CFP WORKING PAPER

Forecasting unemployment in Portugal: A labour market flows approach

Nuno Gonçalves

Domingos Seward

Working Papers

01 | 2019

Conselho das Finanças Públicas
Portuguese Public Finance Council

WORKING PAPERS

01 | 2019

Forecasting unemployment in Portugal: A labour market flows approach

Nuno Gonçalves

Domingos Seward

13 November 2019

CONSELHO DAS FINANÇAS PÚBLICAS

Praça de Alvalade, n.º 6 – 10.º

1700-036 Lisboa

TEL +351 211 024 400

www.cfp.pt

Publication

CFP Working Papers

Lisbon, 2019

Forecasting unemployment in Portugal: A labour market flows approach*

Nuno Gonçalves† Domingos Seward

November 13, 2019

Abstract

This paper applies a labour market flows approach to forecasting the unemployment rate, initially developed by Barnichon and Nekarda (2012) and subsequently extended by Barnichon and Garda (2016), to the Portuguese labour market. We start by implementing a simple two-state labour market forecasting model and then extend it to a three-state labour market forecasting model which incorporates movements in and out of the labour force. We test the forecasting accuracy of each of these models and find that the two-state flow-based forecasting model performs slightly better than the other tested models. We conclude that worker flow data is a valuable input for forecasting the unemployment in Portugal.

Keywords: forecast, labour market dynamics, unemployment rate, worker flows.

JEL codes: C53, E27, J60.

*Valuable inputs, comments, and suggestions by Erica Marujo, Reuben Ellul, Christopher Nekarda, Francisco Ruano, Rui Leite, João Jalles, Luís Centeno, and Alice Paulos are gratefully acknowledged. The views expressed in this article are those of the authors and do not necessarily reflect the views of Conselho das Finanças Públicas.

†Corresponding author. E-mail: ngoncalves@cfp.pt. Address: Conselho das Finanças Públicas, Praça de Alvalade, n.º6 10.º 1700-036 Lisboa.

1. Introduction

Unemployment is an economic indicator of the labour market of paramount importance, for example to evaluate the state of the business cycle, and to assess the slack in the labour market. Forecasting the unemployment rate of an economy is thus a crucial and demanding endeavour, and has become a decisive exercise for both policymakers and macroeconomists alike.

A multitude of forecasting techniques have been developed for this purpose. Traditionally, the unemployment rate has been forecast following two approaches: (*i*) exploiting the historical time-series properties of the rate of unemployment and near-term leading indicators of the labour market to forecast its future values (*e.g.* using simple econometric time-series models), and (*ii*) relying on the relationship between output growth and variations in the unemployment rate, *i.e.* the so-called Okun's law.

Early studies, however, have attempted to address forecasting challenges of the rate of unemployment (see Montgomery et al., 1998 and Rothman, 1998). Most of the focus of these works were on the apparent non-linear, asymmetric behaviour displayed by the unemployment rate over the business-cycle and, consequently, on the adequacy of linear models to forecast this variable. Another approach has applied the so-called neural network models for forecasting purposes. In addition, a recent strand of literature has emerged which uses labour market flows to forecast the unemployment rate. This approach was initiated by Barnichon and Nekarda (2012), and has subsequently been explored and extended by other authors, such as Tasci (2012) and Barnichon and Garda (2016).

As shown by Barnichon and Nekarda (2012) and Barnichon and Garda (2016), the flow-based approach to unemployment forecasting may yield significant forecasting gains *vis-à-vis* traditional forecasting techniques. In this paper, we apply the flow-based approach, as thoroughly discussed in Barnichon and Nekarda (2012) and Barnichon and Garda (2016), to the Portuguese economy. We start by developing a simple two-state model of the labour market. We then extend this model to a three-state model of the labour market, in order to more accurately capture the dynamics in the labour market.

This paper is structured as follows. In section 2, we review the relevant literature on unemployment forecasting methods. Subsequently, in section 3, we outline the two-state forecasting model for the Portuguese labour market. In section 4, we develop the three-state forecasting model for the Portuguese labour market. For both models, we present the methodological approach, the data construction issues, and the discussion of results. Finally, section 5 concludes.

2. Literature Review

In this section, we review the literature on the different approaches which have been developed for forecasting the unemployment rate. We briefly address three strands of literature: (*i*) time-series econometric models, including both linear and non-linear models, and the comparisons which have been developed between each, (*ii*) neural networks models, and (*iii*) the flow-based approach to labour market forecasting.

The focus is on presenting the main findings and on how the forecasting performance of each approach compares with other competing approaches. Moreover, we only address forecasting approaches aimed at forecasting the unemployment rate in the *short-term*.[1]

2.1 Time-series models

Rothman (1998) was among the first to evaluate the out-of-sample forecasting performance of several non-linear time-series models in comparison to linear time-series models. Using post-war aggregate unemployment rates for the USA, the author aims at determining how the non-linearities in the unemployment rate might be helpful in modeling this variable. In total, six non-linear time-series classes of models are applied in the forecasting exercise: exponential autoregressive (EAR), generalised autoregressive (GAR), self-exciting threshold autoregressive (SETAR), smooth threshold autoregressive (STAR), bilinear, and time-varying autoregressive (TVAR) models.[2] The forecasting performance of these models is compared to more conventional linear time-series models, *e.g.* AR models. The author finds that indeed one can achieve improvements in forecast performance by applying non-linear

[1]Several institutions regularly provide forecasts for labour market variables. However, most of them do so by constructing a model for the labour market, *e.g.* the Netherlands Bureau for Economic Policy Analysis (CPB) forecasts the labour supply via a disaggregated forecasting model, which employs statistical filtering techniques and cohort models, and forecasts employment using a macroeconomic demand-driven model. It then derives its unemployment forecast by subtracting the forecasted value of employment from the forecasted value of the labour supply. In this note, we focus on models aimed at solely forecasting the unemployment rate. See Zailstra and Boxhoorn (2017) for a complete overview on short- to medium-term labour supply forecasting models used by fourteen research institutes.

[2]The dynamics which govern such time-series models are "state dependent", whereby they vary according to the past behaviour of the variable.

time-series for the US unemployment rate. Such a finding provides evidence supporting the view that non-linear forecasts may dominate linear forecasts for this variable.[3] Nevertheless, the relative forecasting performance of the models depends on whether one transforms the unemployment rate series to stationarity.[4] The highest forecast improvement was obtained with the subset EAR and GAR models.

In the same vein, Montgomery et al. (1998) compare the forecasting performance of several linear and non-linear time-series models using data on the US unemployment rate. The authors focus on evaluating the forecast performance over economic expansions and recessions, by exploiting the above-mentioned asymmetric behaviour of the unemployment rate. The models analysed include linear univariate autoregressive integrated moving average (ARIMA) models, bivariate vector autoregressive moving average (VARMA) models, threshold autoregressive (TAR) models, and Markov-switching autoregressive (MSA) models. The median of the consensus group forecasts of the Survey of Professional Forecasts (SPF) is also compared. Overall, the authors find that non-linear models can considerably improve the forecasting performance over certain periods.[5]

Johnes (1999) reports a forecasting competition between linear AR, generalised autoregressive conditional heteroskedasticity models (GARCH), SETAR, and neural network models for the UK unemployment rate. The results once again indicate that non-linearities are present and that forecasting performance may be enhanced by modelling such non-linearities in the unemployment rate series.

Proietti (2003) investigates the out-of-sample forecast performance of linear and non-linear structural time-series models for the US unemployment rate. Several linear unobserved components models are used and their forecast accuracy is assessed

[3]This view is rooted on the so-called business-cycle asymmetry hypothesis, which posits that economic expansions take longer but are less sharp than economic downturns. This implies that the unemployment rate increases relatively quickly in recessions, but decreases relatively slowly during expansions. Such a non-linear phenomenon may not be well captured by the conventional linear time-series models.

[4]Rothman (1998) finds that, if the series is not transformed, some of the non-linear models deliver biased estimates.

[5]Such a finding is consistent with evidence reported by Clements and Krolzig (2003) and Koop and Potter (1999), regarding the existence of both statistically and economically significant non-linearities in the US unemployment rate series.

via the application of an extensive rolling forecast exercise. The results corroborate the so-called hysteresis hypothesis[6], since linear models which reflect higher persistence are found to deliver considerably better forecast results. Overall, Proietti (2003) concludes that structural time-series models may prove to be a useful forecasting tool, and can be easily extended to reflect non-linear dynamics.

Drawing on the research developed by Montgomery et al. (1998), Golan and Perloff (2004) apply a non-linear, non-parametric approach to forecast the unemployment rate. The authors extend the nearest-neighbour method by applying a higher dimensional simplex approach, composing a simplex which contains the point they wish to forecast. The authors conclude that, as a result of the non-linearity which characterises the data-generating process, the non-parametric method outperforms other well-known structural and economic-theory-based models.[7]

Milas and Rothman (2008) apply Smooth Transition Vector Error-Correction Models (STVECM's) in a simulated out-of-sample forecasting exercise for the rates of unemployment of several non-European countries.[8] For the USA, the authors find that pooled forecasts constructed by computing the median value across the point forecasts delivered by the linear and STVECM forecasts perform better relative to the linear AR benchmark forecasting model. Moreover, such a finding seems to be most prominent over economic expansions.[9]

Chua et al. (2012) propose a more structural economic-theory-based forecasting model for the unemployment rate, which aims at exploiting its time-series properties, while still satisfying the economic relationships reflected in Okun's law and the Phillips curve. These two economic relationships jointly relate the movements in output, prices, and unemployment. The authors consider the problem of estimating

[6]The hysteresis or persistence hypothesis is based upon the observation of a tendency of the series to remain at a given level that it has reached, with no apparent tendency to revert to a stable implicit level.

[7]Montgomery et al. (1998) note that the superior forecast performance of the proposed highly non-linear, non-parametric approach might be due to the fact that the traditional, relatively straightforward time-series models and even the more complex econometric models are unable to capture the high dimensionality and strong non-linear structure of the series.

[8]These include: the USA, the UK, Canada, and Japan.

[9]The authors believe that, while non-linear forecasting methods might occasionally dominate a linear approach, improvements in forecasting appear to be attained by combining results across the set of linear and non-linear forecasts.

the (unobserved) potential unemployment rate which is consistent with Okun's law and the Phillips curve, and relating it with the actual rate of unemployment. They do so by constructing a latent variable forecasting model which satisfies such economic relationships. The authors present an empirical application which shows that the proposed model outperforms alternative forecasting techniques used for forecasting the rate of unemployment. In particular, the performance of the model is assessed against forecasts provided by Vector-Autorregressive (VAR) and Bayesian Vector-Autorregressive (BVAR) models.

2.2 Neural networks models

Artificial neural networks are based upon the biological and physiological principles underpinning the human brain. Neural networks learn associations, patterns, and functional relations in an inductive fashion by following the data which is supplied (see Haykin, 1994). Neural networks provide certain advantages relative to alternative modeling and forecasting techniques. First, neural networks are not programmed. Rather, neural networks are "trained" by exposing them to individual data examples which will be later used for forecasting or classification purposes. This process is iterated until the network recognises the patterns and relations underpinning the inputs (independent variables) and the outputs (dependent variables). Second, neural networks usually do not require any assumption regarding the data to be forecasted. Therefore, as opposed to other modelling procedures, tests need not be conducted for the assumptions. Lastly, neural networks are able to produce models from imperfect or incomplete data. For most other modelling procedures, missing data can be harmful for the estimation. Thus, neural networks are capable of developing relatively accurate models with a quantification of acceptable fault tolerance, even though some data might be missing.

Aiken (1996) applies neural networks for forecasting the unemployment rate, using leading economic indicator data. The results show that the neural network delivers superior forecasts of the unemployment rate one month ahead relative to multi-linear regression and two naïve forecasting procedures.

Brown and Moshiri (2004) compare the out-of-sample forecast performance delivered by the artificial neural network models[10] relative to other linear and non-linear time-series models popular in the literature. The authors show that the artificial neural networks are capable of forecasting the unemployment rate series as well as, and, in some occasions, even better than the other univariate time-series models under consideration.[11] The authors argue that the artificial neural networks provide a promising solution for the difficulty of forecasting the unemployment rate series over the asymmetric business-cycle, considering that artificial neural networks are non-linear, do not need to rely on the classical regression assumptions, are able to learn the structure of patterns present in the data with a given degree of accuracy, and can then use such a structure for forecasting purposes.

Other applications include Olmedo (2014), which has also employed alternative non-parametric forecasting procedures to the Spanish unemployment rate, namely artificial neural networks, with satisfactory results.[12] Furthermore, Stasinakis et al. (2016) investigate the efficiency of radial basis function neural networks for forecasting the US rate of unemployment, as well as explore the use of Kalman filtering and support vector regression as potential forecast combination procedures. The results show that the proposed neural network statistically outperforms all the other model's individual forecasting performances. In addition, the forecast combinations are shown to be successful, given that both the Kalman filter and the support vector regression procedures improve the accuracy of the forecasts. Nevertheless, the support vector regression is shown to be the single most accurate model of the forecasting competition.

[10]In particular, the authors develop two non-linear artificial neural network models, *i.e.* a back-propagation neural network model (BPNN) and a generalised neural network model (GNN).

[11]These include a linear AR model and two non-linear GAR and EAR models.

[12]The author focuses on the difference between the reconstruction and learning approaches.

2.3 Flow-based approach

Recently, a method that uses data on unemployment flows to forecast the unemployment rate has been developed by Barnichon and Nekarda (2012).[13] This work extends the growing literature aimed at improving forecasts of the rate of unemployment.[14] It also draws heavily on the relatively recent research on labour market flows, which has been the subject of several investigations aimed at evaluating the determinants of labour market fluctuations.[15] The authors argue that the inclusion of labour market flows should be expected to improve forecasts of the unemployment rate because these are the underlying drivers of fluctuations in the aggregate unemployment. Moreover, since the individual flows exhibit distinct time-series properties and the respective contributions vary over the business-cycle, by focusing on the flows one may be able to better capture the asymmetric pattern underlying the unemployment fluctuations.

This method takes advantage of the convergence property, whereby the actual unemployment rate will converge to the rate which is implied by the gross worker flows. This concept is coined the *conditional steady-state unemployment rate* by the authors, the unemployment rate which is pinned down by the flows in and out of unemployment.[16] The model relies on two main elements: (*i*) a non-linear law of motion which captures how the unemployment rate converges toward the steady-state and the convergence speed to steady-state, and (*ii*) a time-series forecasting procedure of the labour market flows. The authors develop both a two-state model of the labour market, which only considers employment and unemployment, and a more general three-state model of the labour market, which considers transitions

[13] This novel approach to unemployment forecasting has been adopted by the International Labor Organization for forecasting unemployment in G7 countries (ILO, 2015). In particular, a multitude of models are constructed that either forecast the unemployment directly or forecast both the unemployment inflow and outflow rates, by applying ARIMA, Vector Autoregressive Model with exogenous variables (VARX), as well as combined forecast procedures.

[14] See, as previously explored, Montgomery et al. (1998), Rothman (1998), Golan and Perloff (2004), Brown and Moshiri (2004), and Milas and Rothman (2008).

[15] See, *e.g.* Petrongolo and Pissarides (2008), Fujita and Ramey (2009), Solon et al. (2009), Shimer (2012), Barnichon (2012), and Elsby et al. (2013).

[16] The authors show that there is a consistent leading relationship between the steady-state unemployment rate and the observed unemployment rate.

into and from inactivity.

In the two-state model, Barnichon and Nekarda (2012) use information on the stocks of unemployment and short-term unemployment to compute the worker flows and respective transition rates, following the methodology put forward by Shimer (2012).[17] In the three-state model, longitudinally matched Current Population Survey (CPS) micro data are used to compute the aggregate labour market transition probabilities.

In both models, a VAR is used to forecast the outflow and the inflow rates.[18] Subsequently, in the two-state model, Barnichon and Nekarda (2012) obtain the forecasts of unemployment by iterating the law of motion (fed by the previously forecasted values of the hazard rates), using Monte-Carlo simulation. In the three-state model, the stocks of unemployment, employment and inactivity satisfy a system of differential equations which can be solved to obtain the forecasts of the three stocks. In turn, these are used for computing the forecasted unemployment rate (and labour force participation rate).

Barnichon and Nekarda (2012) find that the flow-based forecasting model outperforms at short horizons basic time-series models (the univariate VAR and ARIMA models), the Survey of Professional Forecasters, and the Federal Reserve Board's Greenbook. The proposed model displays the highest predictive ability *vis-à-vis* alternative forecasts over business-cycle turning points and recessions.

Barnichon and Garda (2016) evaluate the flow approach to forecasting the unemployment rate for several Organization for Economic Cooperation and Development (OECD) countries.[19] [20] Consistent with the seminal contribution by Barnichon and Nekarda (2012), the authors find that the flow approach improves considerably on

[17] Shimer (2012)'s method relies on duration data, and has subsequently been adopted and extended by, *e.g.* Fujita and Ramey (2009) and Elsby et al. (2013).

[18] Additionally, two leading indicators of labour market flows are included in the VAR: vacancy postings and initial claims for unemployment insurance.

[19] The authors employ the forecasting method for France, Germany, Spain, the UK, Japan, and the USA.

[20] Ellul (2018) applied the method by Barnichon and Nekarda (2012) to the Maltese labour market, and extends it to other estimation techniques, *e.g.* BVAR and Vector Error Correction Models (VECM). The author also reports improvements in forecast accuracy, especially over the short term.

conventional forecasting methodologies in all the countries under the scope of their study. In addition, forecast improvements are highest in the one- to three-months horizons for the USA, and in the one-year horizon for European countries, where the magnitude of labour market flows is comparatively smaller.[21]

Tasci (2012) proposes an empirical method for determining the trend in the rate of unemployment in the long term. Even though Tasci (2012) focuses on the long-run dynamics of the unemployment rate and the respective flow rates, the proposed model can be applied for forecasting the unemployment rate.[22] The author proposed an unobserved components model, whereby it assumes that real Gross Domestic Product (GDP) and the unemployment flow rates are composed of a stochastic trend and a stationary cyclical component. While the stochastic trend of the flow rates follows a random walk, the cyclical component depends on the cyclical component of real GDP.[23] Meyer and Tasci (2015) evaluate the forecasting performance of the model. In particular, the authors evaluate the ability of AR models, professional forecasters, and models which include flows (the more reduced form method by Barnichon and Nekarda, 2012, and the more structural approach by Tasci, 2012) in order to forecast the unemployment rate. The results indicate that any approach which incorporates unemployment flows performs well in the short-term. The flow-based method by Tasci (2012) appears to be suitable in the long-term, and in specific business-cycle points, when the unemployment rate deviates from its "natural" rate. Furthermore, professional forecasts are found to be the best predictor of future unemployment, and the combination of the former with flow-based forecasts yields considerable forecasting improvements.

[21]The authors further show that these improvements stem from (*i*) the use of worker flows and (*ii*) non-linearities present in the unemployment law of motion, particularly, the time-varying steady-state and respective convergence rate.

[22]Sengul and Tasci (2014) follow the methodology put forward by Tasci (2012) to forecast Turkey's unemployment rate. The authors extend the unobserved components model by Tasci (2012) in order to estimate the long-run rate of unemployment while still allowing for changes in the labour force participation. They evaluate the forecasting performance of both the baseline model and the extended model accounting for participation over time. The results indicate that both approaches yield more accurate forecasts *vis-à-vis* a time-series AR model, especially in the near-term horizon.

[23]The model thus incorporates the comovement of the unemployment flows with the aggregate output.

2.4 Summary and conclusions from the literature

Based on the previous evidence on the different methodologies which have been developed for forecasting the unemployment rate in the short-term, we conclude that:

- There are three main approaches to forecasting the rate of unemployment: (*i*) time-series (linear and non-linear) models, (*ii*) neural networks models, and (*iii*) flow-based models.

- Two results are common to most authors: (*i*) by appropriately modelling the non-linearities in the series of the rate of unemployment, one can achieve considerable forecasting improvements relative to simple linear techniques, and (*ii*) combining forecasts from different approaches may enhance forecasting performance.

- The data requirements vary depending on which model, methodology, and variables are applied. For time-series and neural networks models, the data requirements are modest (most of literature applies the unemployment rate series, as well as other leading indicators of the labour market). The two-state flow-based approach to labour market forecasting requires duration data on unemployment (stocks of unemployment and short-term unemployment), while the three-state flow-based approach requires Labour Force Survey (LFS) microdata. In addition, data on other variables may be required, namely for employing the VAR procedure to forecast the unemployment flows.

- One can identify several trade-offs in the strategy to forecast the unemployment rate: (*i*) between structure and flexibility - imposing structure to the forecasting model may improve its performance, but it is prone to the issue of misspecification of the underlying theoretical relationships, which may have undesirable effects in terms of forecasting accuracy; (*ii*) between details and modelling ease - the inclusion of more details by, *e.g.* including more disaggregated data, may also improve forecasts, but modelling them in a consistent framework is generally more demanding.

- Overall, in terms of practical lessons for forecasting the unemployment rate, one may conclude that the flow-based approach appears to be a promising avenue to pursue, due to its modest data requirements (especially in the two-state forecasting model by Barnichon and Nekarda, 2012) and its (apparent) parsimony.[24] [25]

[24] This conclusion is drawn from the results of the pioneering authors of this approach, which report unusually high forecasting improvements, and on the fact that several institutions have declared their intentions to farther their labour market forecasting techniques to include labour market flows (see, *e.g.* Zailstra and Boxhoorn, 2017) or have already adopted this approach (see, *e.g.* ILO, 2015).

[25] The relatively recent research on the determinants of labour market fluctuations (see Shimer, 2003; Petrongolo and Pissarides, 2008; Solon et al., 2009) have highlighted the importance of labour market flows in driving the labour market stocks. The inclusion of such data in forecasting models should thus be expected to yield improvements in performance.

3. A two-state forecasting model of the labour market

3.1 A stock-flow model of unemployment

As extensively discussed by Barnichon and Garda (2016), in the two-state model, individuals are assumed to only occupy two states: employment or unemployment. Let $u_{t+\tau}$ be the unemployment rate at instant $t+\tau$, with t representing the period (*i.e.*, 1 quarter) and $\tau \in [0,1]$ representing a continuous measure of time within the period. The authors further assume that between t and $t+1$ all the unemployed workers find a job according to a Poisson process with a constant arrival rate denoted by f_{t+1} (outflows of unemployment), and all employed workers lose their job according to a Poisson process with constant arrival rate s_{t+1} (inflows of unemployment).

The evolution of the unemployment rate is governed by:

$$\frac{du_{t+\tau}}{d\tau} = s_{t+1}(1-u_{t+\tau}) - f_{t+1}u_{t+\tau}\,. \tag{3.1}$$

The solution to equation (3.1) is given by:

$$u_{t+\tau} = \beta_{t+1}(\tau)u^*_{t+1} + [1-\beta_{t+1}(\tau)]u_t\,, \tag{3.2}$$

where the steady-state unemployment rate (SSUR)[1] is obtained as:

$$u^*_{t+1} \equiv \frac{s_{t+1}}{s_{t+1}+f_{t+1}}\,, \tag{3.3}$$

and the rate of convergence to the SSUR:

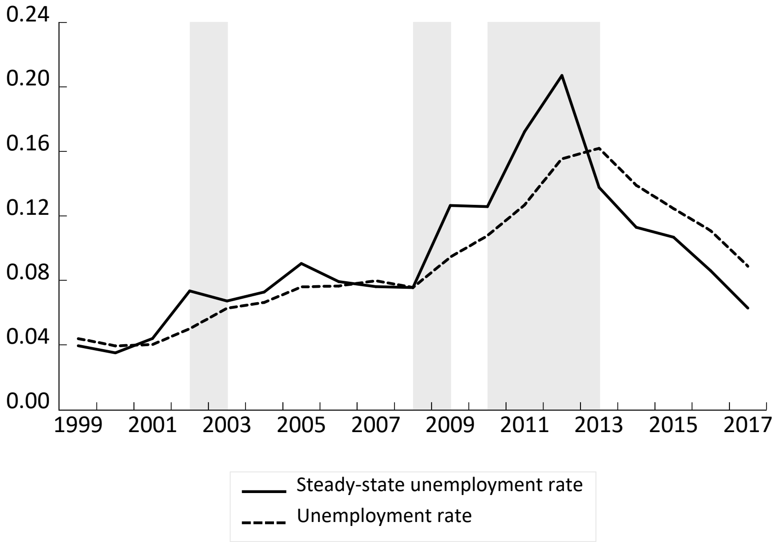
$$\beta_{t+1}(\tau) \equiv 1 - exp[-\tau(s_{t+1}+f_{t+1})]\,. \tag{3.4}$$

Figure 3.1 shows that there is a consistent leading relation between the SSUR and the observed unemployment rate for Portugal, consistent with the results obtained

[1] The SSUR is the rate of unemployment that would eventually be attained if the inflows and outflows of unemployment remained at their current values forever (Barnichon and Garda, 2016).

by Barnichon and Garda (2016) for a range of other OECD countries. This graphical observation is confirmed by Table A.1, which reports the cross-correlations between the unemployment rate and the steady-state unemployment rate.[2]

FIGURE 3.1: The unemployment rate and the steady-state unemployment rate in Portugal



Source: Authors' calculations based on data by INE;
Notes: For clarity, quarterly data have been annualised. The unemployment rate is the dashed line, and the steady-state unemployment rate is the solid line. The shadings signal recessions according to the business-cycle dating methodology by Bry and Boschan (1971).

[2]The steady-state unemployment rate is computed for different values of the parameter d, used in the calculation of the flow rates, as described in section 3.3.

3.2 The flow-based unemployment forecasting model

The applied flow-based unemployment forecasting model consists of two stages: (i) a forecast of the gross worker flows which determine the current and future values of the SSUR, and (ii) an iteration of the unemployment law of motion (see equation 3.2).

3.2.1 First step: to forecast the labour market flows

To forecast the flows, we apply a Vector AutoRegression (VAR), which includes leading indicators of the labour market flows, in accordance with Barnichon and Nekarda (2012). In particular, the VAR considers a vector of the form:

$$\boldsymbol{y}_t = \big(log(s_t), log(f_t), \Delta log(u_t), \Delta log(v_t), \Delta log(gdp_t), \dots\big)', \tag{3.5}$$

where s denotes the separation rate, f denotes the job-finding rate, u denotes the unemployment rate, v denotes the number of vacancies, and gdp denotes the gross domestic product. The following VAR with n lags is estimated:[3]

$$\boldsymbol{y}_t = \boldsymbol{c} + \boldsymbol{\phi}_1 \boldsymbol{y}_{t-1} + \boldsymbol{\phi}_2 \boldsymbol{y}_{t-2} + \cdots + \boldsymbol{\phi}_n \boldsymbol{y}_{t-n} + \boldsymbol{\epsilon}_t \,. \tag{3.6}$$

Since many specifications for the VAR are possible and given that the best-performing specification depends on the country under analysis, the specification in equation (3.5) is illustrative. The VAR specification takes into account several criteria[4], namely the lag length (varying from 1 to 2 quarters), the choice between values in logarithms or in levels, and the choice for taking first-differences. The selected specification corresponds to the model which generates the smallest average Root Mean-Square Errors (RMSE) across the different forecast horizons.

[3] Barnichon and Nekarda (2012) use a 15-year rolling window and Barnichon and Garda (2016) use a 10-year rolling window. Due to sample size considerations, we have used a 10-year rolling window in our study.

[4] Consistent with Barnichon and Garda (2016), the results change little with alternative specifications.

3.2.2 Second step: to iterate the unemployment's non-linear law of motion

Considering the set of worker flows forecasts, the second step consists of iterating on equation (3.2) (the unemployment's law of motion). In particular, given the forecasts of the flow rates $\hat{f}_{t+j|t}$ and $\hat{s}_{t+j|t}$, with $j \in \mathbb{N}$, the j-period-ahead forecast for the rate of unemployment, $\hat{u}_{t+j|t}$, can be constructed recursively from the following equation:

$$\hat{u}_{t+j|t} = \hat{\beta}_{t+j|t}\hat{u}^*_{t+j|t} + (1 - \hat{\beta}_{t+j|t})\hat{u}_{t+j-1|t}\,, \tag{3.7}$$

where:

$$\hat{u}^*_{t+j|t} = \frac{\hat{s}_{t+j|t}}{\hat{s}_{t+j|t} + \hat{f}_{t+j|t}}\,, \tag{3.8}$$

and:

$$\hat{\beta}_{t+j|t} = 1 - exp[-(\hat{s}_{t+j|t} + \hat{f}_{t+j|t})]\,. \tag{3.9}$$

In other words, the unemployment forecast at $t+j$ is given by a weighted average of the previous-period $(t+j-1)$ unemployment forecast (or the observed real-time unemployment rate when $j=1$) and the time $(t+j)$ steady-state unemployment rate, in which the weights are equal to the speed of convergence to the steady-state.[5]

3.3 Data

The flow series are constructed from seasonally-adjusted data[6] on the stocks of unemployment, U_t, and short-term unemployment[7], $U_t^{<d}$, following the methodology put forward by Shimer (2012) and subsequently extended by Elsby et al. (2013) and Barnichon and Garda (2016). The data sources are described in Table A.2.

We calculate the unemployment outflow probability, F, from:

[5] The authors note that the speed of convergence and therefore the weights are also time-varying, which implies that the unemployment law of motion is non-linear.

[6] The data is seasonally-adjusted by applying the Census X-13 method.

[7] Shimer (2012) and Barnichon and Nekarda (2012) use information for unemployed workers less than 5 weeks, *i.e.* $d = 1$ month.

$$F_{t+1} = 1 - \frac{U_{t+1} - U_{t+1}^{<d}}{U_t}\,, \tag{3.10}$$

with $f_{t+1} = -log(1-F_{t+1})/d$ the monthly hazard rate associated with the probability that an unemployed worker at time t completes the spell within the subsequent d months.[8] For robustness, we have computed the flow rates for Portugal for $d = 6, 12$, and 24.[9] The unemployment inflow rate, s, is then obtained by solving equation (3.1) forward over $[t, t+1]$ and finding the value s_{t+1} that solves:

$$U_{t+1} = \frac{\{1 - exp[-(f_{t+1} + s_{t+1})]\}s_{t+1}}{f_{t+1} + s_{t+1}}(U_t + E_t) + exp[-(f_{t+1} + s_{t+1})]U_t\,. \tag{3.11}$$

As discussed by Barnichon and Garda (2016), in this flow accounting model, given the unemployment outflow rate (which also describes movements out of the labour force), and the stock of unemployed workers, the inflow rate describes the observed stock of unemployment in the following quarter. Consequently, the inflow rate incorporates all the movements in unemployment not captured by the outflow rate.

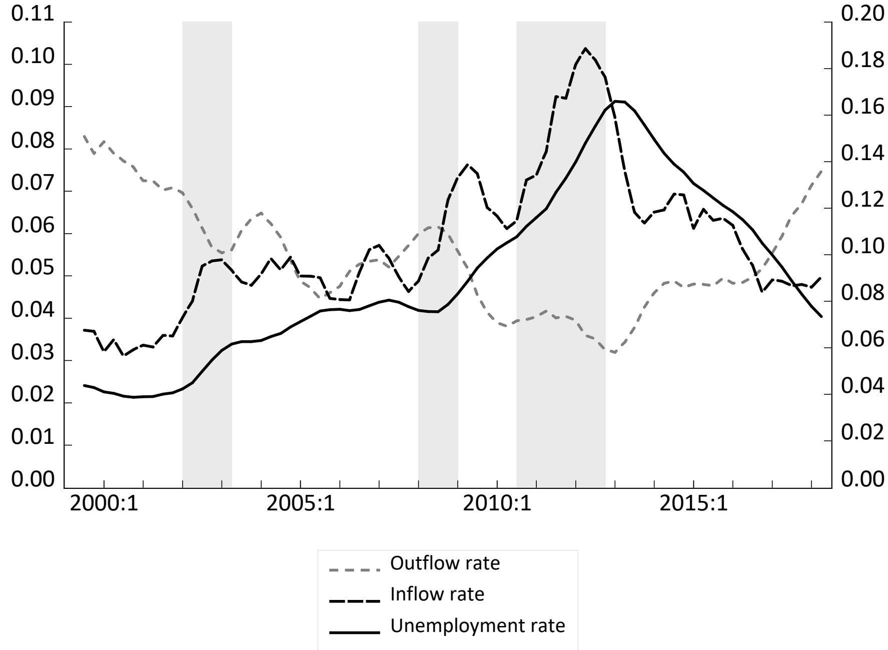
The unemployment rate and the computed flow rates for Portugal are plotted in Figure 3.2. In line with the findings by Elsby et al. (2013) and Barnichon and Garda (2016), the average magnitude of the flow rates place Portugal in the group of Continental European countries, which typically exhibit outflow rates of less than 10% and inflow rates at around 0.5 and 1%, in stark contrast with Anglo-Saxon and Nordic countries, which display outflow and inflow hazards in excess of 20% and 1.5% at a monthly frequency, respectively (as can be readily inferred from Figure B.1). It is worth pointing out that these results influence the business-cycle variation of the convergence rate, since β_t depends on the levels of the corresponding flow rates (see Figure B.2).

Furthermore, consistent with the results presented by Barnichon and Garda

[8] Barnichon and Garda (2016) use $d = 12$ months for Spain, France, and Germany, $d = 6$ months for the UK and Japan, and $d = 1$ for the USA, consistent with the differences in the magnitudes in the flows in each of these countries.

[9] We choose to report the estimation results in section 3.4 only for $d = 12$. We point out that the results do not vary significantly , and the main findings hold for different values of d.

FIGURE 3.2: The unemployment inflow and outflow rates in Portugal



Source: Authors' calculations based on data by INE;
Notes: Quarterly data are smoothed using a four-quarter moving-average procedure. The unemployment rate and the flow rates are plotted in the right axis and left axis, respectively. The unemployment inflow rate is rescaled as $s_t/E(u_t)$, for clarity. The shadings signal recessions according to the business-cycle dating methodology by Bry and Boschan (1971).

(2016) for other OECD countries, we find that, while the inflow and outflow rates move over time according to the business-cycle, they also exhibit considerable persistence, which might be an indication that the contemporaneous observations for the flow rates contain information on the values of the steady-state unemployment rate in the future, and therefore on the future values of the unemployment rate itself (see Table A.3 for the time-series properties of the computed flow rates).

3.4 Discussion of results

In line with Barnichon and Garda (2016), we employ four different forecasting techniques: (*i*) the flow-based forecasting (FbF) model, thoroughly discussed in section 3.2; (*ii*) a simple benchmark AutoRegressive Integrated Moving Average forecasting (ARIMA) model[10], (*iii*) a Vector AutoRegressive (VAR) model which does not include information on labour market flows[11], and (*iv*) a VAR model which includes information on labour market flows. The data used in the VAR estimations (both for the FbF approach and the simple VAR's for forecasting unemployment directly) are presented in Table A.2 in Appendix.

We test the forecasting performance of each model for different specifications as described in section 3.2, and select the best average performing specification for each case. The respective RMSE's are reported in Table 3.1. In order to assess the statistical significance of the results, we conduct the unconditional Giacomini and White (2006) predictive ability test for equal predictive ability between the forecast obtained by the FbF model and the comparison forecast from alternative models.[12]

At a 10 year-rolling window, we find that the FbF model outperforms all the alternative models tested at all forecasting horizons. Moreover, the forecasting gains of the FbF model tend to be higher at two and three quarter horizons.

Several conclusions illustrating the elements underpinning the superior performance of the FbF model are worth pointing out. First, we compare the forecasting performances between the two VAR models and thus assess the value-added of including worker flow data in a forecasting exercise for the unemployment rate. The VAR which does not include worker flow data may be considered a standard approach to forecast the rate of unemployment and lays the foundations of the FbF method. We find that the VAR which includes worker flow data outperforms the

[10]This model is performed in `EViews` via the `autoarma` command, which automatically selects the specification of the model for each forecasting window.

[11]In particular, the job-finding rate and the separation rate.

[12]An alternative predictive ability test would be the Diebold and Mariano (1995) test. As argued by Barnichon and Garda (2016), we choose to report the Giacomini and White (2006) test because, unlike the former, it is robust to not only non-nested, but also nested models (*e.g.* the VAR and FbF models).

Table 3.1: Absolute RMSE of FbF model and RMSE of alternative models relative to FbF model

Forecasting model	Specification	Window	Lags	t+0	t+1	t+2	t+3	t+4	t+8	Average(RMSE)
FbF	$\Delta log(f)\ \Delta log(s)\ \Delta(u)\ \Delta(gdp)\ \Delta(v)\ \Delta log(vabconstruct)$	**40 quarters**	**1**	**0.590**	**0.925**	**1.107**	**1.426**	**1.952**	**4.046**	**1.674**
VAR	$\Delta log(f)\ \Delta log(s)\ \Delta(u)\ \Delta(gdp)\ \Delta(v)\ \Delta log(vabconstruct)$	40 quarters	1	1.154*	1.176	1.295**	1.282**	1.231*	1.081	1.175
				(0.090)	(0.105)	(0.032)	(0.043)	(0.081)	(0.620)	–
VAR no flows	$\Delta(u)\ \Delta(gdp)\ \Delta(v)\ \Delta log(vabconstruct)$	40 quarters	1	1.135	1.154	1.288**	1.299**	1.241**	1.086	1.178
				(0.138)	(0.156)	(0.038)	(0.042)	(0.092)	(0.621)	–
ARIMA	–	40 quarters	–	1.101	1.166	1.304***	1.230**	1.195*	1.108	1.169
				(0.356)	(0.136)	(0.009)	(0.045)	(0.054)	(0.374)	–

Source: Authors' calculations;
Notes: The results concern the best average performing model specifications;
The RMSE's of the VAR with flow data, the VAR without flow data, and the ARIMA models are reported relative to the RMSE's obtained for the FbF model;
Asterisks denote the results of the unconditional Giacomini and White (2006) predictive ability test for equal predictive ability between the forecast obtained by the FbF model and the comparison forecast from alternative models:
* denotes statistically different from FbF at 10% level;
** denotes statistically different from FbF at 5% level;
*** denotes statistically different from FbF at 1% level.

VAR which does not include worker flow data at longer horizons (see Table 3.1), an indication that worker flow data is a valuable input for forecasting the unemployment rate at longer horizons in Portugal.

Second, we compare the forecasting performance between the VAR model including worker flow data with the FbF model and thus assess the importance of the non-linearities included in the unemployment law of motion (in particular, the time-varying nature of the SSUR and the convergence rate, β_t). We conclude that the FbF model outperforms the VAR model at all forecasting horizons and therefore the non-linear relation implied by the theory (section 3.2) is indeed quantitatively valuable for forecasting the rate of unemployment.

Third, we compare the forecasting performance between the benchmark univariate ARIMA model with the FbF model. The former uses only information on the stocks while the latter explores the use of flow variables for forecasting purposes. Since the FbF outperforms the ARIMA model at all forecast horizons, we conclude that the inclusion of flow variables does yield improvements in forecast accuracy.

As argued by Barnichon and Garda (2016), the observation that a stock model (as is the stock-based VAR) cannot perform as well as the FbF model is related with the fact that the labour market flows exhibit distinct time-series properties [13], as well as the contribution by the different flows varies over the business-cycle (see Barnichon, 2012). A model of the stock can capture the average time-series properties exhibited by the stock; however, it is unable to capture different time-series properties across the business-cycle. Particularly, the rate of unemployment exhibits "steepness asymmetry", *e.g.* the increases are steeper compared with decreases. A stock-based model as the VAR model is unable to capture this asymmetry, which is particularly striking during recessionary periods. The FbF model, on the other hand, by including worker flow information that is responsible for the asymmetry of the rate of unemployment, does capture this asymmetric behaviour. In fact, the onset of a recession is usually characterised by a sudden increase in the rate of inflow. We illustrate this pattern by plotting the impulse response functions to

[13] As reported in Table A.3, the autocorrelation observed for the outflow rate is consistently higher *vis-à-vis* the one for the inflow rate.

a one-standard-deviation shock to the unemployment inflow rate from a VAR estimated over the sample period (Figure B.3 in appendix). One can observe that the inflow rate exhibits a considerably sharp increase with a relatively fast reversion back to the mean, whereas the outflow rate exhibits a relatively delayed U-shaped pattern with a considerably slower reversion to the mean.

As noted by Barnichon and Nekarda (2012) and Barnichon and Garda (2016), such distinct impulse responses underpin the steepness asymmetry displayed by the unemployment rate. As a result, after an initial shock to the inflow rate at the beginning of a recession, a FbF model is able to propagate the cyclical pattern of the flows and better describe the steepness asymmetry of the unemployment rate, considering that it forecasts the flow rate using a VAR model. On the other hand, a usual stock-based model should be expected to perform worse across recessionary periods, since it will be unable to capture the asymmetric pattern of the rate of unemployment.

4. A three-state forecasting model of the labour market

4.1 The labour market with three states

An important drawback of the two-state model of the labour market is the fact that it does not consider transitions into and from inactivity. These transitions are particularly important in the context of the Portuguese economy. In fact, we observe that gross flows into and out of the labour force are quantitatively larger compared with those into and out of the unemployment. Figure B.4 presents the average gross worker flows for transitions across the three labour market states, expressed as a total number of individuals in thousands, as a percentage of the labour force, and as a hazard rate. Indeed, we find that, on average across the sample period, gross unemployment inflows originating from inactivity surpass those originating from employment both in absolute terms (approximately, 64 thousand *versus* 56 thousand, respectively) and as a hazard rate (approximately, 3 % *versus* 1.6 %, respectively). A model of the Portuguese labour market which captures these transitions should thus be expected to better describe its dynamics and therefore perform better for forecasting purposes.

The present section outlines a model which considers transitions across all the three labour market states, as proposed by Barnichon and Nekarda (2012). As noted by the authors, a key advantage of such a model is the fact that it captures more accurately the labour market dynamics. Considering that all six flows should exhibit distinct time-series properties, a three-state model of the labour market may be able to improve unemployment forecasts relative to a two-state model. Furthermore, a three-state model allows for computing forecasts of the labour force participation rate.

In order to generalise the two-state model to three states, one must specify a system of differential equations describing the evolution of the stocks of unemployment, U, employment, E, and inactivity, N (Barnichon and Nekarda, 2012). Over months t and $t+1$, individuals may change from state $a \in \{E, U, N\}$ to state $b \in \{E, U, N\}$.

The transitions across states are governed by a Poisson process with a constant arrival rate given by λ^{ab}_{t+1}. The stocks of unemployment, employment, and inactivity thus satisfy the following system:

$$\begin{aligned}
\dot{U}_{t+\tau} &= \lambda^{EU}_{t+1}E_{t+\tau} + \lambda^{NU}_{t+1}N_{t+\tau} - (\lambda^{UE}_{t+1} + \lambda^{UN}_{t+1})U_{t+\tau} \\
\dot{E}_{t+\tau} &= \lambda^{UE}_{t+1}U_{t+\tau} + \lambda^{NE}_{t+1}N_{t+\tau} - (\lambda^{EU}_{t+1} + \lambda^{EN}_{t+1})E_{t+\tau} \\
\dot{N}_{t+\tau} &= \lambda^{EN}_{t+1}E_{t+\tau} + \lambda^{UN}_{t+1}U_{t+\tau} - (\lambda^{NE}_{t+1} + \lambda^{NU}_{t+1})N_{t+\tau}\,.
\end{aligned} \tag{4.1}$$

By applying the initial and terminal conditions, the forecasts of the three stocks for the next period may be solved as a function of the transition probabilities (see appendix C). These solutions can be applied to generate the next quarter forecasts of the unemployment rate and the labour force participation rate, as presented below:[1]

$$\widehat{u}_{t+1|t} = \frac{\widehat{U}_{t+1|t}}{\widehat{U}_{t+1|t} + \widehat{E}_{t+1|t}}\,, \tag{4.2}$$

and

$$\widehat{lfpr}_{t+1|t} = \frac{\widehat{U}_{t+1|t} + \widehat{E}_{t+1|t}}{\widehat{E}_{t+1|t} + \widehat{U}_{t+1|t} + \widehat{N}_{t+1|t}}\,. \tag{4.3}$$

Similar to the two-state model, a VAR is used to produce forecasts of the transition rates, which are then applied to construct forecasts beyond one quarter. In particular, the following VAR with n lags is estimated:

$$\boldsymbol{y}_t = \boldsymbol{c} + \boldsymbol{\phi}_1\boldsymbol{y}_{t-1} + \boldsymbol{\phi}_2\boldsymbol{y}_{t-2} + \cdots + \boldsymbol{\phi}_n\boldsymbol{y}_{t-n} + \boldsymbol{\epsilon}_t\,, \tag{4.4}$$

where:

$$\boldsymbol{y}_t = log(\lambda^{EU}_t, \lambda^{UE}_t, \lambda^{EN}_t, \lambda^{NE}_t, \lambda^{NU}_t, \lambda^{UN}_t, u_t, ...)'\,. \tag{4.5}$$

As before, several specifications are tested and we choose the specification with the best average forecasting performance.

[1] As noted by Barnichon and Nekarda (2012), since we forecast population shares, we implicitly assume that population growth has no impact on the forecasts.

4.2 Data

In the three-state forecasting model of the Portuguese labour market, we use aggregate transition rates across the three labour market statuses: employment, unemployment, and inactivity. The aggregate labour market transition rates are computed using the Portuguese Labour Force Survey microdata.[2]

The transition rates are constructed from labour market gross worker flows as $\lambda_t^{ab} = F_t^{ab}/(\sum_h F_t^{ah}), a, b, h \in \{E, U, N\}$, where F_t^{ab} denotes the number of persons moving from status a to status b in quarter t, and $\sum_h F_t^{ah}$ denotes the total number of persons moving out of status a in quarter t. We use LFS microdata from 1998:1 to 2018:1, covering two LFS series (from 1998:1 to 2010:4 and from 2011:1 to 2018:1). The Portuguese LFS has been subject to a methodological redesign in 2011:1, with a considerable impact in the estimated aggregate transition rates. Thus, for forecasting purposes, we interpolate backward the second series based on the variations of the first series. The series of the transition rates used in the estimations are plotted in Figure B.5 in appendix.

4.3 Discussion of results

On average, across the sample period under consideration, we find that the three-state labour market forecasting model for Portugal (FbF-3) performs worse relative to the simpler the second-state forecasting model (FbF-2) (see Table 4.1). By restricting the sample to the period previous to the methodological redesign (*i.e.*, from 2008:1 to 2010:4), we find that the results obtained by the FbF-3 improve considerably on the FbF-2. Conversely, the results obtained by the FbF-3 for the period following the methodological redesign are much less accurate relative to the other alternative forecasting models, as evaluated by their respective RMSE's. This finding most likely reflects the series break for the labour market transition rates.

As argued by Barnichon and Nekarda (2012), the two proposed flow-based mod-

[2]See appendix D for a brief description of the Portuguese LFS and an extensive overview of the method used for the computation of the transition rates.

els present advantages and disadvantages in relation to each other. On the one hand, the FbF-2 is rather parsimonious and easier to implement relative to the FbF-3. Furthermore, even though the duration-based unemployment flow rates are not directly measured, but instead inferred via a theoretical model, these are considerably less noisy in comparison to the aggregate transition rates computed using the the Portuguese LFS.

On the other hand, the FbF-3 is a more realistic representation of the Portuguese labour market, considering the non-negligible impact of the gross worker flows into and out of the labour force for unemployment variations. The FbF-3 is capable of producing internally consistent forecasts for several key labour market indicators, namely, the unemployment rate, the labour force participation rate, and the employment-population ratio. Although, on average, the FbF-3 performs worse for unemployment rate forecasts relative to the alternative models, including the simpler FbF-2, these results most likely reflect the series break for the aggregate transition rates estimated using the Portuguese LFS, following the methodological redesign in 2011:1.

Table 4.1: Absolute RMSE of FbF-2, FbF-3, and alternative forecasting models

Forecasting model	t+0	t+1	t+2	t+3	t+4	t+8	Average(RMSE)
FbF-2	0.59	0.93	1.11	1.43	1.95	4.05	1.67
FbF-3	0.98	1.22	1.45	1.81	2.31	4.23	2.00
FbF-3 (2008:1-2010:4)	0.63	0.87	1.00	1.03	1.11	2.03	1.11
FbF-3 (2011:1-2018:1)	1.09	1.35	1.60	2.07	2.70	5.08	2.32
VAR	0.68	1.09	1.43	1.83	2.40	4.37	1.97
VAR no flows	0.67	1.07	1.43	1.85	2.42	4.39	1.97
ARIMA	0.65	1.079	1.44	1.75	2.33	4.48	1.96

Source: Authors' calculations;
Notes: The results concern the best average performing model specification.

5. Conclusion

This paper applies the "flow approach" to forecasting the rate of unemployment, initially developed by Barnichon and Nekarda (2012) and subsequently extended by Barnichon and Garda (2016), to the Portuguese labour market. We start by implementing a two-state labour market forecasting model and then extend it to a three-state labour market forecasting model which incorporates movements in and out of the labour force.

We find that the two-state flow-based forecasting model performs slightly better than the other tested models. Therefore, worker flow data is a valuable input for forecasting the unemployment in Portugal. We illustrate the advantages of a flow-based forecasting model *versus* a stock-based forecasting model and conclude that the flow-based forecasting model is able to propagate the cyclical pattern of the flows and better describe the steepness-asymmetry exhibited by the unemployment rate, which underpins its superior forecasting performance. As noted by Barnichon and Nekarda (2012) and Barnichon and Garda (2016), a crucial advantage of the two-state flow-based forecasting model is its modest data requirements, which means that it may easily be applied to and comparable amongst different countries.

We have extended the two-state flow-based forecasting model to include transitions into and out of the labour force, a striking feature of the Portuguese labour market. The three-state flow-based model outperforms the simpler two-state model in forecasting the unemployment rate only if we restrict the sample period to 2008:1-2010:4. This result reflects the methodological break affecting the Portuguese Labour Force Survey in 2011:1, rendering the two LFS series incompatible (see INE, 2015). Further refinements to the three-state forecasting model are expected to lead to forecasting improvements, notwithstanding the data issues surrounding its implementation.

Bibliography

Aiken, M. (1996), 'A neural network to predict civilian unemployment rates', *Journal of International Information Management* .

Barnichon, R. (2012), 'Vacancy posting, job separation and unemployment fluctuations', *Journal of Economic Dynamics and Control* **36**(3), 315–330.

Barnichon, R. and Garda, P. (2016), 'Forecasting unemployment across countries: The ins and outs', *European Economic Review* **84**(C), 165–183.

Barnichon, R. and Nekarda, C. J. (2012), 'The Ins and Outs of Forecasting Unemployment: Using Labor Force Flows to Forecast the Labor Market', *Brookings Papers on Economic Activity* **43**(2 (Fall)), 83–131.

Blanchard, O. and Portugal, P. (2001), 'What hides behind an unemployment rate: Comparing portuguese and u.s. labor markets', *American Economic Review* **91**(1).

Bleakley, H., Ferris, A. E. and Fuhrer, J. C. (1999), 'New data on worker flows during business cycles', *New England Economic Review* (Jul), 49–76.

Brown, L. and Moshiri, S. (2004), 'Unemployment variation over the business cycles: a comparison of forecasting models', *Journal of Forecasting* **23**(7), 497–511.

Bry, G. and Boschan, C. (1971), *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*, National Bureau of Economic Research, Inc.

Chua, C. L., Lim, G. C. and Tsiaplias, S. (2012), 'A latent variable approach to forecasting the unemployment rate', *Journal of Forecasting* **31**(3), 229–244.

Clements, M. P. and Krolzig, H.-M. (2003), 'Business Cycle Asymmetries: Characterization and Testing Based on Markov-Switching Autoregressions', *Journal of Business & Economic Statistics* **21**(1), 196–211.

Diebold, F. X. and Mariano, R. S. (1995), 'Comparing predictive accuracy', *Journal of Business & Economic Statistics* **13**(3), 253–263.

Ellul, R. (2018), Forecasting unemployment rates in Malta: A labour market flows approach, CBM Working Papers WP/03/2018, Central Bank of Malta.

Elsby, M. W. L., Hobijn, B. and ahin, A. (2013), 'Unemployment Dynamics in the OECD', *The Review of Economics and Statistics* **95**(2), 530–548.

Fujita, S. and Ramey, G. (2009), 'The Cyclicality Of Separation And Job Finding Rates', *International Economic Review* **50**(2), 415–430.

Giacomini, R. and White, H. (2006), 'Tests of conditional predictive ability', *Econometrica* **74**(6), 1545–1578.

Golan, A. and Perloff, J. M. (2004), 'Superior Forecasts of the U.S. Unemployment Rate Using a Nonparametric Method', *The Review of Economics and Statistics* **86**(1), 433–438.

Haykin, S. (1994), *Neural Networks: A Comprehensive Foundation*, 1st edn, Prentice Hall PTR, Upper Saddle River, NJ, USA.

ILO (2015), 'World employment and social outlook: Trends'.

INE (2015), 'Instituto nacional de estatística, documento metodológico do inquérito ao emprego versão 3.1'. [Online; accessed 15/05/2019].
URL: *http://smi.ine.pt/UploadFile/Download/1964*

Johnes, G. (1999), 'Forecasting unemployment', *Applied Economics Letters* **6**(9), 605–607.

Koop, G. and Potter, S. M. (1999), 'Dynamic Asymmetries in U.S. Unemployment', *Journal of Business & Economic Statistics* **17**(3), 298–312.

Martins, F. and Seward, D. (2019), Into the heterogeneities in the portuguese labour market: an empirical assessment, Banco de Portugal Working Papers 8:2019.

Meyer, B. and Tasci, M. (2015), Lessons for forecasting unemployment in the United States: use flow rates, mind the trend, FRB Atlanta Working Paper 2015-1, Federal Reserve Bank of Atlanta.

Milas, C. and Rothman, P. (2008), 'Out-of-sample forecasting of unemployment rates with pooled STVECM forecasts', *International Journal of Forecasting* **24**(1), 101 – 121.

Montgomery, A. L., Zarnowitz, V., Tsay, R. S. and Tiao, G. C. (1998), 'Forecasting the u.s. unemployment rate', *Journal of the American Statistical Association* **93**(442), 478–493.

Neves, D. (2014), Labour market flows after the redesign of the portuguese labour force survey, Master's thesis, ISEG, Universidade de Lisboa.

Olmedo, E. (2014), 'Forecasting Spanish Unemployment Using Near Neighbour and Neural Net Techniques', *Computational Economics* **43**(2), 183–197.

Petrongolo, B. and Pissarides, C. A. (2008), 'The Ins and Outs of European Unemployment', *American Economic Review* **98**(2), 256–262.

Proietti, T. (2003), 'Forecasting the US unemployment rate', *Computational Statistics & Data Analysis* **42**(3), 451 – 476.

Rothman, P. (1998), 'Forecasting Asymmetric Unemployment Rates', *The Review of Economics and Statistics* **80**(1), 164–168.

Sengul, G. and Tasci, M. (2014), Unemployment Flows, Participation, and the Natural Rate for Turkey, Working Papers (Old Series) 1422, Federal Reserve Bank of Cleveland.

Shimer, R. (2003), The Cyclical Behavior of Equilibrium Unemployment and Vacancies: Evidence and Theory, NBER Working Papers 9536, National Bureau of Economic Research, Inc.

Shimer, R. (2012), 'Reassessing the ins and outs of unemployment', *Review of Economic Dynamics* **15**(2), 127 – 148.

Solon, G., Michaels, R. and Elsby, M. W. L. (2009), 'The Ins and Outs of Cyclical Unemployment', *American Economic Journal: Macroeconomics* **1**(1), 84–110.

Stasinakis, C., Sermpinis, G., Theofilatos, K. and Karathanasopoulos, A. (2016), 'Forecasting US Unemployment with Radial Basis Neural Networks, Kalman Filters and Support Vector Regressions', *Computational Economics* **47**(4), 569–587.

Tasci, M. (2012), The ins and outs of unemployment in the long run: unemployment flows and the natural rate, Working Papers (Old Series) 1224, Federal Reserve Bank of Cleveland.

Zailstra, A. and Boxhoorn, K. (2017), Labour supply and the business cycle: Lessons from labour market flows and international forecating practices, Technical report, Netherlands Bureau for Economic Policy Analysis, CPB.

Appendices

A. Tables

Table A.1: Cross-correlation between the unemployment rate and the steady-state unemployment rate

i	$Corr(u, SSUR(-i))$		
	$d = 24$	$d = 12$	$d = 6$
0	0.796	0.824	0.866
1	0.834	0.859	0.895
2	0.863	0.883	0.912
3	0.884	0.900	0.917
4	0.886	0.896	0.906

Source: Authors' calculations;
Note: Cross-correlations between the unemployment rate at quarter t (u_t) and the steady-state unemployment rate at quarter $t - i$ ($SSUR(-i), i \in \{0, 1, 2, 3, 4\}$), using quarterly data over 1999:1 to 2018:3. For robustness, the computations are performed for different values of the parameter $d \in \{6, 12, 24\}$.

Table A.2: Description of the data sources

Variable	Source	Units	Abbreviation
Unemployment rate	Statistics Portugal (INE)	% of labour force	u
Unemployment	Statistics Portugal (INE)	thousands of persons	U
Unemployment by duration	Statistics Portugal (INE)	thousands of persons	$U^{<d}$
Employment	Statistics Portugal (INE)	thousands of persons	E
Inactivity	Statistics Portugal (INE)	thousands of persons	N
Real Gross Domestic Product	Statistics Portugal (INE)	millions of EUR (base year=2011)	GDP
Vacancies	OECD Statistics	number of persons	v
Gross Valued Added by the Construction Sector	Statistics Portugal (INE)	millions of EUR	$vabconstruct$
Job finding rate	Own calculations	% of unemployment	f
Separation rate	Own calculations	% of employment	s
Transition rate from state a to b	Own calculation based on the LFS	% of state a	λ^{ab}

Notes: All the data is measured at a quarterly frequency.

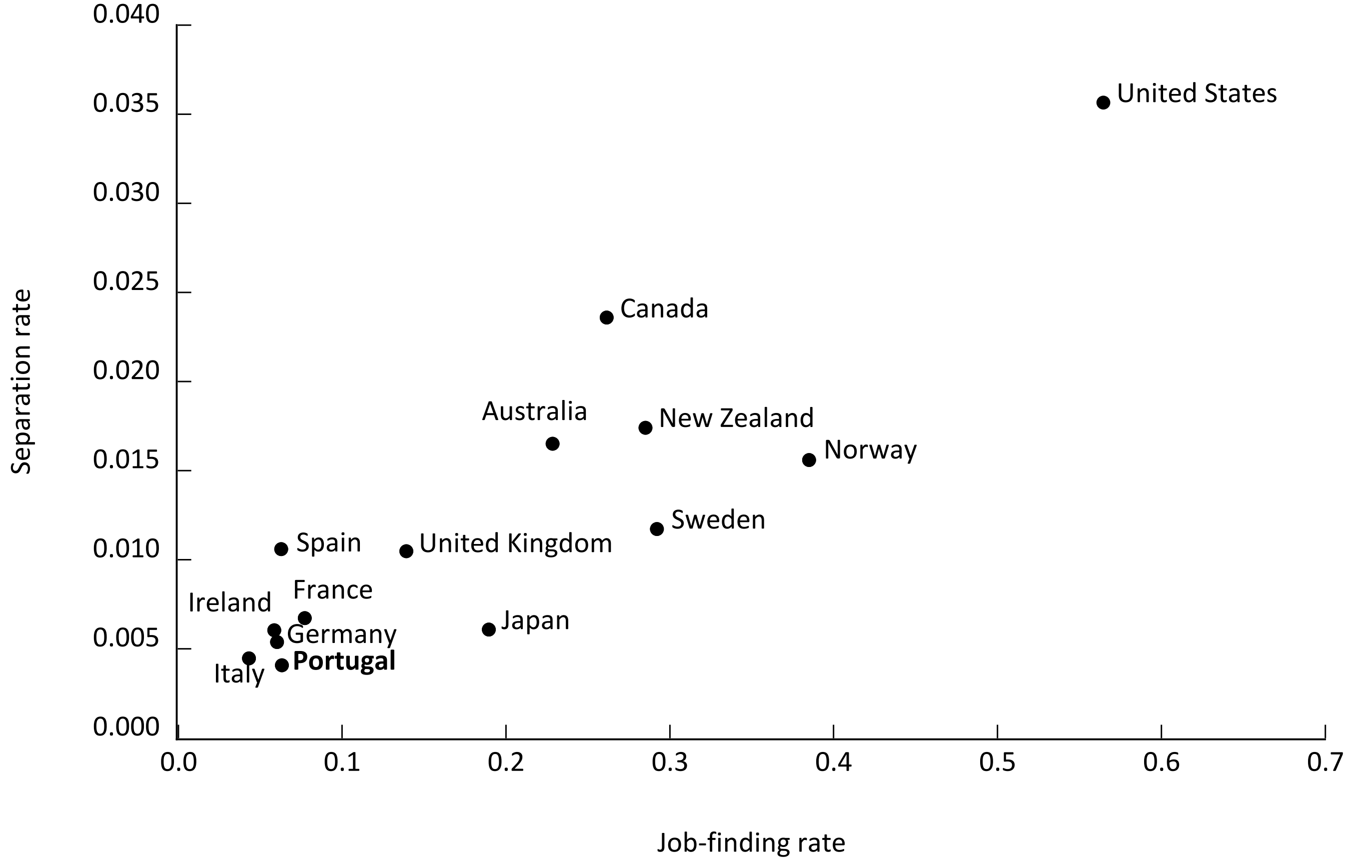
Table A.3: Time-series properties of the flow rates

	$d = 24$		$d = 12$		$d = 6$	
	f_t	$s_t/E(u_t)$	f_t	$s_t/E(u_t)$	f_t	$s_t/E(u_t)$
$E(X)$	0.050	0.053	0.055	0.057	0.071	0.073
$sd(X)$	0.010	0.023	0.014	0.019	0.027	0.023
$\rho(X)$	0.932	0.810	0.911	0.740	0.894	0.842

Source: Authors' calculations;
Note: Quarterly data over 1999:1 to 2018:3.

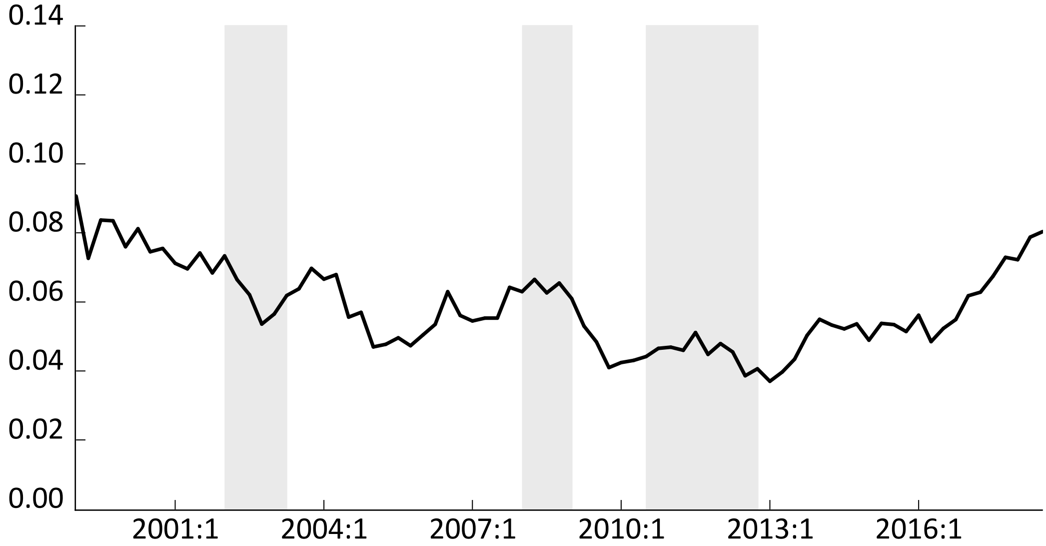
B. Figures

Figure B.1: Job-finding and separation rates for several OECD countries



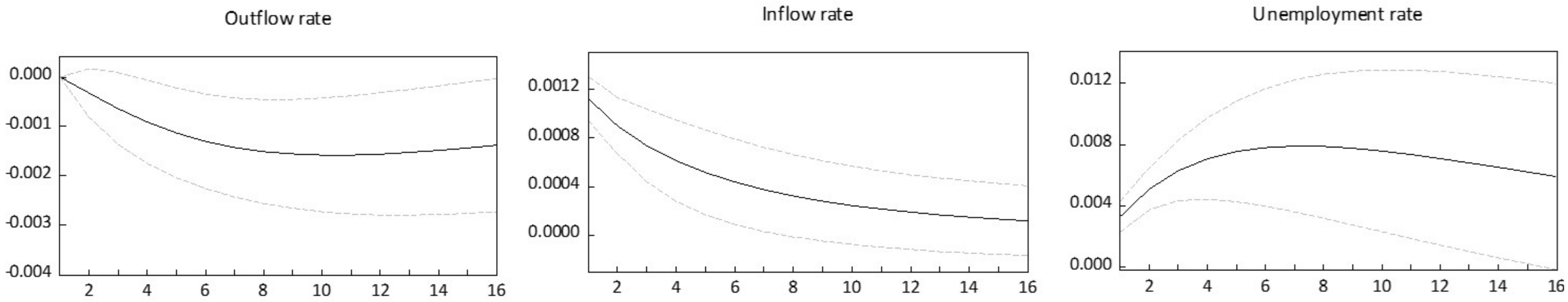
Source: Authors' calculations based on Elsby et al. (2013) and Barnichon and Garda (2016);
Notes: The results concern the averages of monthly data for the job-finding and separation rates, computed following the methodology by Elsby et al. (2013). The time-period depends for each country. For Portugal, the starting year and the ending year are 1986 and 2009, respectively.

Figure B.2: The convergence rate to steady-state unemployment, β_t



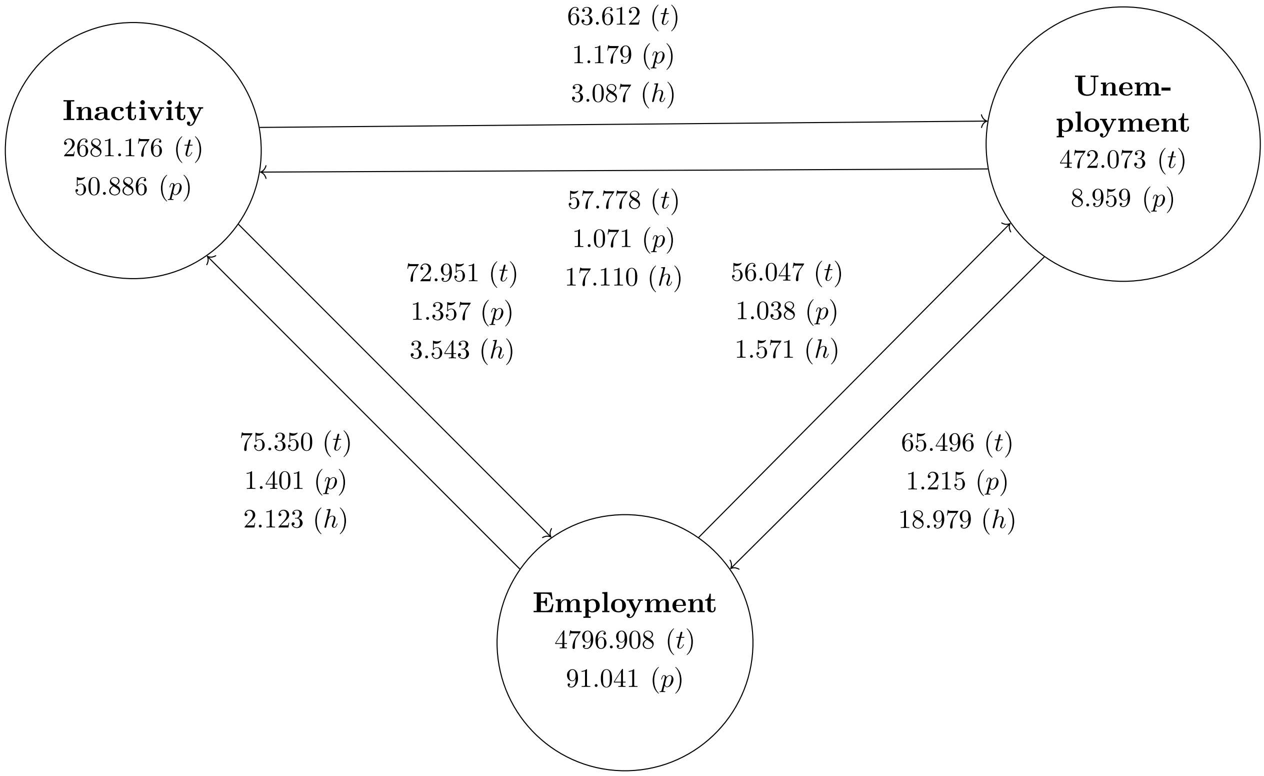
Source: Authors' calculations based on data by INE;
Notes: The convergence rate to the steady-state unemployment rate is computed as $\beta_t = 1 - exp[-(s_t + f_t)]$, using quarterly data from 1999:1 to 2018:3. The shadings signal recessions according to the business-cycle dating methodology by Bry and Boschan (1971).

FIGURE B.3: Impulse response functions to a one standard deviation shock to the unemployment inflow rate



Source: Authors' calculations;
Notes: The impulse response functions are computed to a 1-standard-deviation shock to the unemployment inflow rate, from the estimation of a VAR given by $\boldsymbol{y}_t = log(f_t, s_t, u_t)'$, with one lag. The VAR is estimated on the quarterly data across the whole sample period. Dashed lines represent $\pm$ 2 standard-error confidence intervals.

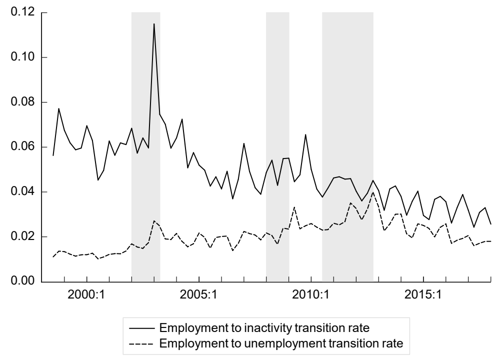
Figure B.4: Average quarterly worker flows by labour market status, Labour Force Survey, 1998:1 to 2018:1



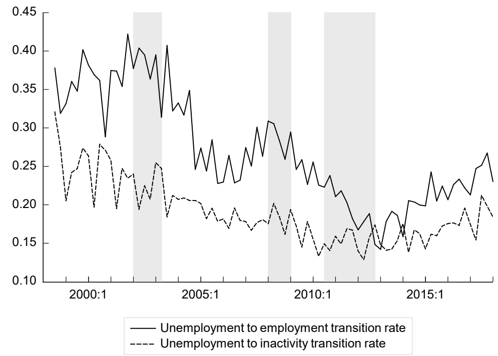
Source: Authors' calculations based on the LFS;
Notes: The worker gross flows are expressed as total number of individuals in thousands (t), as a percentage of the labour force (p), and as a hazard rate (h). The statistics are the quarterly averages of the period from 1998:1 to 2018:1.

FIGURE B.5: Extrapolated transition rates by labour market status of origin, 1998:2-2018:1

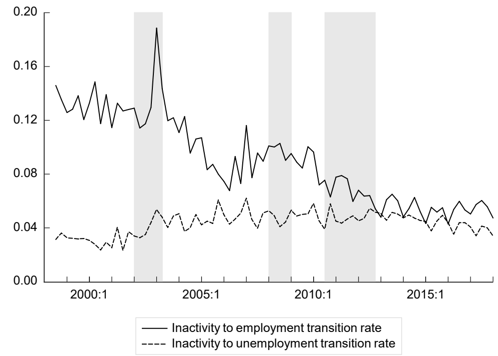
(a) Transition rates out of employment



(b) Transition rates out of unemployment



(c) Transition rates out of inactivity



Source: Authors' calculations based on the LFS;
Notes: The second LFS series (from 2011:1) is extrapolated backwards based on the variations of the transition rates in the first series (from 1998:2 to 2010:4). The shadings signal recessions according to the business-cycle dating methodology by Bry and Boschan (1971).

C. Solution to the three-state model

In this section, we present the analytical solution for equation (4.1), as described by Barnichon and Nekarda (2012).

Let $\boldsymbol{Y}_{t+\tau} = (U_{t+\tau}, E_{t+\tau}, N_{t+\tau})'$. Hence, equation (4.1) may be expressed as:

$$\dot{\boldsymbol{Y}}_{t+\tau} = \boldsymbol{A}_t \boldsymbol{Y}_{t+\tau} \, , \tag{C.1}$$

where

$$\boldsymbol{A}_t = \begin{bmatrix} -\lambda_t^{UE} - \lambda_t^{UN} & \lambda_t^{EU} & \lambda_t^{NU} \\ \lambda_t^{UE} & -\lambda_t^{EU} - \lambda_t^{EN} & \lambda_t^{NE} \\ \lambda_t^{UN} & \lambda_t^{EN} & -\lambda_t^{NE} - \lambda_t^{NU} \end{bmatrix} . \tag{C.2}$$

$\boldsymbol{A}_t$ has one eigenvalue which equals zero, since its columns must sum to zero. Let $\boldsymbol{Q}_t$ be the matrix of eigenvectors of matrix $\boldsymbol{A}_t$ associated to the eigenvalues $[r_{1t}, r_{2t}, 0]$, one may express the solution to equation (C.1) as:

$$\boldsymbol{Y}_{t+\tau} = \boldsymbol{Q}_t \begin{bmatrix} c_1 \times exp\{\tau(r_{1t})\} \\ c_2 \times exp\{\tau(r_{2t})\} \\ c_3 \end{bmatrix} , \tag{C.3}$$

where c_1, c_2, and c_3 are the usual constants of integration. As noted by Barnichon and Nekarda (2012), the two non-zero eigenvalues are negative and may be expressed as a function of the hazard rates, namely:

$$\begin{aligned} r_{1t} \approx -\beta_{1t} \equiv \lambda_t^{UE} + \lambda_t^{UN} \\ r_{2t} \approx -\beta_{2t} \equiv \lambda_t^{EU} + \lambda_t^{EN} + \lambda_t^{NE} + \lambda_t^{NU} \, . \end{aligned} \tag{C.4}$$

In order to obtain the values represented by c_1, c_2, and c_3, one may use the initial conditions $\boldsymbol{Y}_t = (U_t, E_t, N_t)'$ and the terminal conditions $\boldsymbol{Y}_t \rightarrow (U_t^*, E_t^*, N_t^*)'$ as $t \rightarrow \infty$ and the vector of the steady state numbers of the stocks, containing U_t^*, E_t^*, and N_t^*.

Hence, the steady-state stocks are provided by:

$$\begin{aligned}
U_t^* &= k\frac{s_{t+1}}{s_{t+1}+f_{t+1}+o_{t+1}} \\
E_t^* &= k\frac{f_{t+1}}{s_{t+1}+f_{t+1}+o_{t+1}} \\
N_t^* &= k\frac{o_{t+1}}{s_{t+1}+f_{t+1}+o_{t+1}} \,,
\end{aligned} \tag{C.5}$$

with k representing a constant which we set such that the steady-state stocks U_t^*, E_t^*, and N_t^* sum to the total working-age population[1] in quarter t, P_t, and s_{t+1}, f_{t+1}, and o_{t+1} are given by:

$$\begin{aligned}
s_{t+1} &= \lambda_{t+1}^{EN} \times \lambda_{t+1}^{NU} + \lambda_{t+1}^{NE} \times \lambda_{t+1}^{EU} + \lambda_{t+1}^{NU} \times \lambda_{t+1}^{EU} \\
f_{t+1} &= \lambda_{t+1}^{UN} \times \lambda_{t+1}^{NE} + \lambda_{t+1}^{NU} \times \lambda_{t+1}^{UE} + \lambda_{t+1}^{NE} \times \lambda_{t+1}^{UE} \\
o_{t+1} &= \lambda_{t+1}^{EU} \times \lambda_{t+1}^{UN} + \lambda_{t+1}^{UE} \times \lambda_{t+1}^{EN} + \lambda_{t+1}^{UN} \times \lambda_{t+1}^{EN} \,.
\end{aligned} \tag{C.6}$$

The one-quarter ahead forecasts for the stocks of unemployment, employment, and inactivity may then be obtained as:

$$\begin{aligned}
U_{t+1} &= q_{11t} \times c_1 \times exp\{-\beta_{1t}\} + q_{12t} \times c_2 \times exp\{-\beta_{2t}\} + U_t^* \\
E_{t+1} &= q_{21t} \times c_1 \times exp\{-\beta_{1t}\} + q_{22t} \times c_2 \times exp\{-\beta_{2t}\} + E_t^* \\
N_{t+1} &= q_{31t} \times c_1 \times exp\{-\beta_{1t}\} + q_{32t} \times c_2 \times exp\{-\beta_{2t}\} + N_t^* \,,
\end{aligned} \tag{C.7}$$

wherein q_{ijt} is the element (i, j) of matrix $\boldsymbol{Q}_t$ and the values c_1 and c_2 are obtained as:

$$\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}^{-1} \times \begin{bmatrix} U_t \\ E_t \end{bmatrix} \,. \tag{C.8}$$

[1] Defined as the total population aged between 15 and 74.

D. Computing labour market transition rates

In this section, we describe the method used to compute the gross worker flows and labour market transition rates applied in the three-state forecasting model of the Portuguese labour market (see section 4).[1]

The aggregate labour market transition rates are computed using the Portuguese Labour Force Survey microdata. The Portuguese LFS is a household survey performed by Statistics Portugal (INE) at a quarterly basis. Its main goal is to characterise the Portuguese labour market. The LFS complies to the guidelines specified by Eurostat.

The LFS gathers personal data on various characteristics of the interviewees, which enable the estimation of quarterly stocks for employment, unemployment, and inactivity. This information is then used for the computation of several labour market indicators.

Methodologically, the INE surveys around 20,000 households each quarter. The population which the aim of the LFS is the group of residents in the national territory, whereas the sample unit taken into account in the statistical inference procedure is the household as a main residence. The household units contained in the probabilistic sample are drawn through a multistage stratified sampling process, which ensures that every individual is representative of a subgroup of the total population. Therefore, every individual is linked to a statistical weight used for performing the inference to the population as a whole.

The sample used in the LFS is a rotating sample. It consists of six rotations which conform to a rotation scheme: every quarter 5/6 of the total sample are retained, while the remaining 1/6 are replaced, *i.e.* once a given household is selected it should be interviewed for six quarters consecutively. Since one may assess the labour market status of 5/6 of the interviewees in the sample across adjacent quarters, we were able to compute the gross worker flows and labour market transition rates used in the three-state labour market forecasting model. The INE provided us access to the LFS microdata from 1998:1 to 2018:1.

[1]The method follows closely Blanchard and Portugal (2001), Neves (2014), and Martins and Seward (2019).

The rotating sample of the Portuguese LFS (each quarter 5/6 of the sample is retained, whilst the remaining 1/6 is replaced) implies that the LFS respondents should be interviewed for six consecutive quarters (INE, 2015). In this paper, we apply a backward matching technique, whereby quarters 2, 4, and 6 are matched in a backward fashion with the corresponding counterparts.[2]

In order to assess the labour market status of each LFS respondent over the six quarters they are kept in the sample: (*i*) from 2011:1, we make use of the LFS individual's unique identifier[3], and (*ii*) from 1998:1 to 2010:4, we make use of a set of identifying variables, including the accommodation, the location, the household, and an individual within the household identifiers. Consistent with previous studies which use the Portuguese LFS, we also validate the matched labour market individual transitions at each quarter based on the respondent's reported age and gender.

By applying the statistical weights associated with each respondent and assessing the transitions across quarters, we are able to report the respective gross worker flows. In order to compute the labour market transition rates, we follow Shimer (2012). Therefore, the labour market transition rates from status a to status b at quarter t are calculated as the ratio of the respective flows over the sum of outflows from status a:

$$\lambda_t^{ab} = \frac{F_t^{ab}}{\sum_h F_t^{ah}}, a, b, h \in \{E, U, N\} \tag{D.1}$$

The Portuguese LFS presents us with two main data issues: (*i*) a scramble of the household identifier performed by INE from 1998:4 to 1999:1, which implies the existence of a missing value, and (*ii*) a survey methodological redesign performed in 2011:1, which renders the two LFS series not comparable. In both cases we employ an imputation method based on a moving-average. Moreover, for forecasting purposes, and considering the substantial differences in magnitudes for some of the transition rates following the LFS redesign, we extrapolate backwards the second

[2]One should keep in mind that the so-called backward and forward matching procedures are not merely conventions, considering that the computed transition rates will depend on the adopted method. See Bleakley et al. (1999) for a comprehensive discussion on this issue.

[3]In 2011:1, the INE completely revised the LFS and introduced an individual's unique identifier, which renders the matching procedure simpler.

LFS series (from 2011:1) based on the variations of the transition rates in the first series (from 1998:2 to 2010:4).