Reserve Bank of New Zealand
Analytical Notes

Using Job Transitions Data As A Labour Market Indicator

AN2020/02

Christopher Ball, Adam Richardson and Thomas van Florenstein Mulder

January 2020

Reserve Bank of New Zealand Analytical Note Series
ISSN 2230-5505

Reserve Bank of New Zealand
PO Box 2498
Wellington
NEW ZEALAND

www.rbnz.govt.nz

RESERVE BANK
OF NEW ZEALAND
TE PŪTEA MATUA

1. Introduction

The Reserve Bank of New Zealand (RBNZ) is now mandated to support the maximum sustainable level of employment. This new objective has lead us to look for new sources of data that allow us to better understand and predict labour market outcomes.

This paper explores the usefulness of a large administrative data set collected by the Inland Revenue Department (IRD). This data records changes in tax codes for all tax payers in New Zealand over time. We can therefore track each taxpayer's movements across different employment states, e.g. a movement from full time employment to studying. This data set covers all registered tax payers in New Zealand and is therefore a wide covering data set. This IRD data is currently lagged. Therefore, this paper aims to assess the benefit of the dataset if it was made available in a timelier manner. One way to do this is to test the usefulness of the dataset in nowcasting the unemployment rate.

Richardson et al. (2018) shows that machine learning models can lend to nowcast accuracy when combined with large datasets, therefore we use their methodology in conjunction with the job-to-job transitions data to assess the usefulness of this data set in predicting labour market outcomes.

The results indicate that combining the administrative dataset with the machine learning models allows us to nowcast unemployment more accurately than conventional benchmarks. This suggests that timelier job-to-job transitions data could help improve our assessment of the labour market in New Zealand.

The paper is outlined as follows. Section 2 presents the data and the empirical methodology followed in this exercise. Section 3 presents the results and section 4 concludes.

2. What is the job transitions data?

This analysis uses the anonymised linked administrative data available in the Integrated Data Infrastructure (IDI).[1] The administrative tax data from the IRD is part of the IDI. These tax data

---

[1]The results in this report are not official statistics. They have been created for research purposes from the Integrated Data Infrastructure (IDI) managed by Statistics New Zealand. The opinions, findings, recommendations and conclusions expressed in this report are those of the authors, not Statistics NZ or the Reserve Bank of New Zealand. Access to the anonymised data used in this study was provided by Statistics NZ in accordance with security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business or organisation and the results in this report have been anonymised to protect these groups from identification. Careful consideration has been given to the privacy, security and confidentiality issues associated with using administrative and survey data in the IDI. Further detail can be found in the privacy impact assessment for the IDI available from www.stats.govt.nz. The results are

contain one record per employee, employer, and quarter from April 1999 for almost every NZ tax payer.[2] We can categorise each record as income obtained from one of the following sources: pension, ACC, paid parental leave, student allowance, benefit, withholding payment or wage/salary. Some individuals appear multiple times per quarter and therefore we prioritise their income sources in the same order as presented above. Therefore if a person is recorded twice in the quarter, firstly for an income from a wage/salary and then secondly to an income from a pension, they will only be recorded once as having an income from a pension. The ordering loosely reflects increasing attachment to the labour market.

We then use these categories to construct the transitions data. We can track whether an individual changes income category from quarter to quarter e.g. from benefit in the first quarter to wage/salary in the second quarter, or whether they stay in the same category e.g. benefit in the first quarter to benefit in the second quarter. If they are recorded as having been in the wage/salary category in one quarter and they are still in the wage/salary category the following quarter we can identify whether their employer is the same as the previous period and separate them. There is a total of roughly 65 possible transitions between states. This gives us a time series that has a count for each of the 65 transition states per quarter.

We transform the data from gross flows to proportion of transitions and we remove any transition states that have no data. This leaves us with around 50 transition states, these will be our explanatory variables used in nowcasting the unemployment rate. This data set covers the 1999Q3 to 2018Q1 period.

This data is currently only available with a substantial lag of six months.[3] This exercise can be considered as an assessment of the quasi real-time nowcast performance of the employment transitions data if it were made available in a timelier manner.

3. Nowcast Evaluation Exercise

---

based in part on tax data supplied by Inland Revenue to Statistics NZ under the Tax Administration Act 1994. This tax data must be used only for statistical purposes, and no individual information may be published or disclosed in any other form, or provided to Inland Revenue for administrative or regulatory purposes. Any person who has had access to the unit-record data has certified that they have seen, read, and have understood section 81 of the Tax Administration Act 1994, which relates to secrecy. Any discussion of data limitations or weaknesses is in the context of using the IDI for statistical purposes, and is not related to the data's ability to support Inland Revenue's core operational requirements.

[2] For social transfer payments the government department responsible for the payments is identified in the employer field.

[3] Because of this, at present our exercise cannot be considered truly real-time however we emphasise that if transitions data were available within the quarter to which it corresponds, then this is a real-time forecast exercise.

To create our unemployment nowcast we use several popular machine learning models. These models include: lasso, ridge, elastic net, support vector machines (SVM), least squares boosted trees and partial least squares. We compare the nowcasts from these models to that of a benchmark AR(1). As a cross check we compare to an ARIMA model and a random walk model. An overview of these models is provided in the technical appendix. We train the models using the vintage of the unemployment rate that was available to the forecaster at each individual quarter estimated. We also seasonally adjust each estimation vintage separately to preserve as much of the real-time nature of this exercise as possible.

In order to assess the accuracy of our models we run a quasi-real-time out of sample forecast performance exercise for all of the algorithms. Then we measure the algorithms nowcast errors to the benchmark nowcast errors using ratios. Ratios under one indicate the machine learning algorithm is able to nowcast the unemployment rate more accurately than the benchmark model. A more in depth description of this process is available in the technical appendix.

4. How do the nowcasts from our proposed exercise compare to the benchmark model?

In this section we present the results of the expanding window nowcast performance test. Table one shows the RMSE's of the individual models in the first column and then the ratio of the individual models performance relative to the AR(1) benchmark model in the second column. In the third column we present the p-value associated with the significance test of each model.

Table 1: Forecast performance of machine learning models against benchmark

Models	RMSE	Ratio	P-Value
Elastic-Net	0.30	0.63	0.06
Lasso	0.33	0.69	0.09
LS-Boost	0.40	0.83	0.11
Ridge	0.45	0.94	0.45
PLS	0.45	0.94	0.66
SVM	0.46	0.96	0.77
AR(1)	0.48	1.00	-

The results indicate all the models that use the job transitions data are able to outperform the AR(1) for nowcasting the unemployment rate, with the elastic net performing the best overall.[4] If we then average the forecasts from the machine learning models we again outperform the AR(1) benchmark (table 2).

Table 2: Forecast performance of model average against benchmark

Models	RMSE	Ratio	P-Value
Model Average	0.33	0.69	0.03

The results from the model average indict that we can improve the nowcast accuracy of the AR(1) by 31%, this result is statistically significant at the 3% level.[5]

5. Conclusion

This note has assessed the usefulness of job transitions data for nowcasting unemployment rates in New Zealand. We used popular machine learning methods and found that they were able to produce nowcasts that are more accurate than conventional econometric models. This strategy could be helpful for policy makers to gain a better understanding of the current state of the labour market and therefore we conclude that more-timely job transitions data could be valuable.

In this paper we have tested its ability to nowcast the unemployment rate, however there is scope for us to test its ability to not only nowcast but also to forecast further horizons. This data could also help forecast other labour market variables as well, such as wage inflation or employment growth.

[4] We have also conducted the same experiment using both a random walk and an ARIMA model as our benchmark, we find that the AR(1) is the most competitive benchmark model and therefore we present the results relative to the AR(1).

[5] We have tested other alternative methods for forecast aggregation and found that the simple average gives the best accuracy.

References

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine learning 20 (3), 273–297.

Diebold, F. X., Mariano, R. S., July 1995. Comparing Predictive Accuracy. Journal of Business & Economic Statistics 13 (3), 253–263.

Friedman, J. H., 2001. Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189–1232.

Gunn, S. R., et al., 1998. Support vector machines for classification and regression. ISIS technical report 14 (1), 5–16.

Hansen, P. R., Lunde, A., Nason, J. M., 2011. The model confidence set. Econometrica 79 (2), 453–497.

Hecht-Nielsen, R., 1992. Theory of the backpropagation neural network. In: Neural networks for perception. Elsevier, pp. 65–93.

Hoerl, A. E., Kennard, R. W., 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12 (1), 55–67.

Kim, K., 2003. Financial time series forecasting using support vector machines. Neurocomputing 55 (1), 307 – 319.

McAdam, P., McNelis, P., 2005. Forecasting inflation with thick models and neural networks. Economic Modelling 22 (5), 848 – 867.

Richardson, A., Mulder, T., et al., 2018. Nowcasting New Zealand GDP using machine learning algorithms.

Scholkopf, B., Smola, A. J., 2001. Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press.

Schölkopf, B., Smola, A. J., 2003. A short introduction to learning with kernels. In: Advanced lectures on machine learning. Springer, pp. 41–64.

Teräsvirta, T., van Dijk, D., Medeiros, M. C., 2005. Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination. International Journal of Forecasting 21 (4), 755 – 774.