Economic Letter 2021/09

Forecasting unemployment with a model of labour flows

James P Vere
Senior Economist

October 2021

Abstract

This article applies Barnichon and Nekarda's (2012) two-state model of labour flows to the case of Hong Kong in two steps. In the first step, time series data on the employed and unemployed are used to estimate transition probabilities in and out of employment. In the second step, the transition probabilities are forecast with a VAR model, which yields one-step ahead forecasts of employment, unemployment, and the non-seasonally adjusted unemployment rate. Between 2001 and 2020, the model achieves a mean absolute error of 0.20 percentage point, performing notably better than a benchmark AR model.

以勞動力流向模型預測失業率

摘要

本文章以兩個步驟將 Barnichon and Nekarda's (2012)就勞動力流向的雙狀態模型應用於香港。第一個步驟是利用就業和失業人數的時間序列估算就業和失業兩種狀態之間的轉換機率(即由就業變為失業，或由失業變為就業的機率)。第二個步驟是以向量自回歸模型預測這些轉換機率，從而為就業人數、失業人數以及非季節性調整的失業率作出向前一步的預測。由 2001 年至 2020 年，這模型的平均絕對誤差為 0.2，表現明顯優於標準的自回歸模型。

I. INTRODUCTION

1. The unemployment rate has been a mainstay of labour market statistics since the 1930s, when the modern definition of the labour force was developed at the US Census Bureau.[1] Forecasts of the unemployment rate are of special interest during recessions, when nonlinear methods (e.g. threshold autoregressive models) designed to capture inherent asymmetries—rapid rises during contractions, followed by slower declines as the economy recovers—often perform better than simpler linear models, but they still lag significantly behind consensus professional forecasts.[2] More recently, however, Barnichon and Nekarda (2012) showed that an explicit model of flows in and out of employment, in which the stocks of employed and unemployed evolve according to VAR-modelled transition probabilities, is able to outperform both benchmarks.[3] This note applies Barnichon and Nekarda's method to the Hong Kong case.

2. This note is organised as follows. The next section describes the model, and the third section discusses the empirical results. The fourth section concludes.

II. METHODOLOGY

II.A. Modelling Labour Flows

3. Conceptually, a full model of labour flows would encompass three different pools of potential labour: the employed, the unemployed, and those out of the labour force entirely. With three different states, nine different transition probabilities would be needed to describe how the three pools evolve from one period to the next. However, estimating nine different transition probabilities can only be done with panel data on the labour force[4], which is not feasible in Hong Kong. Consequently, it is only possible to proceed with a two-state model, where the same pool of workers transitions between employment and unemployment. That said, though the three-state model may seem more appealing in theory, the two-state model actually performed better in Barnichon and Nekarda's (2012) analysis, possibly because transitions in and out of the labour force are less responsive to the business cycle.

[1] Card, D. 2011. "Origins of the Unemployment Rate: The Lasting Legacy of Measurement without Theory." *American Economic Review*, 101(3), pp. 552-557.

[2] Montgomery, A. L., V. Zarnowitz, R. S. Tsay and G. C. Tiao. 1998. "Forecasting the U.S. Unemployment Rate." *Journal of the American Statistical Association* 93(442), pp. 478-493.

[3] Barnichon, R. and C. J. Nekarda. 2012. "The Ins and Outs of Forecasting Unemployment: Using Labor Force Flows to Forecast the Labor Market." *Brookings Papers on Economic Activity*, Fall, pp. 83-117.

[4] Toikka, R. S. 1976. "A Markovian Model of Labor Force Decisions by Workers." *American Economic Review* 66(5), pp. 821-834.

4. In a two-state model, the pools of employed E and unemployed U evolve according to the formula

$$\begin{bmatrix} E_t \\ U_t \end{bmatrix} = \begin{bmatrix} (1 - P_{U,t}) & P_{E,t} \\ P_{U,t} & (1 - P_{E,t}) \end{bmatrix} \begin{bmatrix} E_{t-1} \\ U_{t-1} \end{bmatrix}$$

where t is a time subscript, P_E is the probability a person who is unemployed at time $t-1$ will become employed by time t, and P_U is the probability that a person who is employed at time $t-1$ will become unemployed by time t. Thus, the off-diagonal elements of the matrix (the transition probability matrix) are probabilities of switching from one pool to the other, while the diagonal elements are probabilities of remaining in the same pool. Since everyone has to end up in one pool or another, the columns of the transition probability matrix must sum up to one, and the two probabilities P_E and P_U characterise the whole system.

5. Time series data on E_t and U_t, by themselves, are not sufficient to identify $P_{E,t}$ and $P_{U,t}$. Intuitively, this is because E_t and U_t only provide information on net flows in and out of employment, not gross flows. For instance, if the pools of employed and unemployed were exactly the same size in periods $t-1$ and t, it would be impossible to tell whether this is because $P_{E,t}$ and $P_{U,t}$ are zero, or because they are positive but simply cancel each other out. However, if data on the pool of unemployed by duration are available, they can be used to solve for $P_{E,t}$, and then the flow equations can be used to solve for $P_{U,t}$.

6. In particular, given quarterly data on unemployment by duration, anyone who has been unemployed for at least three months is a job seeker in both the past quarter ($t - 1$) and the current quarter (t). Meanwhile, U_{t-1} is the total number of job seekers in the quarter $t - 1$. The ratio is $1 - P_{E,t}$, or the probability of staying unemployed between quarters $t - 1$ and t:

$$1 - P_{E,t} = \frac{U_{t,>3\ months}}{U_{t-1}}$$

$$P_{E,t} = \frac{U_{t-1} - U_{t,>3\ months}}{U_{t-1}}$$

7. In Hong Kong, data on the pool of unemployed by duration are available from the *Quarterly Report on General Household Survey* prepared by the Census and Statistics Department (C&SD).

8. After solving for $P_{E,t}$, $P_{U,t}$ can be worked out from the flow equations in para. 4. Either row could be used for this purpose, but the second is more convenient algebraically. Working from the equation for the pool of unemployed[5],

$$U_t = P_{U,t} \cdot E_{t-1} + \left(1 - P_{E,t}\right) \cdot U_{t-1}$$

$$P_{U,t} = \frac{U_t - \left(1 - P_{E,t}\right) \cdot U_{t-1}}{E_{t-1}}$$

Thus, the time series of E_t and U_t, together with the previous estimate of $P_{E,t}$, is sufficient to identify $P_{U,t}$.

II.B. Forecasting Labour Flows

9. By themselves, $P_{E,t}$ and $P_{U,t}$ are not necessary to project the unemployment rate in quarter t, which is already known from the information available ($\frac{U_t}{E_t+U_t}$, the unemployment rate without seasonal adjustment). Predicting how the unemployment rate will evolve, however, requires forecasts of $P_{E,t}$ and $P_{U,t}$.

10. One-step ahead forecasts of $P_{E,t}$ and $P_{U,t}$ can be constructed with a simple VAR model, which, similar to Barnichon and Nekarda's (2012) model, is specified with two lags. The exogenous variables in the VAR model are lag quarterly year-on-year real GDP growth (expressed as a difference in logs), the lag ratio of quarterly employment vacancies to the size of the labour force ($\frac{Vacancies_t}{E_t+U_t}$), and a set of quarterly indicators. The quarterly vacancy statistics are from C&SD's *Quarterly Report of Employment and Vacancies Statistics*, and are added in as an additional leading indicator to help predict how the transition probabilities will evolve in the future.

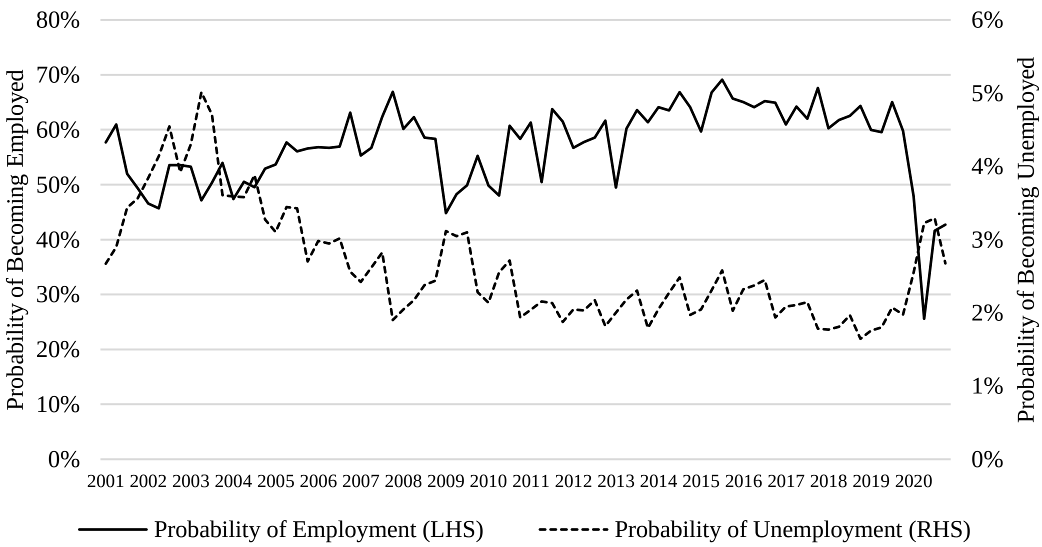
11. Once the once-step ahead forecasts of $P_{E,t}$ and $P_{U,t}$ are in hand, future values of E_t and U_t then follow from E_{t-1}, U_{t-1} and the flow equations in para. 4. The future values of E_t and U_t in turn determine the non-seasonally adjusted unemployment rate. For longer-term forecasts, $Vacancies_t$ can be added as an endogenous variable in the VAR. Future values of year-on-year GDP growth require an overall view on the economic outlook that is beyond the scope of the model, but could be sourced from forecasts published elsewhere.

[5] This equation says that unemployment pool at time t is the employed pool at time $t-1$ multiplied by the probability of unemployment (i.e., new entrants to the pool of unemployed), plus the unemployed pool at time $t-1$ multiplied by the probability of not finding employment (i.e., those who stay in the pool of unemployed).

III. RESULTS

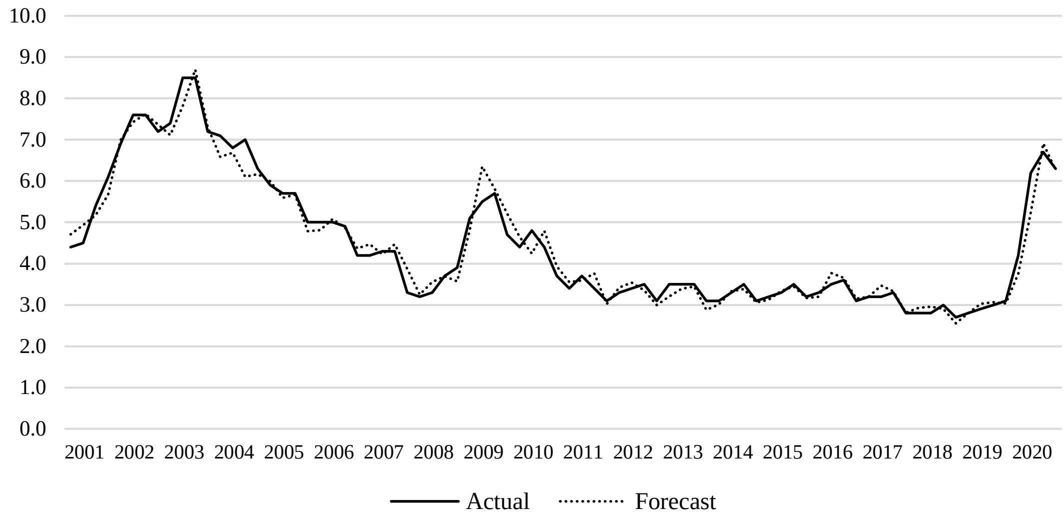
12. Estimates of P_E and P_U, the quarterly probabilities of transitioning from unemployed to employed and vice versa that underpin the flow model, are shown in **Chart 1**. A few observations are pertinent. First, the probability of becoming employed by the next quarter—which ranges around 50-60% for most of the two decades from 2001-2020—is much higher than the probability of becoming unemployed, which is intuitive because periods of employment are generally much longer than periods of unemployment. Second, as one would expect, the probabilities of employment and unemployment tend to move in opposite directions. Mirroring overall economic conditions, there are noticeable spikes in unemployment (and downward movements in employment) in 2002-03, 2009, and 2020 that mark the run-up to SARS, the global financial crisis, and the COVID-19 pandemic. Third, between 2003 and 2009, there is an extended secular decline in the probability of unemployment, reflecting Hong Kong's relatively rapid pace of economic growth in that period.

Chart 1: Probabilities of Employment and Unemployment



13. As mentioned in para. 10, the probabilities in **Chart 1** can be forecast one period ahead with a VAR model that includes real GDP growth, vacancy statistics, and quarterly indicators as exogenous variables. These in turn generate predicted totals of the numbers of employed and unemployed, which imply a non-seasonally adjusted unemployment rate. The unemployment rates forecast in this fashion, together with the actual unemployment rates, are shown in **Chart 2**.

Chart 2: Actual and Forecast Non-Seasonally Adjusted Unemployment Rates



14. **Chart 2** shows that the model is able to track unemployment fairly closely. There is a large miss of 1.0 percentage point in the second quarter of 2020, which is not surprising because the impact of COVID-19 was still mostly unforeseen at that time. Overall, the mean absolute error of the model is 0.20, significantly better than a benchmark AR(2) model[6] that yielded a mean absolute error of 0.25.

IV. CONCLUSIONS

15. Modelling labour flows in Hong Kong is challenging from a data perspective as, without a panel, it is impossible to directly observe transitions from one labour force status to another. However, with cross-sectional data and information on the duration of unemployment, it is possible to estimate a simplified two-state model of employment and unemployment, which performs reasonably well as a forecasting tool. The implied transition probabilities are also of useful reference value in summarising local labour market conditions at various times.

[6] Meyer and Tasci (2015), in reviewing various methods of forecasting unemployment, remark that even a simple AR(1) model "often becomes a measure of economists' collective ignorance, as it is hard to beat."

Meyer, B. and M. Tasci. 2015. "Lessons for Forecasting Unemployment in the United States: Use Flow Rates, Mind the Trend." Federal Reserve Bank of Atlanta Working Paper 2015-1.