QSBO as a forecasting tool

Paper for the 2011 NZAE conference

June 2011

About NZIER

NZIER is a specialist consulting firm that uses applied economic research and analysis to provide a wide range of strategic advice to clients in the public and private sectors, throughout New Zealand and Australia, and further afield.

NZIER is also known for its long-established Quarterly Survey of Business Opinion and Quarterly Predictions.

Our aim is to be the premier centre of applied economic research in New Zealand. We pride ourselves on our reputation for independence and delivering quality analysis in the right form, and at the right time, for our clients. We ensure quality through teamwork on individual projects, critical review at internal seminars, and by peer review at various stages through a project by a senior staff member otherwise not involved in the project.

NZIER was established in 1958.

Authorship

Prepared by:	Peter O'Connor and Dr Qing Yang
Quality approved by:	
Date:	17/06/2011 12:27 p.m.
Version:	1.0
Acknowledgements:	

8 Halswell St, Thorndon
P O Box 3479, Wellington
Tel: +64 4 472 1880
Fax: +64 4 472 1211
econ@nzier.org.nz
www.nzier.org.nz

Executive Summary

NZIER's *Quarterly Survey of Business Opinion* (QSBO) has provided a wealth of information over the previous 50 years as an indicator of economic activity. While its predictive capabilities are well known, the QSBO has largely been used to only forecast the next quarter's results. This paper investigates using the QSBO to forecast GDP and inflation over the following year. NZIER have developed forecasts using VAR with highly disaggregated QSBO data.

QSBO as a forecasting tool

Domestic trading activity is the best measure of GDP activity in the economy, while average costs provide a measure for inflation. These include breaking down series by sector, firm size and region that have been used in this analysis.

Methodology

Forecasts have been estimated using VAR. Each model has a similar structure at a different lag of GDP and inflation. Due to the numerous data series in the QSBO, these have been reduced into principal components. This is particularly useful as series are highly correlated.

Results

The QSBO data provides a robust predictor of GDP and inflation, particularly for the following two quarters. Developing the model revealed several findings on the drivers and best predictors of GDP and inflation from the QSBO series. A summary of the findings are:

- Labour market indicators, average costs, and selling prices add the most predictive value of the general economic indicators.
- Services sector indicators are strong predictors of GDP growth. This is unsurprising because the services sector represents a large portion of the New Zealand economy.

Contents

Appendices

Figures

Tables

1. Introduction

NZIER's *Quarterly Survey of Business Opinion* (QSBO) has provided a wealth of information over the previous 50 years as an indicator of economic activity. While its predictive capabilities are well known, the QSBO has largely been used to only forecast the next quarter's results. This paper investigates using the QSBO to forecast GDP and inflation over the following year. NZIER have developed forecasts using VAR with highly disaggregated QSBO data.

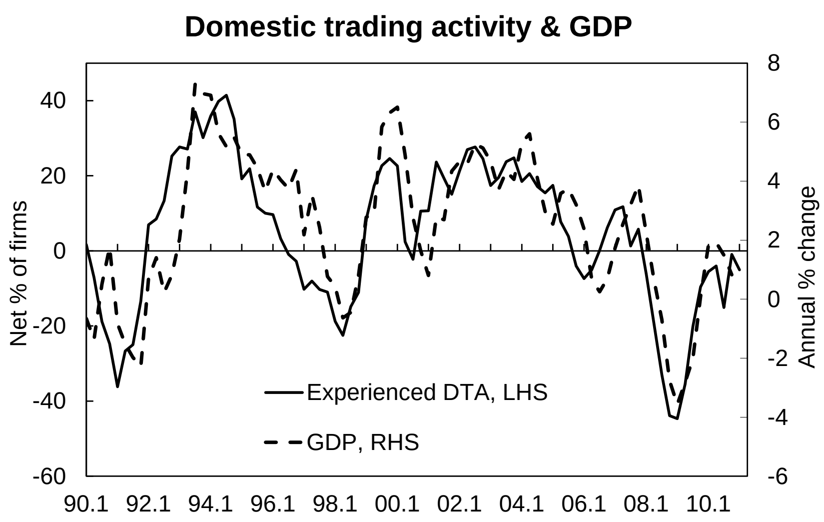
2. QSBO as a predictive tool

Every quarter, thousands of chief executives are surveyed as to the performance of their firm compared to the previous quarter. The survey results produce a vast array of data on business opinion and the aggregated results are released to members of NZIER. This includes the net percentages of the responses sent in from survey participants. Domestic trading activity provides the best measure of GDP, while average costs provide a measure for inflation.

However, this vast array of data has been underutilised in the past. Relationships have been established between QSBO data and data released from Statistics New Zealand. The relationships have been quite robust over time (Figure 1), but can be improved upon. There are two limitations of these basic relationships:

- a single series from the QSBO is unable to fully reflect the quarterly variation
- the relationships can only be used to forecast for the current and expected quarters.

Figure 1 Domestic trading activity & GDP



Source: Statistics NZ, NZIER

This analysis tests the forecasting potential of the QSBO of up to a year ahead of statistical releases. To do this, disaggregated data has been 'built up' using numerous series to provide a robust forecasting model. Data is disaggregated by sector, firm size and region.

2.1 What makes the QSBO so good?

Long-term data series are not common in New Zealand. Those that do span a long time often consist of break points where calculation of the series has changed. Whereas the QSBO provides a robust 50 year history; its greatest strength. The questions have run relatively unaltered over the 50 years, with a few questions added and the responses adjusted for questions identifying characteristics of the firm.

With such a long series of data, models generated using QSBO will be founded on drivers of GDP over a long time horizon rather than the best fit of a short time series. Short term forecasts are already used, but as a predictive tool, we have the data to forecast over a whole year.

Some limitations exist to the use of long term time series. There have been structural changes to inflation over the past 50 years. The last major change was in the late 1980s when inflation targeting was introduced. Since 1991, inflation rates have been low and stable, but are not comparable to inflation rates of the 1970s. This means a model cannot easily be fitted over the 50 year history.

2.2 Data within the QSBO

The QSBO contains numerous questions that can be used for forecasting. These series can be highly correlated, and only the key drivers[1] of inflation and GDP have been included in the analysis. These include:

- labour market conditions: primarily employment, overtime worked, and labour turnover
- general economic conditions: investment intentions, expected overtime worked and profitability
- average cost and selling prices.

These variables are broken down by sector for their influence on GDP and inflation. Within each sector there are two variables, **experienced** changes and **expected** changes. The three sectors are manufacturers and builders, merchants and services.

3. Methodology

Forecasts have been estimated using VAR. Each model has a similar structure using different time lag for GDP and inflation. All the models were specified as:

[1] Determined by their predictive capability.

$$y_{t+h}^{h} = \mu + \alpha(L)y_t + \beta(L)Z_t + \varepsilon_{t+h}^{h}$$

where h is the forecasting horizon, y_{t+h}^{h} is the projection of GDP/inflation h quarters ahead, $\alpha(L)$ and $\beta(L)$ are lag polynomials, Z_t is a vector of principal components from the QSBO, and μ is a constant.

This approach is not new to forecasting GDP and inflation. Stock and Watson (1999) used this model to forecast inflation in the US, while Marcellino, Stock, and Watson (2003) used this model to forecast GDP growth and inflation in a Euro wide economy.

3.1 Suitability of the model

In order to implement the VAR model, we have to decide the order of integration of the two series: GDP growth and CPI index. In the case of GDP growth, both Dickey-Fuller test and Phillips-Perron test have rejected the existence of unit root, so that y_t denotes the growth rate of GDP. Inflation rates before 1990 are very volatile due to not having inflation targeting, where statistical tests[2] suggest that the inflation series had a structural change between 1986 and 1989. Because of this, we only fit the model for inflation from 1990. Statistical tests have also rejected the existence of a unit root[3] so that y_t denotes the inflation rate.

3.2 Principal Component Analysis (PCA)

Many of the QSBO indicators are highly correlated with each other. To overcome this phenomenon we use Principal Component Analysis (PCA). This reduces the number of indicators by combining the joint (or correlated) effects of these indicators. PCA is widely used in forecasting where there are many predictors (see Stock (2006)[4] for a review).

Before the data can be modelled, the variables have been constructed into nine principal components. VAR (Vector Autoregression)[5] can then be used to estimate the relationships between GDP growth, inflation rate and the QSBO's principal components.

Principal Component Analysis (PCA) is a multivariate analysis technique that was first introduced by Pearson in 1901 (Pearson 1901) and developed independently by Hotelling in 1933. PCA involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables

[2] Statistical tests suggested by Clemente, Montanes, and Reyes (1998) and Andrews and Zivot (1992). Testing for a unit root in variables with a double change in the mean. *Economics Letters* 59, 175-182.

[3] Matheson (2006) has also forecasted inflation rate as a series with an integration order of 0.

[4] Stock(2006), Forecasting with Many Predictors, in Handbook of Economic Forecasting.

[5] See Stock and Waston (2001), Vector Autogression, *Journal of Economic Perspectives*, Vol 15 (4), 101-115.

called principal components, which are linear combinations of the variables that explain the maximum amount of variance in the original variables.[6]

The first component accounts for most of the variance in the variables. Then the second component accounts for the largest share of the remaining variance, and so on.

We apply PCA to economy wide, manufacturing, merchants, and services sector indicators separately. The results of the PCA are in Appendix B.

There were three principal components among the economy wide indicators.

- **genpc1:** capturing labour market conditions
- **genpc2:** capturing general economic conditions
- **genpc3:** capturing average cost and selling prices.

There were two principal components for the sectorial breakdowns.

- **exper:** capturing experienced changes
- **expect:** capturing expected changes.

The relationship was not as clear for the services sector, but the same two principal components were retained. The services sector is full of highly variable firms, ranging from a large number of small firms, to some of the largest firms in the economy. The performance of firms varies between small and large firms. This can also relate to location where the larger firms tend to be located (or centralised) in Auckland.

4. Results

VAR analysis results are shown in the tables below. The QSBO data provides a robust predictor of GDP and inflation, particularly for the following two quarters. Figure 2shows how well the VAR model is able to predict GDP up to a year ahead of official statistics. Figure 3 shows the results for inflation.

Developing the model revealed several findings on the drivers and best predictors of GDP and inflation from the QSBO series. A summary of the findings are:

- labour market indicators, average costs, and selling prices add the most predictive value of the general economic indicators
- services sector indicators are strong predictors of GDP growth. This is unsurprising because the services sector represents a large portion of the New Zealand economy

[6] See the appendix for details of the mathematics of PCA.

- activity **expectations** of merchants provide a stronger predictor of GDP than **experienced** activity
- manufacturers and builders are the weakest sector in predicting GDP
- sectoral indicators are poor predictors of inflation.

Table 1 Regression results for GDP

	Current quarter		**Next quarter**		**Two quarters ahead**		**Three quarters ahead**	
	Coef.	P>z	Coef.	P>z	Coef.	P>z	Coef.	P>z
Lag1 GDP	0.616***	0	0.312***	0	0.180*	0.07	-0.405***	0
Lag2 GDP	-0.186**	0.04	-0.061	0.57	-0.517***	0	0.132	0.35
Lag3 GDP	0.015	0.87	-0.475***	0	0.12	0.31	0.077	0.59
Lag4 GDP	-0.182***	0.01	0.251***	0	0.131	0.14	0.054	0.61
genpc1	0.450*	0.07	0.611**	0.04	0.485	0.14	1.009***	0.01
genpc2	-0.395	0.3	-0.222	0.62	0.077	0.88	1.580***	0.01
genpc3	-0.401**	0.02	-0.439**	0.02	-0.719***	0	-1.063***	0
m&bexper	0.707***	0.01	0.574*	0.08	-0.031	0.93	-1.209***	0.01
m&bexpect	0.17	0.54	0.468	0.14	0.257	0.47	-0.709*	0.1
mercexper	0.056	0.81	0.32	0.23	0.431	0.15	0.628*	0.07
mercexpect	0.197	0.33	0.588***	0.01	0.842***	0	0.835***	0.01
servexper	0.560***	0.01	0.536**	0.03	0.843***	0	0.607*	0.06
servexpect	0.344**	0.03	0.337*	0.07	0.659***	0	0.374	0.14
constant	1.480***	0	1.917***	0	2.102***	0	2.210***	0
R^2	0.8301		0.7693		0.7207		0.6107	

Notes: (1) ***, **, * represent the 1%, 5% and 10% level of significance respectively.

Source: NZIER

Table 2 Regression results for Inflation

	Current quarter		Next quarter		Two quarters ahead		Three quarters ahead	
	Coef.	P>z	Coef.	P>z	Coef.	P>z	Coef.	P>z
Lag1 CPI	0.718***	0	0.514***	0	0.580***	0	0.05	0.74
Lag2 CPI	0.004	0.98	0.16	0.27	-0.274**	0.05	-0.081	0.7
Lag3 CPI	0.164	0.22	-0.335**	0.03			0.136	0.53
Lag4 CPI	-0.243***	0.01	0.173*	0.08			0.177	0.22
genpc1	-0.775***	0.01	-0.561*	0.1	-1.094***	0.01	-0.878*	0.07
genpc2	-0.462	0.14	-0.455	0.19	-0.999**	0.03	-0.885*	0.1
genpc3	1.195***	0	1.491***	0	1.942***	0	2.246***	0
m&bexper	0.093	0.65	0.534**	0.02	0.439	0.17	0.506	0.17
m&bexpect	-0.106	0.65	-0.022	0.93	0.091	0.79	0.207	0.6
mercexper	0.251	0.21	0.175	0.44	0.519*	0.09	0.537	0.11
mercexpect	0.017	0.91	-0.032	0.85	0.334	0.12	0.485**	0.04
servexper	0.486*	0.06	0.293	0.3	0.521	0.14	0.413	0.31
servexpect	-0.018	0.91	-0.301*	0.08	-0.014	0.95	-0.176	0.51
constant	2.053***	0	2.384***	0	3.416***	0	3.461***	0
R^2	0.853		0.8157		0.6744		0.6193	

Notes: (1) ***, **, * represent the 1%, 5% and 10% level of significance respectively.

Source: NZIER

Figure 2 Forecast and Real GDP

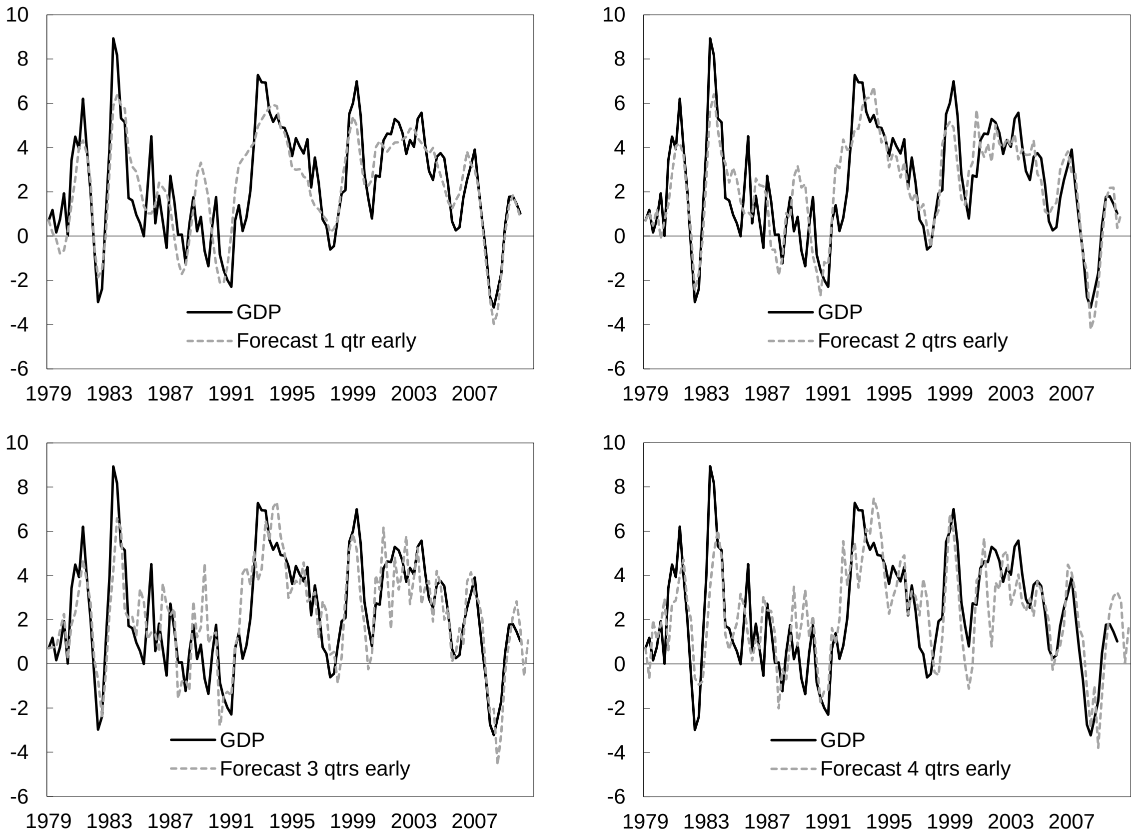
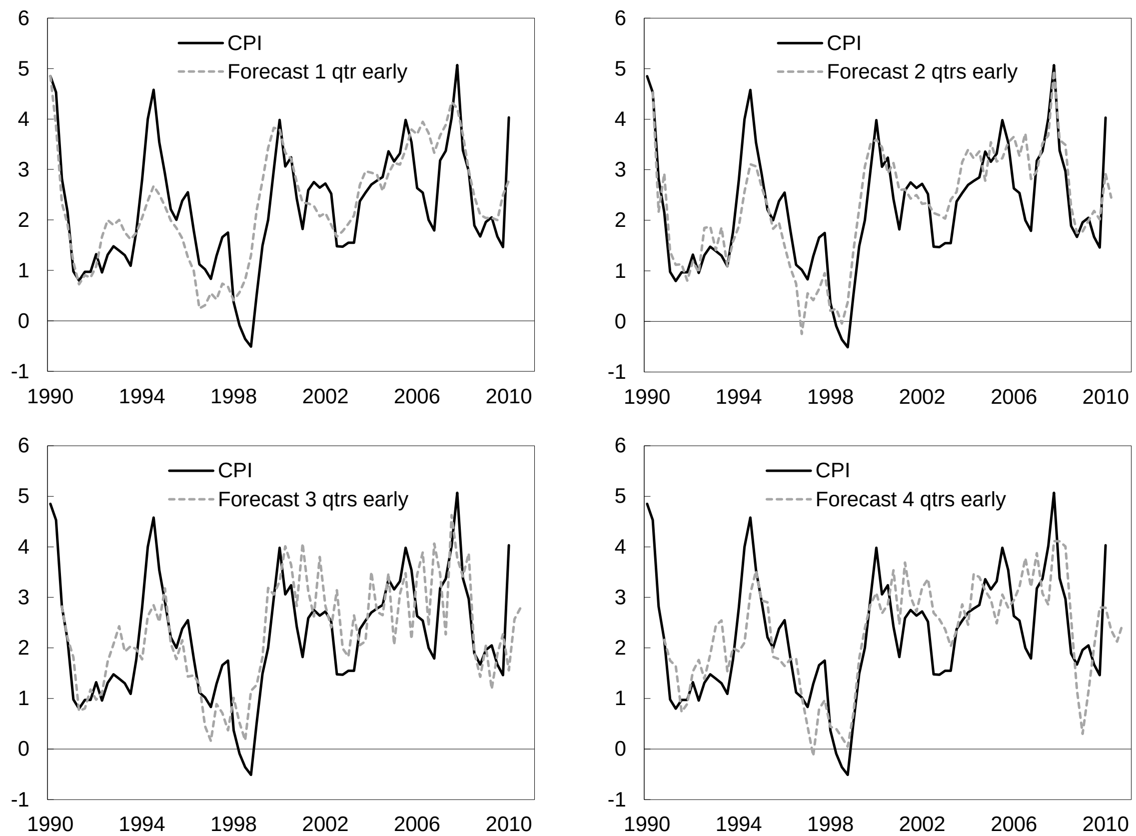


Figure 3 Forecast and Real Inflation

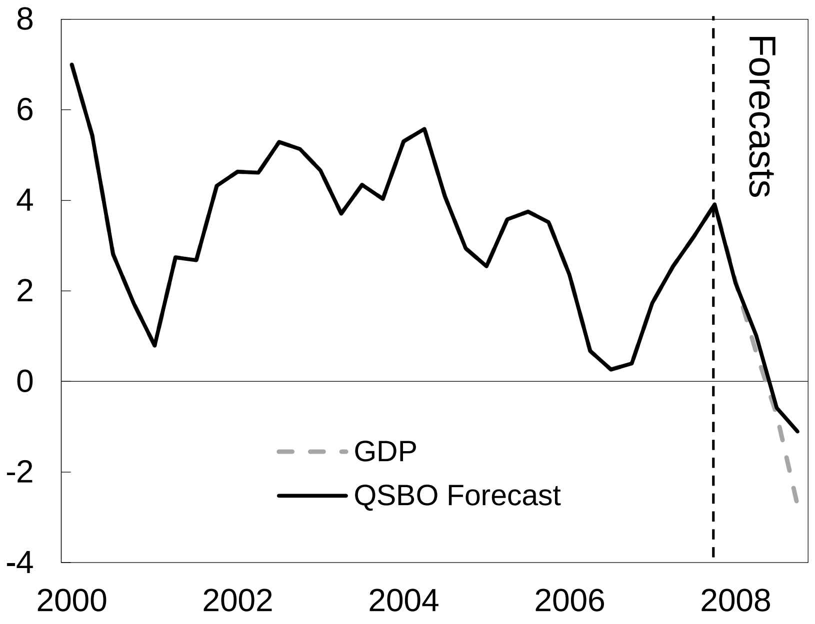


5. Testing historical forecasts

This section investigates how the QSBO would have fared when forecasting some of the more volatile periods of GDP. For example, the global financial crisis provides a unique situation of a time that was very difficult for forecasters to predict. The QSBO was able to provide a timely and accurate picture of GDP in 2008 after the April 2008 release.

Figure 4 depicts what the forecast model predicted for the following four quarters after the April 2008 QSBO release. As March 2008 GDP was not released at this time, the forecasts are for the 2008 calendar year. The forecasts for March, June and September were very accurate, but diverged for December 2008. Given the volatile conditions, this is not unexpected. What is astounding is that the QSBO provided a timely and accurate prediction of the extent of the GFC impact a year in advance.

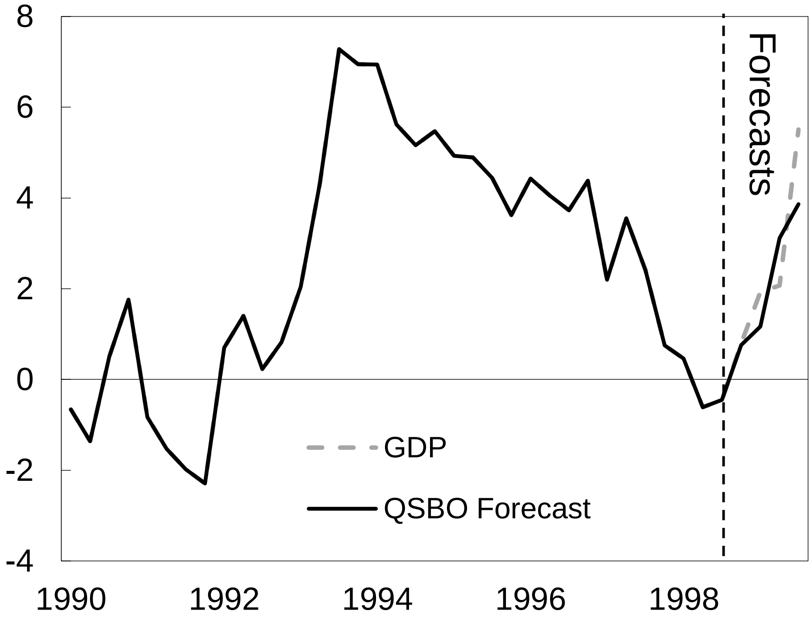
Figure 4 Actual versus forecast GDP 2008



Source: Statistics NZ, NZIER

Figure 5 depicts what the QSBO forecast model predicted for the following four quarters after the January 1998 release. This time, the forecasts are just after a mild recession and are predicting the strength of the recovery. The QSBO again gives a strong depiction of the path of GDP over the following year.

Figure 5 Actual versus forecast GDP 1998-1999



Source: Statistics NZ, NZIER

6. Conclusion

The QSBO provides a strong predictive tool. Forecasts have been created for GDP and inflation and are accurate up to a year ahead of statistical releases, with accuracy peaking in the short term. VAR combined with principal component analysis provides a strong modelling framework to utilise the QSBO's 50 year history. This yields robust forecasts that have shown to be accurate, even during volatile times. There are plenty more opportunities to utilise the QSBO's data that will be explored in the future.

7. References

Banerjee, A., Marcellino, M. & Masten, I., 2005. Leading Indicators for Euro-area Inflation and GDP Growth*. *Oxford Bulletin of Economics and Statistics*, 67(s1), 785-813.

Bloor, C. & Matheson, T., 2010. Real-time conditional forecasts with Bayesian VARs: An application to New Zealand. *The North American Journal of Economics and Finance*, In Press, Corrected Proof.

Buckle, R.A. et al., 2007. A structural VAR business cycle model for a volatile small open economy. *Economic Modelling*, 24(6), 990-1017.

Burns, A.F. & Mitchell, W.C., 1946. Measuring Business Cycles. Available at: http://econpapers.repec.org/bookchap/nbrnberbk/burn46-1.htm [Accessed February 17, 2011].

Chamberlin, G., 2007. Forcasting GDP using external data sources. *Economic & Labour Market Review*, 1_(8), 18-23.

Cheung, C. & Demers, F., 2007. *Evaluating Forecasts from Factor models for Canadian GDP Growth and Core Inflation*, Bank of Canada. Available at: http://ideas.repec.org/p/bca/bocawp/07-8.html [Accessed February 18, 2011].

Kabundi, A., 2004. Estimation of Economic Growth in France Using Business Survey Data. In *Estimation of Economic Growth in France Using Business Survey Data*. IMF.

Mariano, R.S. & Tse, Y., 2008. *Econometric forecasting and high-frequency data analysis*, World Scientific.

Narayan, P.K., Narayan, S. & Prasad, A., 2008. A structural VAR analysis of electricity consumption and real GDP: Evidence from the G7 countries. *Energy Policy*, 36(7), 2765-2769.

Pesaran, M.H., Schuermann, T. & Smith, L.V., Forecasting economic and financial variables with global VARs. *International Journal of Forecasting*, 25(4), 642-675.

Qin, D. et al., 2008. Automatic leading indicators versus macroeconometric structural models: A comparison of inflation and GDP growth forecasting. *International Journal of Forecasting*, 24(3), 399-413.

Schumacher, C., 2007. Forecasting German GDP using alternative Factor models based on large datasets. *Journal of Forecasting*, 26(4), 271-302.

Schumacher, C. & Dreger, C., 2002. Estimating Large-Scale Factor models for Economic Activity in Germany: Do They Outperform Simpler Models? *HWWA Discussion Paper no. 199*. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=335062 [Accessed February 2, 2011].

Sims, C.A., 1980. Macroeconomics and Reality. *Econometrica*, 48(1), 1-48.

Stock, J.H. & Watson, M.W., 2005. An empirical comparison of methods for forecasting using many predictors. *Manuscript, Princeton University*.

Stock, J.H. & Watson, M.W., 2002a. Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association*, 97(460), 1167-1179.

Stock, J.H. & Watson, M.W., 2002b. Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics*, 20(2), 147-162.

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

Stone, R., 1947. On the Interdependence of Blocks of Transactions. *Supplement to the Journal of the Royal Statistical Society*, 9(1), 1-45.

Tallman, E.W., 1999. Vector autoregressions: forecasting and reality. *Federal Reserve Bank of Atlanta*. Available at: http://www.frbatlanta.org/frbatlanta/filelegacydocs/robtallman.pdf.

Appendix A Literature review

A review of the literature is provided of the approaches used to forecast GDP and inflation in recent decades. Burns and Mitchell (1946) identified the regularity of business cycles co-movement with different economic variables. Since their finding a wealth of knowledge developed to forecast short and long run economic variables. Models have developed extensively during recent years, resulting in a considerable body of literature which includes both theoretical and empirical analysis.

VAR is a widely used tool in empirical macroeconomic forecasting since its introduction by Sims (1980). VAR's strengths are in finite samples with limited explanatory variables. The reliability of forecasting results depends on the choice of variables, and optimum performance is achieved with fewer than ten variables (Qin et al. 2008). Because of this we utilised PCA in our model to reduce the number of explanatory variables to nine.

The most recent extension of the VAR model is proposed by Bloor and Matheson (2010). They attempt to forecast New Zealand GDP and inflation via Bayesian Vector Autoregression (BVAR). Bayesian methods were used on a large panel to impose tighter priors with the number of variables included in the model. Bayesian shrinkage was also applied which gives more weight to larger principal components (Bloor & Matheson 2010). The data sets include New Zealand quarterly data[1] from March 1990 to September 2008. They produced three different results by using a small size (5 variables), a medium size (13 variables) and a large size (35 variables) of BVAR respectively. The results suggested that the large BVAR provided the most reliable forecast of macroeconomic activities.

A further New Zealand study used a 13 variable structured VAR model to investigate business cycles. The model was capable of explaining shocks to the business cycle in our small, volatile open economy (Buckle et al. 2007).

The introduction of factor analysis and principal component analysis are two well-known methods for summarising variation and covariation among large numbers of variables. The use of factor models is not restricted by the number of variables. Meanwhile, the use of the PCA is better with a small number of variables. The difference can be diminished with a larger number of variables, but the factor model is naturally more data-driven. The approach was originally proposed by Stock and Watson (1989).

Factor models and VAR has been extensively used to forecast economic variables at the macro level of many different countries (J. H Stock & M. W Watson 2005; James H Stock & Mark W Watson 2002a; James H Stock & Mark W Watson 2002b; James H. Stock & Mark W. Watson 1999). The rationale behind the model is that the co-

[1] Quarterly data includes business and consumer confidence, housing and labour market indicators, consumption, investment, production, financial markets, and the world economy.

movement of the variables have a common factor which can be captured by one single latent variable.

There are three assumptions that need to be satisfied which are: a priori distinction between coincident and leading variables; no correlation between common and idiosyncratic components at all leads and lags; and no mutual correlation in between idiosyncratic components (Schumacher & Dreger 2002). One drawback of the factor model is that setting up a large dataset to run an effective model is resource intensive (Cheung & Demers 2007).

Another recent study did a comparison between a static factor model, a dynamic principal component model, and a subspace factor model to forecast German GDP (Schumacher 2007). The study concluded that the performance of forecasting relies heavily on the choice of appropriate information criteria for the auxiliary parameters of the model. The generalized dynamic factor model is based on PCA (Cheung & Demers 2007).

PCA was introduced into econometric modelling by Stone (1947). Initially, it was frequently used in psychology research to identify latent factors. This technique is famous for reducing a set of large data collections into a more manageable form, especially for dealing with problems of multi-collinearity and shortage of degrees of freedom (Mariano & Tse 2008). A major difference between a factor model and PCA is that the latter does not require an a priori distinction between coincident and leading variables. Additionally, it also allows as many principal components as they are indicators (Kabundi 2004).

Another forecasting issue to confront is the choice of data. Many studies use both business and consumer survey data or monetary and fiscal data (Chamberlin 2007). The advantage of using survey data is that they are released in advance of the official statistics. To have the best understanding of the economy, there are a large number of available indicators that can be drawn from business surveys and financial markets (Chamberlin 2007).

Appendix B Principal component analysis

B.1 Mathematics of Principal Component Analysis

Suppose there are M potential quarantine pests, each with N risk factors. The pests risk attributes can be organised as follow:

$$X = \begin{pmatrix} X_1 \\ X_2 \\ . \\ . \\ . \\ X_N \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1M} \\ x_{21} & x_{22} & \cdots & x_{2M} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ x_{N1} & x_{N2} & \cdots & x_{NM} \end{pmatrix}$$

where x_{ij} denotes the risk scores given by the risk assessors for the pest j in regard to the *ith* risk factors. PCA seeks to find a set of new variables $Y = (Y_1 \quad Y_2 \quad \ldots \quad Y_P)^T$, which are linear combinations of X as follows:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_P \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1N} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2N} \\ . & & & \\ . & & & \\ . & & & \\ \alpha_{P1} & \alpha_{P2} & \cdots & \alpha_{PN} \end{pmatrix} \times \begin{pmatrix} X_1 \\ X_2 \\ . \\ . \\ . \\ X_N \end{pmatrix}$$

where α_{ij} is the weight value that reflects the contribution of X_j to Y_i, the *ith* principal component, satisfying:

$$\sum_{i=1}^{N} \alpha_{ij}\alpha_{jk} = 0 \qquad j \neq k$$

$$\sum_{i=1}^{N} \alpha_{ij}\alpha_{jk} = 1 \qquad j = k$$

α_{ij} is also termed factor I's.

As principal components are linear combinations of variables, the coefficient of each variable is the scoring coefficient on the principal components.

The new variables $Y = (Y_1 \quad Y_2 \quad \ldots \quad Y_P)^T$ [2] are themselves uncorrelated, but retain maximally the variance of observations. This linear combination can be found by solving the following eigensystem subject to the above constraints:

[2] T: transpose

$$(C-\lambda I)A=0$$

where λ are the eigenvalues and $\lambda_1 > \lambda_2 > ... > \lambda_p > 0$, C is the covariance matrix, and A are the eigenvectors.

The ratio of variance explained by the first q principal components can be expressed as:

$$R_q^2 = \frac{\lambda_1 + \lambda_2 + ... + \lambda_q}{\lambda_1 + \lambda_2 + ... + \lambda_p} \qquad q \le p$$

The first principal component accounts for the largest share of the total variance, the second principal component accounts for the largest share of the remaining variance, and so on.

B.2 How many principal components (PCs) to retain

The use of more components increases the model's explanatory power, but does not achieve model simplification. In contrast, using or choosing fewer components results in reduced explanatory power for the model. In deciding how many factors to retain and extract, there are generally three tests. The first test is the scree test. The scree test is a graphic method for determining the number of factors. The eigenvalues are plotted in the sequence of the principal factors. The number of factors is chosen where the plot levels off to a linear decreasing pattern. The second test is proposed by Everitt and Dunn (1992) suggesting to discard all components accounting for less than $(70/n)\%$ of the overall variance, where n is the number of PCs. The third test is proposed by Hotelling (1933) suggesting to keep the first few PCs that explains more than 85% of the total variance.

B.3 Factor loadings

Factor loading is a term used to refer to factor pattern coefficients or structure coefficient, which multiply with PCs to produce measured variables. Furthermore, it represents the correlations between the original variables and the new principal components. Mathematically, it can be shown as:

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_N \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1N} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2N} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \alpha_{P1} & \alpha_{P2} & \dots & \alpha_{PN} \end{pmatrix}^{-1} \times \begin{pmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_P \end{pmatrix} = \begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1p} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2p} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \beta_{N1} & \beta_{N2} & \dots & \beta_{NP} \end{bmatrix} \times \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_p \end{bmatrix}$$

where β_{ij} is the factor loading or pattern coefficient for factor X_i on principal component factor Y_j.

The new axes, or dimensions, are uncorrelated with each other, and are selected according to the amount of the total variance that they describe. Normally this results in there being a few large axes accounting for most of the total variance, and a large number of small axes accounting for very small amounts of the total variance. These small axes are normally discounted from further consideration, so that the data set having P correlated variables has been transformed to a data set having N uncorrelated axes, or principal components, where N is usually less than P.

The fact that the N axes are uncorrelated is often a very useful property if further analysis is planned. Much attention focuses on the relationship of the principal components to the original variables. For example, which of the original axes contributed the largest variance to each of the principal components.

B.4 Principal components

Principal Component Analysis for Economy Wide Indicators

Variable	Factor1	Factor2	Factor3
Genbus	-0.0963	0.8804	-0.3694
Labskill	-0.9407	-0.1869	-0.1371
Labunskill	-0.967	-0.1068	0.1325
Investbuild	0.6911	0.6553	0.0525
Investplant	0.5517	0.7949	-0.046
Nosexper	0.8506	0.3855	0.1752
Nosexpect	0.7118	0.5917	0.1529
Overtimeexper	0.8098	0.5074	0.028
overtimeexpect	0.5868	0.7501	-0.0072
Ltexper	0.9564	0.1051	-0.1611
Ltexpect	0.8197	0.3587	-0.2859
Avecostexper	-0.1074	-0.2297	0.9558
Avecostexpect	-0.0778	-0.1561	0.9711
Avepriceexper	0.0428	-0.0631	0.983
avepriceexpect	0.054	0.068	0.9762
Profitexper	0.5732	0.7212	-0.0374
Profitexpect	0.3655	0.8253	-0.315

PCA for Manufacturing Sector Indicators

Variable	Factor1	Factor2
Mandbdta	0.9067	0.4092
Mandbdtae	0.4245	0.8974
Mandbsdta	0.9041	0.2434
Mandbsdtae	0.4925	0.7761
Mandbmdta	0.8282	0.3068
Mandbmdtae	0.3089	0.7698
Mandbldta	0.8182	0.4853
Mandbldtae	0.3561	0.8879
Mandbadta	0.8998	0.311
Mandbadtae	0.4388	0.8148
Mandbwdta	0.7511	0.4077
Mandbwdtae	0.1921	0.7584
Mandbcdta	0.748	0.3441
Mandbcdtae	0.3948	0.701
Mandbrdta	0.7515	0.4225
Mandbrdtae	0.3285	0.8098

PCA for Merchants Sector Indicators

Variable	Factor1	Factor2
Merchantsdta	0.9213	0.3608
Merchantsdtae	0.3816	0.9158
Merchantssdtaa	0.8716	0.304
Merchantssdtae	0.3626	0.8498
Merchantsmdtaa	0.8173	0.2597
Merchantsmdtae	0.3992	0.7043
Merchantsldtaa	0.8535	0.3839
Merchantsldtae	0.3511	0.8732
Merchantsadtaa	0.8561	0.349
Merchantsadtae	0.4655	0.7629
Merchantswdtaa	0.7386	0.3612
Merchantswdtae	0.0239	0.8623
Merchantscdtaa	0.7635	0.3191
Merchantscdtae	0.4315	0.6867
Merchantsrdtaa	0.8611	0.1911
Merchantsrdtae	0.5003	0.6502

PCA Analysis for Services Sector Indicators

Variable	Factor1	Factor2
Servicesdta	0.7726	0.5749
Servicesdtae	0.9052	0.3405
Servicessdta	0.7362	0.5071
Servicessdtae	0.8337	0.1736
Servicesmdta	0.5662	0.66
Servicesmdtae	0.5439	0.5415
Servicesldta	0.7377	0.5262
Servicesldtae	0.8232	0.3562
Servicesadta	0.8144	0.2927
Servicesadtae	0.8782	0.0863
Serviceswdta	0.6696	0.4597
Serviceswdtae	0.7951	0.1978
Servicescdta	0.1167	0.8584
Servicescdtae	0.2565	0.7132
Servicesrdta	0.6512	0.5112
Servicesrdtae	0.686	0.2965