lie6$	-0.199	-0.246	-0.311	-0.358	-0.366

sample period: 1991Q3 to 2005Q3

4 Forecasting performance

Each leading index is included in the following simple forecasting model:

$$demp_t = \alpha + \beta(L)dlie_t + \gamma(L)demp_t + \epsilon_t \tag{3}$$

where $demp_t$ is the first difference in the natural logarithm of New Zealand employment at time t. Except for $lie1$ which is already in change form $dlie_t$ is the first difference in the natural logarithm of the leading index of employment ($dlie_t = lie1_t, dlie2_t, ...dlie6_t$). L is the lag operator ($L > 0$). α is a constant, $\beta(L)$ and $\gamma(L)$ are vector lag polynomials coefficients and ϵ_t is a white noise error. The short term employment forecasting model in equation (3) is estimated in EViews 5.1. All equations exhibit stable coefficients and the null hypothesis of normally distributed errors cannot be rejected. Table 3 shows the specification chosen for each model as well as an AR benchmark model. The table also gives the regression results. The regression R^2 is given below the model name (column (1)) and the parameter estimates with p-values in parenthesis are below the model specification (column (2)).

Table 3 shows that the model including $lie2$ explains the largest and that including $lie4$ explains the smallest share of quarterly variations in employment growth. $lie2$ is also the only index that has predictive power at three different lags, at lags 1, 3, and 4. $lie4$ has the lowest statistical significance. A zero coefficient of the index can only be rejected at a level of significance of 12 percent.

To assess the relative out-of-sample forecasting performance, the six models are compared to a benchmark AR(1,3) model. The forecasts here are generated using a fixed specification, rolling window, and time varying coefficients approach. With this technique, the six models are first estimated over the period 1991Q3 to 1998Q1. The estimated coefficients from each model are used to forecast employment growth one quarter ahead, *i.e.*, 1998Q2. The models are then rolled forward one quarter to the sample period 1990Q3 to 1998Q2 and re-estimated. Then a new set of one quarter ahead forecasts, *i.e.*, 1998Q3, are generated. This process is repeated until the last observation is reached. This leads to 30 one quarter ahead out-of-sample forecasts. $lie5$ and $lie6$ are reconstructed every quarter with a rolling window where the first window is 1990Q1 to 1998Q1 and 1990Q3 to 1998Q1. In the strict sense of the word the forecasts are not out-of-sample because the dataset includes historically revised series and only $lie5$ and $lie6$ are re-estimated every quarter.

The specification of the six models remains unchanged over the forecasting exercise. This may be undesirable in light of changes in the significance of the composite

Table 3: Forecasting models specification

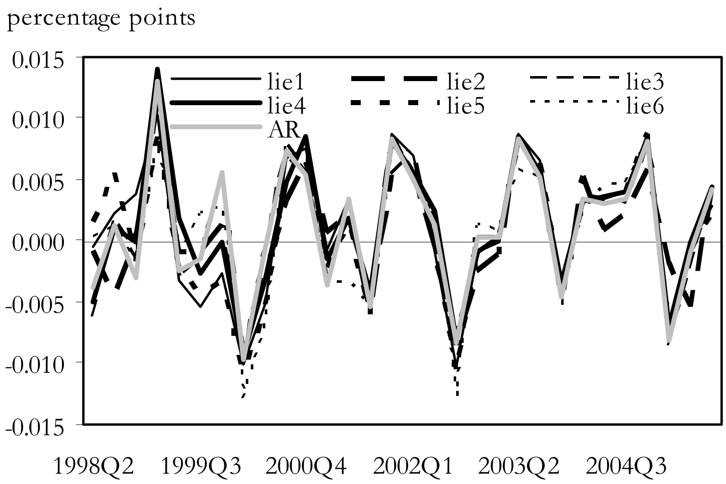
Model (R^2) (1)		Specification (2)		
$lie1:$	$demp_t$	$= \alpha^1 +$	$\beta^1_1 lie1_{t-1} + \gamma^1_3 demp_{t-3}$	$+\epsilon^1_t$
(0.264)		-0.028 (0.43)	0.000 (0.08); 0.420 (0.00)	
$lie2:$	$demp_t$	$= \alpha^2 +$	$\beta^2_1 lie2_{t-1} + \beta^2_3 lie2_{t-3} + \beta^2_4 lie_{t-4} + \gamma^2_3 demp_{t-3}$	$+\epsilon^2_t$
(0.420)		0.003 (0.00)	0.104 (0.05); 0.187 (0.00); 0.103 (0.06); 0.203 (0.09)	
$lie3:$	$demp_t$	$= \alpha^3 +$	$\beta^3_3 lie3_{t-3} + \gamma^3_1 demp_{t-1} + \gamma^3_3 demp_{t-3}$	$+\epsilon^3_t$
(0.309)		0.003 (0.01)	0.099 (0.08); 0.217 (0.07); 0.313 (0.01)	
$lie4:$	$demp_t$	$= \alpha^4 +$	$\beta^4_3 lie4_{t-3} + \gamma^4_3 demp_{t-3}$	$+\epsilon^4_t$
(0.256)		0.004 (0.00)	0.173 (0.12); 0.378 (0.00)	
$lie5:$	$demp_t$	$= \alpha^5 +$	$\beta^5_1 lie5_{t-1} + \gamma^5_3 demp_{t-3}$	$+\epsilon^5_t$
(0.272)		0.003 (0.00)	-0.117 (0.06); 0.479 (0.00)	
$lie6:$	$demp_t$	$= \alpha^6 +$	$\beta^6_1 lie5_{t-1} + \beta^6_2 lie5_{t-2} + \beta^6_3 lie5_{t-3} + \gamma^6_3 demp_{t-3}$	$+\epsilon^6_t$
(0.313)		0.003 (0.00))	-2.91 (0.06); 4.20 (0.09); -2.933 (0.05); 0.419 (0.00)	
AR	$demp_t$	$= \alpha^a +$	$\gamma^a_1 demp_{t-1} + \gamma^a_3 demp_{t-3}$	
(0.269)		0.002 (0.1)	0.226 (0.07); 0.377 (0.00)	

sample period: 1991Q3 to 2005Q3

index coefficients. Stock & Watson (1999), for example, reassess the lag structure in every quarter. The procedure applied here is a rolling window which allows estimated coefficients to change over time thus limiting the impact of retaining a fixed specification.

Figure 2 shows the out-of-sample forecast errors of employment growth from the six indexes and the AR benchmark. All six models exhibit similar errors with the smallest errors generated by the *lie*2 model.

Figure 2: Forecast errors



sample period: 1998Q2 to 2005Q3

To compare the forecasting performance of the six models more rigorously, Table 4 gives the mean absolute forecast error (MAE) and the root mean squared forecast error (RMSE) for each model relative to the AR benchmark. A value smaller than 1 indicates lower forecast errors compared to the benchmark while values greater than 1 indicate larger forecast errors compared to the benchmark.

Table 4 shows that the models including *lie*2 and *lie*4 generate lower and all other

Table 4: Forecast error statistics

	$lie1$	$lie2$	$lie3$	$lie4$	$lie5$	$lie6$
(1)	(2)	(3)	(4)	(5)	(6)	(7)
MAE	1.042	0.859	1.039	0.947	1.015	1.004
RMSE	1.034	0.886	1.035	0.742	1.011	1.024

sample period: 1998Q2 to 2005Q3

models generate larger forecast errors than the benchmark. The model including $lie1$ exhibits the largest forecast errors. Further, $lie5$ and $lie6$ generate similar forecast errors.

Table 5: Confusion matrix

Model	$\{up, up\}$	$\{down, down\}$	$\{up, down\}$	$\{down, up\}$
(1)	(2)	(3)	(4)	(5)
$lie1$	25	2	1	2
$lie2$	24	2	2	2
$lie3$	24	1	2	3
$lie4$	24	1	2	3
$lie5$	25	2	1	2
$lie6$	24	2	2	2
$AR(1,3)$	24	1	2	3

sample period: 1998Q2 to 2005Q3

Finally, Table 5 reports the confusion matrices for the seven models. The confusion matrix records the number of times a model correctly predicts the sign of next period's employment growth over the testing sample. Column (2) records the number of times a model correctly predicted an increase in employment growth, while column (3) reports how often the model correctly forecasted a fall. For example, the *lie*1 model correctly predicted twenty-five rises and two declines in employment. Columns (4) and (5) report the number of times a model missed the direction of employment changes. Column (4) (Column (5)) records the number of actual moves in employment growth that were up (down) while the predicted changes were down (up). Adding all values in the confusion matrix gives the overall number of out-of-sample forecasts. A word of caution is warranted when analyzing the confusion matrix. This test is somewhat simplistic as it does not distinguish whether a forecast was only slightly wrong or completely missed.

Table 5 suggests that all six models perform well in forecasting the sign of next period's employment growth. The models including *lie*1 and *lie*5 give the lowest number of false directions at 3 each, followed by *lie*2 and *lie*6 with 4 false directions each. At 5 the benchmark AR, the *lie*3 and *lie*4 have the largest number of false directions.

The figures and tables suggest three main results. First, the composite index of leading indicators constructed using the traditional NBER method (*lie*2) is the most useful tool for forecasting quarterly employment growth in New Zealand. The index provides early signals of cyclical changes and has the highest overall correlation with employment growth. Including the index in a short term forecasting model explains a larger share of quarterly variations in employment growth than a model including any of the other five indexes and generates lower forecasting errors. Second, utilizing larger datasets does not seem to add to the forecasting performance of an index. $lie5$ which includes the entire dataset of 95 variables generates larger errors than $lie2$. Moreover, $lie6$ which includes all 95 variables plus each variable at lags 1 and 2 and $lie5$ generate similar errors. Third, using the concordance method to weigh series seems to add to forecasting performance. Exogenously determining the component weights as in $lie2$ leads to better results than implicitly determining the weights as in $lie3$. Further, $lie4$ whose component series are also weighted by the concordance with employment generates lower errors than the model including $lie3$.

5 Conclusion

This paper constructed six leading indexes of employment in New Zealand for the period 1990Q1 to 2005Q3. The purpose of the paper was twofold. The first aim was to build an effective forecasting tool for quarterly employment growth. The second aim was to assess the relative performance of indexes constructed using various methods available in the literature.

The paper has three main findings. First, a composite index of leading indicators using the traditional NBER method is the most useful tool in providing early warning signals of cyclical changes in employment and in forecasting one quarter ahead employment growth in New Zealand. Second, increasing the size of the dataset by utilizing all available data and by including lags of all variables does not appear to boost the usefulness of the index. Third, exogenously determining the weights of an index's component series can strengthen forecasting performance.

Two extensions of the paper are to expand the dataset and to introduce more dynamics other than including lags in the dataset in the latent factor model. The dataset could be extended by including series at monthly frequencies and shorter time series. Stock & Watson (2002*b*) develop an algorithm that can be used for principal component analysis on such irregular datasets. Another extension of the paper is to follow Forni, Hallin, Lippi & Reichlin (2000) and introduce dynamics into the kalman filter leading indexes. Both extensions are subject of future research.

References

Brisson, M., Campbell, B. & Galbraith, J. W. (2003), 'Forecasting some low-predictability time series using diffusion indices', *Journal of Forecasting* **22**(6/7), 515–531.

Claus, E. (2001), 'Constructing NEO: A near-term employment outlook', *Department of Finance Working Paper* **2001-07**.

Claus, E. (2006), Two leading indexes of New Zealand employment, *Report prepared for the New Zealand Department of Labour.*

Claus, E. & Claus, I. (2002), 'How many jobs? A leading indicator model of New Zealand employment', *New Zealand Treasury Working Paper* **02/13**.

Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2000), 'The generalized dynamic factor model: Identification and estimation', *The Review of Economics and Statistics* **82**(4), 540–554.

Harding, D. & Pagan, A. R. (2002), 'Dissecting the cycle: A methodological investigation', *Journal of Monetary Economics* **49**(2), 364–381.

Holmes, R. A. & Shamsuddin, A. F. M. (1993), 'Evaluation of alternative leading indicators of British Columbia employment', *International Journal of Forecasting* **9**(1), 77–83.

Kennedy, J. E. (1994), 'The information in diffusion indexes for forecasting related economic aggregates', *Economics Letters* **44**(1-2), 113–117.

Kitchen, J. & Monaco, R. (2003), 'Real-time forecasting in practice', *Business Economics* **38**(4), 10–19.

Moore, G. H. (1983), 'Using a leading employment index to forecast employment in 1983', *Monthly Labor Review* **106**(5), 30–32.

Stock, J. H. & Watson, M. W. (1989), New indexes of coincident and leading economic indicators, *in* 'NBER Macroeconmics Annual', MIT Press, Cambridge, Massachusetts, pp. 351–394.

Stock, J. H. & Watson, M. W. (1999), 'Forecasting inflation', *Journal of Monetary Economics* **44**(2), 293–335.

Stock, J. H. & Watson, M. W. (2002*a*), 'Forecasting using principle components from a large number of predictors', *Journal of the American Statistical Association* **97**(460), 1167–1179.

Stock, J. H. & Watson, M. W. (2002*b*), 'Macroeconomic forecasting using diffusion indexes', *Journal of Business and Economic Statistics* **20**(2), 147–162.

The Conference Board (2006), How to compute diffusion indexes, *http://www.conference-board.org.*

Notes

[1]See http://www.stats.govt.nz/default.htm.

[2]Exceptions are, for example, Moore (1983), Holmes & Shamsuddin (1993), Kennedy (1994), Claus (2001), and Claus & Claus (2002).

[3]The Conference Board index has a threshold of 0.05 percent for assigning increases and decreases in the component series; see The Conference Board (2006).

[4]Holmes & Shamsuddin (1993) present empirically determined weights; the first is using the R^2 from regressing component series on cyclical variations in employment and the second is principal component analysis. Similarly, Claus (2001) uses correlation coefficients and Kitchen & Monaco (2003) the regression R^2s between employment and the component series. Stock & Watson (1999) employ three different procedures. In the first and second procedure the weight are chosen so that the composite index has the same sample mean or sample median and the third procedure also employs regresssion R^2s. A problem with using either regression R^2s or correlation coefficients is the loss in degrees of freedom. This means that too many relationships are estimated with a limited dataset leading to potentially questionable results.

[5]The choice of one to four quarters is somewhat arbitrary. The maximum lag of four quarters was chosen as leads of more than one year are generally difficult to establish.

[6]For a more detailed description see Claus (2006).

[7]See Table 1 pp. 362–363 in Stock & Watson (1989) for a list of the variables.

A Data and data sources

The 95 variables that we considered can be divided into seven broad categories.

(i) Labour market indicators

hours

unemployment rate

participation rate

ANZ job vacancy rate

ANZ job vacancy rate, spliced – three cities (Auckland, Wellington, Canterbury)

Quarterly Survey of business opinion (QSBO)

QSBO: find skilled labour

QSBO: find unskilled labour

QSBO: next three months, number of employees

QSBO: limiting factor – labour

QSBO: past three months – number of employees

QSBO: past three months – overtime worked

QSBO: next three months – overtime worked

(ii) Domestic activity indicators

company tax

permanent and long term migration – net actuals

external migration – arrivals, total

long term migration – arrivals

new dwelling consents

total dwellings

REINZ number of dwelling sales

food price index

livestock slaughter

total electricity generation – sales to customers

new vehicles car registration

Southern oscillation index

(iii) Trade indicators

merchandise imports

merchandise exports

arrivals, overseas visitors

West Texas intermediate oil price

(iv) Foreign activity indicators and commodity prices

US S&P500 equity price index

Australian equity price index – all ords

MSCI equity price index – Australia

MSCI equity price index – world

MSCI equity price index – United States

Top 5 trading partners real growth

ANZ commodity price index

CRB commodity price index

Economist commodity price index – total

Economist commodity price index – industrials

Economist commodity price index – non food agriculture

Economist commodity price index – metals

Economist commodity price index – food

Goldman Sachs commodity price index – total

Goldman Sachs commodity price index – agriculture

Goldman Sachs commodity price index – livestock

Dubai oil

Brent oil price

(v) Consumer and business confidence indicators

Quarterly Survey of business opinion (QSBO)

QSBO: general business situation

QSBO: limiting factor – capital

QSBO: limiting factor – other

QSBO: new investment – buildings

QSBO: new investment – plant and machinery

QSBO: past three months – average costs

QSBO: past three months – average selling price

QSBO: past three months – profitability

QSBO: next three months – average costs

QSBO: next three months – profitability

CAPU: next three months – profitability

(vi) Financial variables

1 year government bond yield

2 year government bond yield

5 year government bond yield

10 year government bond yield

30 day bank bill yield

60 day bank bill yield

90 day bank bill yield

10 year government bond yield – 5 year government bond yield

10 year government bond yield – 2 year government bond yield

10 year government bond yield – 1 year government bond yield

10 year government bond yield – 90 day bank bill yield

10 year government bond yield – 60 day bank bill yield

10 year government bond yield – 30 day bank bill yield

5 year government bond yield – 2 year government bond yield

5 year government bond yield – 1 year government bond yield

5 year government bond yield – 90 day bank bill yield

5 year government bond yield – 60 day bank bill yield

5 year government bond yield – 30 day bank bill yield

call rate

(vii) Monetary variables

NZD/AUD exchange rate – average 11am

NZD/USD exchange rate – average 11am

trade weighted index (TWI)

trade weighted index (TWI) / consumer price index (CPI)

total billings on New Zealand credit cards / CPI total advances on credit cards outstanding / CPI

notes and coins held by the public / CPI

M1

M2

M3

M1 / CPI

M2 / CPI

M3 / CPI

private sector credit

resident private sector credit

domestic credit

resident domestic credit

New Zealand stock exchange index, top 10 companies