ACCESSING THE IMPACT OF COVID-19 ON THE PORTUGUESE UNEMPLOYMENT RATE

Diogo Queiroz Miguel

Dissertation report presented as partial requirement for obtaining the Master's degree in Information Management with specialization in Knowledge Management and Business Intelligence

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

BOOK SPINE

2021

Accessing the impact of Covid-19 on the Portuguese unemployment rate

Diogo Queiroz Miguel

MGI

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

ACCESSING THE IMPACT OF COVID-19 ON THE PORTUGUESE UNEMPLOYMENT RATE

by

Diogo Queiroz Miguel

Dissertation report presented as partial requirement for obtaining the Master’s degree in Information Management, with a specialization in Knowledge Management and Business Intelligence

Advisor: Pedro Coelho

Co-Advisor*:* Bruno Damásio

FEB 2023

ACKNOWLEDGEMENTS

I would like to thank professors Pedro Coelho and Bruno Damásio for all the help provided in making this paper. I would also like to thank my parents for all the support and motivation provided during this year. Finally, a huge thank you to my friend Daniela Marques, who has been a pillar of support and whose help during this period I could never repay.

ABSTRACT

We analyze the possibility of Vector Autoregressive models being good estimators for the unemployment rate in Portugal, by studying their ability to understand the impact the COVID-19 pandemic had on the unemployment rate. We make use of Bayesen Stochastic Search Variable Selection and bootstrapping techniques for forecasting, comparing the results of these models with two benchmark techniques, ARIMA and Artificial Neural Networks. The model performance is tested through the RMSE, MSE and MAE of the estimations, and we compare the forecasting quality through a Diebold-Mariano test. We conclude that the VAR methodology can provide better forecasts than the benchmark models when combined with the Bayesian approach, both for shorter and longer forecasting horizons. We also conclude that COVID-19 did not provide the expected shock to the Portuguese unemployment rate.

KEYWORDS

COVID-19; unemployment; vector autoregressive models; forecasting.

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ARIMA	Autoregressive Integrated Moving Average
CPI	Consumer Price Indices
INE	Statistics Portugal
IEFP	Instituto do Emprego e Formação Profissional
VAR	Vector Autoregressive Models

1. INTRODUCTION

The impact of COVID-19 has been strongly felt over the last two years. Since the first detected case on the 2nd of March 2020, the country has been through two lockdowns that closed schools and companies, restricted access to public spaces and changed entire work and home routines, which, while controlling the progression of the virus, helped increase economic repercussions (Vieira & Meirinhos, 2021) by disrupting production and supply market chains and causing several public health emergencies (Nigam et al., 2022).

The unemployment rate also increased exponentially during this period, having only started to decrease recently due to the increase in vaccinations and the end of restrictions (INE, 2021). Not only that, but many other macroeconomic measures also suffered drastic shifts due to the pandemic, such as exports, inflation and household consumption, for example. There is still a lot of uncertainty regarding the full impact of the pandemic on the economy in Portugal and unemployment and therefore, it is required that we can quantify said impact.

1.1. BACKGROUND AND PROBLEM IDENTIFICATION

There have been a variety of studies depicting the impact of the pandemic on unemployment. Barro et al. (2020) estimated upper limits for the infection outcome of the pandemic, by comparing it to the Great Influenza Pandemic of 1918, showing possible worst-case scenarios for the impact of COVID-19. Other papers aim at more specific impacts such as Şahin et al. (2020) which used historical flow rates of unemployment to predict the shift of this indicator in the U.S or Lai et al. (2021) who used a hybrid approach of prediction by combining an ARIMA model with several non-linear advanced techniques, such as Artificial Neural Networks (ANN) and support vector machines, to predict the unemployment rate in several Asian countries.

However, it can be proven relevant to access how multiple variables affected each other during this period, to predict unemployment. Trying to understand the relation between several macroeconomic variables and producing a singular model based on it is not a novel idea. Stock & Watson (2002) have previously provided studies on this topic. However, their research was made much before the pandemic and therefore does not include a crucial element we intend to study. Understanding these relationships can provide a greater knowledge of how unemployment might change in the future, given the state of the pandemic.

According to (Zivot & Wang, 2006), VAR models are one of the best and easiest models to use for the analysis of multivariate time series, providing better forecasts compared to univariate time series models. It can be relevant to understand if this approach is indeed more suited to evaluate the evolution of unemployment in Portugal and to confirm how the economy shifted given the state of the pandemic.

Therefore, our study aims to construct several VAR models that analyze the underlying relationships between COVID-19, unemployment and several other macroeconomic data from Portugal, to describe the future evolution of unemployment. We intend to study if a VAR methodology is indeed more suited to forecast unemployment, compared to other forecasting techniques, such as

ARIMA and Artificial Neural Networks, and calculate the impact the COVID-19 pandemic had on the unemployment rate.

1.2. STUDY OBJECTIVES

The main objectives of this study include the identification of suitable variables to predict unemployment, variables that show the progression of the pandemic and the construction of adequate models to forecast unemployment. Accordingly, we will apply the following research objectives:

- **(RO1).** Identify relevant macroeconomic variables that can explain the unemployment rate in Portugal.
- **(RO2).** Find relevant dimensions that explain the progression of the pandemic in Portugal.
- **(RO3).** Develop models that predict the unemployment rate in Portugal.
- **(RO4).** Understand how the pandemic affected the unemployment rate and impacted other macroeconomic variables.
- **(RO5).** Validate the quality of the models produced.

2. LITERATURE REVIEW

In this section, we present the main literature evaluated to define the methodology used for the construction of models that forecast unemployment. The main tool used to find these studies was Google Scholar. The main keywords used were: unemployment: VAR; ARIMA; Neural Network.

2.1. WHY PREDICT UNEMPLOYMENT?

It has been shown that the unemployment rate of a country is an important tool that can provide good insights into the development of a country and allows for accurate decision-making and the design of policymaking (see Claveria, 2019; Giles et al., 2005; Katris, 2020). Usualy expansionary periods are associated with low unemployment levels, while recessions bring higher unemployment rates and lower expected wages (Martins & Damásio, 2020). The world's economy was put on a halt for most of 2020 and 2021 which created a completely unexpected and tragic recession, which led to spillovers into many other major sectors of the economy, affecting financial markets, travel industries, the health sector, among others (Ozili, 2020). The impact of these spillovers resulted in a large increase in job losses across the world (Blustein et al., 2020), including in Portugal. The European Commission estimated that almost 11% of employment in Portugal is in sectors that were forced to shut down almost entirely due to the COVID-19 outbreak (Duarte, 2020), after the intense lockdown restrictions implemented by governments.

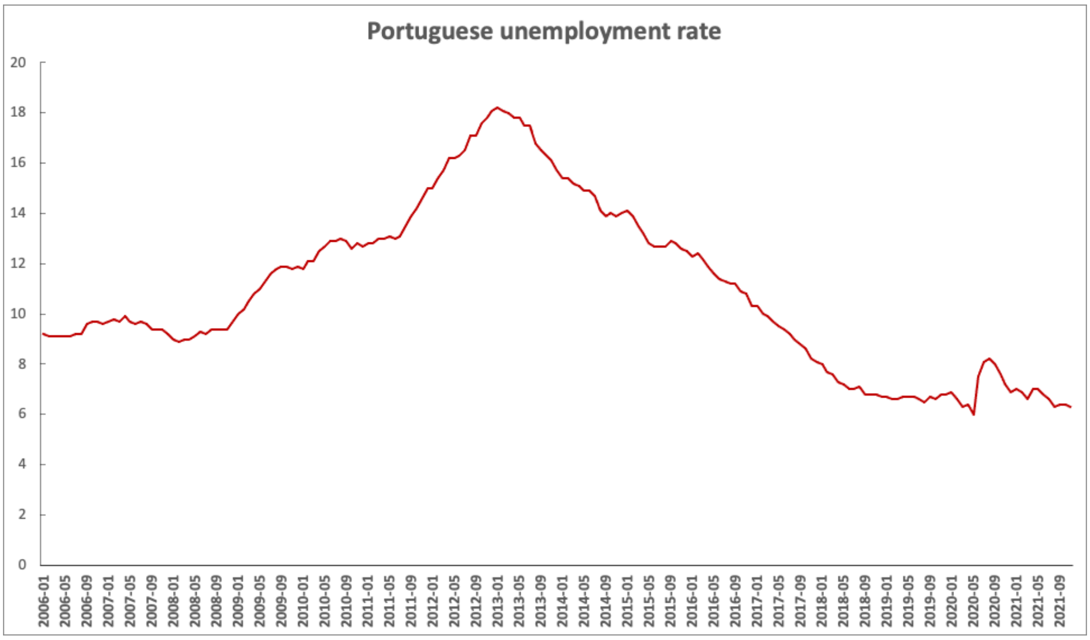


Fig 1. Portuguese unemployment rate.
Note: The unemployment rate is measured as a percentage of the Portuguese labour force.

2.2. MODEL SELECTION

The selection of VAR models for this analysis is also based on literature. As seen above, VAR models are incredibly suitable for the forecast of multivariate time series since they can interpret the underlying relationships between variables (Sims, 1980). As seen in the literature, VAR models can extract incredibly powerful insights from economic information (see Balcerzak & Zurek, 2011; Darko, 2018; Dolado & Jimeno, 1997; Naccarato et al., 2018).

It has been shown that they provide several improvements over ARIMA models, which can only forecast univariate time series (Thomakos & Guerard, 2004). However, they provide some shortcomings when compared to machine learning models, such as Neural Networks. (Mulaudzi & Ajoodha, 2020) shows that Neural Networks are more suitable to deal with nonlinear data. VAR model forecasts tend to worsen as the predictive horizon increases, something that Neural Network predictors can overcome (Marcellino et al., 2006; Olmedo, 2014). We intend to compare these approaches to confirm if the VAR model is indeed more suitable for the forecast of Portuguese unemployment.

2.3. UNDERSTANDING VARIABLE RELEVANCE

To better select suitable indicators that could help forecast unemployment, we analysed multiple studies that provide relevance to the selection.

Inflation has also been affected by the pandemic. Reinsdorf (2020) shows that the CPIs, used to obtain inflation tendencies, are among the most affected statistics by the sudden shock caused by the COVID-19 pandemic since it led to a drastic change in consumer spending patterns. These changes in demand lead to a decrease in revenues, which reduces inflation. According to the Phillips curve, there is a negative relationship between inflation and unemployment (Phillips, 1958) and therefore it should be considered for the analysis we are attempting.

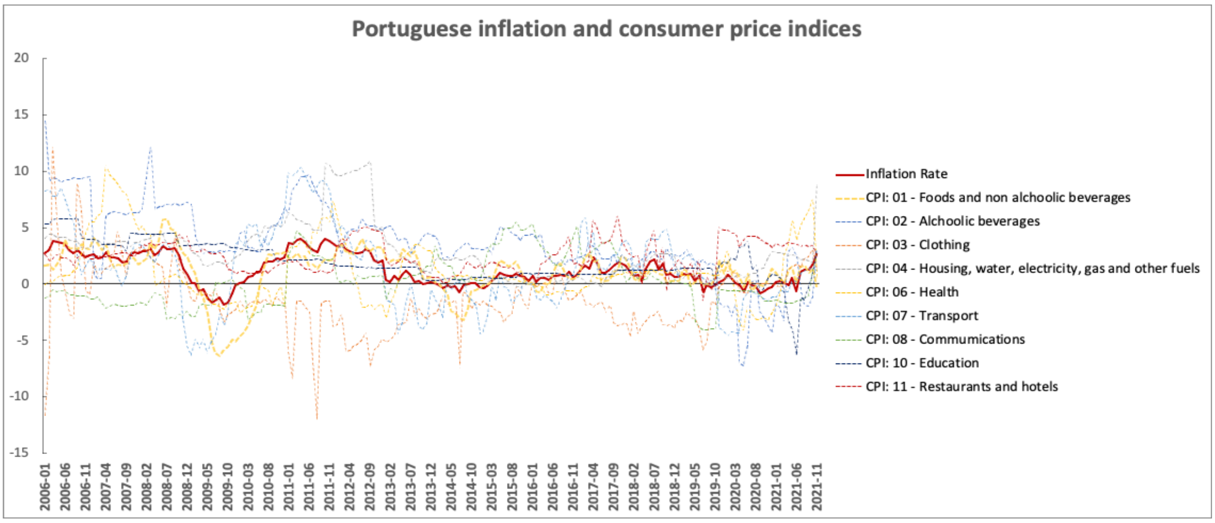


Fig 2. Portuguese inflation rate and CPIs.
Note: The inflation rates and CPIs are seasonally adjusted,
showing the percentage change over the previous year.

Data collected by INE shows that unemployment rose to 7.3% in November 2020, which is 0.5 percentual points higher than one year before, with the tourism and hospitality industry being one of the most affected areas, having almost immediate revenue losses after the first cases were detected in the country (Lopes et al., 2021). It is therefore important to understand the quantity in which these revenues were affected by the pandemic and how much it correlated with the increase in unemployment.

Stock & Watson (1999) have previously provided an analysis of inflation, using the VAR approach. Their analysis used a large array of macroeconomic data to predict inflation, taken from the DRI-McGraw Hill Basic Economics database, which encompasses a multitude of topics, from production, consumption, exchange rates, retail and trade, variables which can prove to be extremely relevant to study in our analysis. Therefore, some of the variables in this study will be based on the ones analyzed by Stock & Watson (1999).

Other than that, some other variables might also be useful to analyze, such as exports and imports. Studies such as (Liu et al., 1997) suggest a causal relationship between economic growth and the level of exports and imports of a country. Ègner (2001) then provides an analysis of Okun's law, which represents the relationships between the economic growth of a country and unemployment, showing that there is a negative relationship between the two. With this in mind, we can place the hypothesis that exports and imports will harm unemployment.

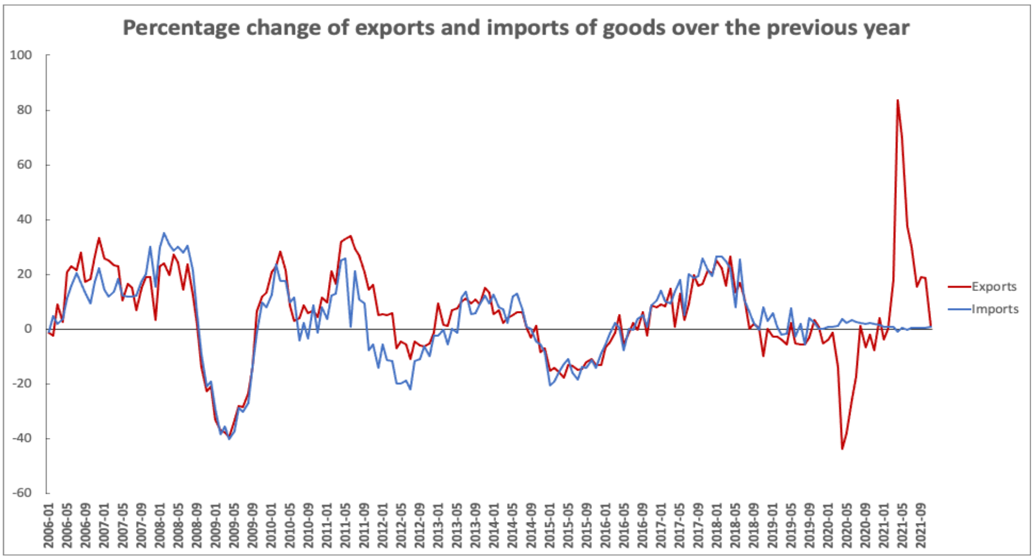


Fig 3. Percentage change of exports and imports of goods over the previous year in Portugal.
Note: The values are seasonally adjusted for both variables

Industrial production will also be analyzed. Gunduz (2020) provides a similar study to the one made by Ègner. In his paper, he comes to similar conclusions, while also showing that unemployment will decrease with the increase in production.

Finally, vehicle registrations and fatalities may also help predict unemployment. A study conducted by He (2016), analyzing the fatality rate of vehicle crashes during the Great Recession shows that each percentage point increase in the unemployment rate predicts a significant 2.9 percent decrease in the fatality rate. Another study, conducted by (Sivak & Schoettle, 2009) using data from October 2007 to April 2009 showed that the number of new vehicles purchased in the U.S. was negatively correlated with the unemployment rate. Therefore, we can also place the hypothesis that an increase in both these variables will lead to a decrease in Portuguese unemployment.

We will also be making use of the number of unemployed people registered with the IEFP since it provides us with an essential metric of the job market in Portugal. As seen in (Coelho, 2003), the probability of more unemployed workers filling job offer inscriptions at IEFP increases with the increase of the unemployment rate. It is therefore relevant to understand if the relation of this variable with unemployment provides relevant conclusions.

3. DATA AND METHODOLOGY

This section presents the input series used to construct the required models for the prediction of unemployment, as well as the required transformations to them, followed by a presentation of the methodology used to access said models.

We intend to construct a multitude of VAR models, both with and without the inclusion of COVID-19 data since they capture the dynamics of linear interactions between variables (Pfaff, 2008). To provide meaningful comparisons to them, we also produced benchmark models, using ARIMA and Artificial Neural Network techniques. We used R language to produce all the models in this analysis.

3.1. DATA

Our variables include monthly information regarding several economic indicators from January 2007 to November 2021, that can prove relevant to predict unemployment. COVID-19 data will be used as exogenous data, since it may impact the way economic data affects unemployment and was collected for the same period as the endogenous data, being then lagged by two months to allow for future out-of-sample forecasting.

We calculated year-on-year rates for the macroeconomic variables that required it, as well as month-to-month rates for all COVID-19 variables, to remove any seasonal effects they might have. Missing values of some of the variables have been filled with the mean of the series. We then made sure to remove highly correlated variables from the analysis, by applying a Spearman correlation test (Artusi et al., 2002), and removed any variable that had a correlation higher than 0.7 with another one. Information regarding both sets of variables can be visualized in Table 1 and Table 2.

Table 1. Macroeconomic indicators evaluated

Series	**Source**	**Measurement**
Unemployment rate	PORDATA	Percentage of the labour force
Number of unemployed people registered with the IEFP	IEFP	Number
Electricity consumption [3]	Eurostat	GWh
Electricity production [3]	Eurostat	GWh
Public Debt [2 4]	Banco de Portugal	Percentage of PIB
Inflation rate [1]	Banco de Portugal	Percentage change over the previous year
Index of retail trade volume [4]	OECD	Index 2015 = 100
Total manufacturing production [1 4]	OECD	Percentage change over the previous year
Total industrial production [1 3]	OECD	Percentage change over the previous year

Total construction [1] [3]	OECD	Percentage change over the previous year
Exports in goods [1] [4]	OECD	Percentage change over the previous year
Imports in goods [1]	OECD	Percentage change over the previous year
Consumer Price indices of Food and non-alcoholic beverages [1] [4]	OECD	Percentage change over the previous year
Consumer Price Indices of Alcoholic beverages, tobacco and narcotics [1]	OECD	Percentage change over the previous year
Consumer Price Indices Clothing and footwear [1]	OECD	Percentage change over the previous year
Consumer Price Indices Housing, water, electricity, gas and other fuels [1]	OECD	Percentage change over the previous year
Consumer price indices of Health [1]	OECD	Percentage change over the previous year
Consumer price indices of Transport [1] [4]	OECD	Percentage change over the previous year
Consumer price indices of Communication [1]	OECD	Percentage change over the previous year
Consumer price indices of Education [1]	OECD	Percentage change over the previous year
Consumer price indices of Restaurants and hotels [1]	OECD	Percentage change over the previous year
Total rail passenger transport	OECD	Passenger-kilometres, Millions
First registration of brand-new passenger cars	OECD	Number
Service Payment index – Total [4]	INE	Index 2015 = 100
Service Payment Index - Wholesale	INE	Index 2015 = 100
Consumption of goods per household [1] [2]	OECD	€
Total assets of investment funds	Banco de Portugal	€ Millions
Investment funds in shares and other equity [4]	Banco de Portugal	€ Millions
Investment funds in Cash and deposits	Banco de Portugal	€ Millions

[1] *Seasonally adjusted data*

[2] *Quarterly data*

[3] *Missing values filled with the mean of the series*

[4] *Variable removed due to high correlation with another one*

Table 2. COVID-19 indicators evaluated

Series	Source	Measurement
Total registered cases	Our World in Data	Number
Sum of new daily cases [1]	Our World in Data	Number
Total registered deaths [1]	Our World in Data	Number
Sum of new daily deaths [1]	Our World in Data	Number
Total number of patients admitted into hospitals	Our World in Data	Number
Total number of patients admitted into ICU's [1]	Our World in Data	Number
Total number of tests [1]	Our World in Data	Number
Sum of new daily tests [1]	Our World in Data	Number
Total number of administered vaccinations	Our World in Data	Number
Sum of new daily administered vaccines [1]	Our World in Data	Number

[1] *Variable removed due to a high correlation with another one.*

3.2. METHODOLOGY

With the final set of variables defined, we started by analyzing the possibility of the models being estimated using the rolling window method. In time series analysis, recent lags often have higher predictive power than older ones, meaning that a model that is applied to the whole series may provide a wrong analysis of the characteristics of the series. To do so, we employed rolling windows of 171 observations. There is a necessity to set them with this many values, since we are working with COVID-19 vaccination information which only exists past 2021 and needs to be addressed in the models, or else they run into linear correlation issues.

We also intend to make a more thorough selection of variables for the VAR and Artificial Neural Network models, while using the entire set of data in the ARIMA models. The idea here is to produce a very basic benchmark model that will be compared with the robust VAR models. The Artificial Neural Network is then used as a powerful comparison to confirm if VAR models can indeed be a good methodology to predict unemployment. We base our choice of variables on the literature review of this paper as well as on the Granger causality test between unemployment and the remaining variables being studied. This test is used to determine if one time series is useful for forecasting the other. If the p-value is less the 0.05, for example, then the hypothesis that the times series is not useful for the forecasting of the other is rejected (Rossi, 2013). We make sure to choose variables that Granger cause unemployment, since it is shown that using variables as covariates to forecast another can significantly improve the forecast error of said dependent variable (Damásio & Nicolau, 2014) and that an absence of statistical evidence for the test may be due to non-existing relationships between them (Damásio & Mendonça, 2022).

We then confirm the quality of the models by looking at the Box-Pierce test of the residuals, guaranteeing there is no autocorrelation and analyse the RMSE, MSE and MAE of the out-of-sample predictions to confirm their quality, the R^2 to see if the model provides a good representation of the variance of the data and finally by applying a Diebold-Mariano test, which compares the forecasting quality of the models and shows if there are significant changes between them.

3.2.1 ARIMA MODEL

ARIMA models are classified as $ARIMA(p,d,q)$, where p denotes the maximum order of the AR polynomial, d refers to integrated parts of the data set and q denotes moving average parts of the data set, having p,d,q as nonnegative integers. (Mondal et al., 2014)

ARIMA models are widely used to predict linear time series and offer great flexibility in univariate time series (Mondal et al., 2014). It is widely used as it uses only information from the previous period to forecast the next. However, financial time series, as the ones we are dealing with, are known to exhibit non-linearities, having non-normal distributions and being quite volatile, meaning these models are neither robust nor efficient for our analysis. If a simple ARIMA can produce better results than the robust VAR models presented further, then we will be able to conclude that the VAR approach is not appropriate to predict unemployment.

ARIMA models have the following equation:

$$y_t = \alpha + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t \quad [1]$$

where β_i are the parameters of the autoregressive part, θ_i are the parameters of the moving average part and ϵ_t are the error terms.

To produce a comparison benchmark, we produced three ARIMAX models, which allow the inclusion of exogenous variables. We produced a basic model, with no exogenous variables, one with the macroeconomic data captured and one with both macroeconomic data and COVID-19 related variables. These models will be referred to as AR_1, AR_2 and AR_3 from here on.

3.2.2 ARTIFICIAL NEURAL NETWORKS

For several years now, Artificial Neural Networks (ANN) have been used to implement information processing systems, inspired by biological neural structures (Vaz et al., 2021). ANN provide several advantages over other statistical models and prediction techniques. Rather than being programmed, they are trained by exposing the network to individual examples of data. The process is repeated until the Artificial Neural Network recognizes underlying patterns between inputs and outputs (Aiken, 1996). This process also does not require any assumptions about the data, contrary to some of the models being studied here.

There are several types of ANN and the choice of the type to use depends on the problem being studied. Here, we implement a backpropagation algorithm. These neural networks process is divided into two steps (Wang & Zheng, 2009): the forward propagation of information, where the information is provided by the input layer, preprocessed by each hidden layer and passed to the output layer. If the actual output is not equal to the expected output, then the error is sent through the back-propagation process, where the error at the output layer transfers to the input layer through the hidden layer, updating each preprocessed value. These two stages are repeated until the error reduces to an acceptable value.

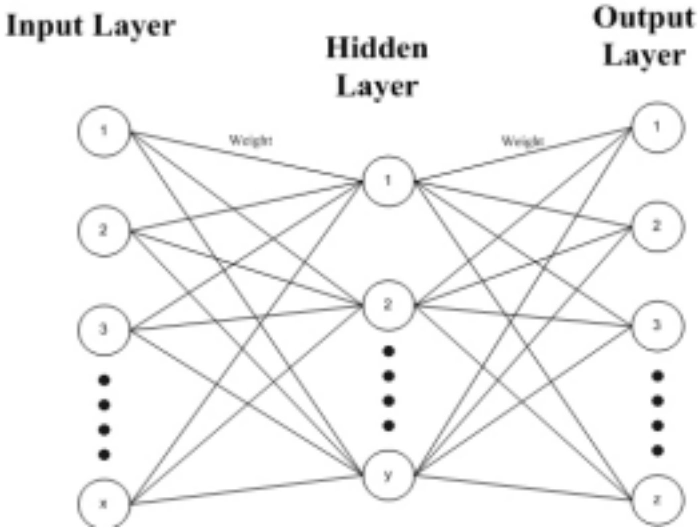


Fig 4. Backpropagation neural network representation.
Note: The input of a node is determined by the weighted sum of the preceding nodes, including a bias term. Image extracted from (Tanty & Desmukh, 2015)

We calculated two different Artificial Neural Networks to predict unemployment. The first one uses only the macroeconomic data identified in section 3.1., while the second one also uses the COVID-19 data provided. Both Artificial Neural Networks are referred to as NN_1 and NN_2 respectively. Both models used 2 hidden layers since it has been proven by the theory that such a model can approach any continuous function with any precision (Wang & Zheng, 2009).

3.2.3 VAR MODELS

Introduced by (Sims, 1980), these models explain the endogenous variables solely by their own history, apart from deterministic regressors, incorporating non-statistical a priori information (Pfaff, 2008) and provides a systematic way to capture linear dynamics in multivariate processes (Damásio et al., 2018).

A VAR model can be defined as the following equation:

$$\boldsymbol{y_t} = \boldsymbol{c} + \boldsymbol{A_1}\boldsymbol{y_{t-1}} + \cdots + \boldsymbol{A_p}\boldsymbol{y_{t-p}} + \boldsymbol{u_t} \qquad [2]$$

where $\boldsymbol{y_t} = (y_{1t}, \dots, y_{kt}, \dots, y_{Kt})$ represents the set of K endogenous variables, $\boldsymbol{c}$ is the k-vector of constants serving as the intercept of the model, $\boldsymbol{A_i}$ are (K x K) coefficient matrices for $i = 1, \dots, p$ and $\boldsymbol{u_t}$ is a K-dimensional process with $E(\boldsymbol{u_t}) = \boldsymbol{0}$ and time invariant positive definite covariance matrix $E(\boldsymbol{u_t}\boldsymbol{u_t^T}) = \sum_u$ (Pfaff, 2008).

Another important characteristic of the VAR model is also its stability. This means it generates a stationary time series with time-invariant means, variances and covariance structure, given sufficient starting values (Pfaff, 2008).

We used this methodology to generate predictions using the variables best suited to predict unemployment. Two models were produced, both without the impact of exogenous COVID-19 related variables and with them, which from here on are referred to as VAR_1 and VAR_2 respectively. Both models were produced using 3 lags, based on the $AIC(n)$, $HQ(n)$ and $SIC(n)$ criterion of selection. All criteria are convergent, which is reassuring since a single criterion is a weak basis to judge a model (Damásio & Mendonça, 2019).

3.2.4 VAR WITH BOOTSTRAP RESAMPLE

Traditional VAR models are mostly limited to 6 to 8 variables, due to the sampling error making results unpredictable in larger models (Sims & Zha, 1998), which can lead to autocorrelation in the residuals of larger models. Due to this, we make use of the bootstrap simulation techniques, which have proven to be a good method to improve the results of autocorrelation tests and deal with normality issues, as seen in (Hatemi-J, 2004). There are several bootstrap techniques as outlined in (Efron & Tibshirani, 1986). Here, however, we make use of the standard bootstrap method, as seen in (Benkwitz et al., 2000) and follow the method presented in this paper and adapted below.

Consider the model:

$$\boldsymbol{y_t} = \sum_{i=i}^{p} \boldsymbol{A_i y_{t-i}} + \boldsymbol{u_t} \tag{3}$$

1) Estimate a VAR model of order p using OLS, obtain the coefficients $\widehat{\mathbf{A}}_{\boldsymbol{i}}$ and calculate the residuals $\widehat{\boldsymbol{u}}_{\boldsymbol{t}}$.
2) With replacement, draw n random samples from $\widehat{\boldsymbol{u}}_{\boldsymbol{t}}$, obtaining $\boldsymbol{u}_{\boldsymbol{t}}^{*}$.
3) Generate the series $\boldsymbol{y}_{\boldsymbol{t}}^{*} = \sum \widehat{\boldsymbol{A}}_{\boldsymbol{i}} \boldsymbol{y}_{\boldsymbol{t-1}}^{*} + \boldsymbol{u}_{\boldsymbol{t}}^{*}$ and estimate $\widehat{\boldsymbol{A}}^{*}$ for each random sample.
4) Construct an estimate of the covariance matrix $\widehat{\boldsymbol{\Omega}}^{*}$, which can be computed as

$$cov(\boldsymbol{y}_{\boldsymbol{t}}^{*}, \boldsymbol{y}_{\boldsymbol{t-k}}^{*}) = \frac{1}{B} \sum_{b=1}^{B} (\boldsymbol{y}_{\boldsymbol{tb}}^{*})(\boldsymbol{y}_{\boldsymbol{t-kb}}^{*})^{T} \tag{4}$$

where $\boldsymbol{y}_{\boldsymbol{tb}}^{*}$ is the bth bootstrap replication.

5) The confidence intervals of the bootstrap are then constructed based on the above statistics.

We used 1000 bootstrap replication to generate the new bootstrapped series and compute the confidence intervals for the VAR model predictions.

3.2.5 BAYESIAN VAR MODELS

Another method to deal with the larger array of variables presented is to apply vector autoregression with Bayesian shrinkage (BVAR) which allows estimations when we have more parameters than observations (Bańbura et al., 2010).

These methods differ from traditional VAR models in that they combine prior information along with the likelihood function to obtain a posterior distribution for the parameters of interest, improving forecasting performance, by using the Bayes theorem (Cuestas & Ordóñez, 2018):

$$\pi(\vartheta|y) = \frac{f(y|\vartheta)\pi(\vartheta)}{f(y)} \quad [5]$$

where ϑ refers to the set of parameters we are interested in estimating, $\pi(\vartheta|y)$ is the posterior distribution of these parameters, conditional to the dataset y, $f(y|\vartheta)$ is the likelihood function, $\pi(\vartheta)$ is a set of priors with a probability distribution and $f(y)$ is the marginal density.

Several different priors have been defined in literature and can be applied to these models. The most common one is the Minnesota prior defined by (Litterman, 1986), as it can handle the issue of unit roots in the data. However, it requires prior knowledge of the residual variance-covariance matrix, which may lead to incorrect assumptions.

Therefore, we test the possibility of using a Stochastic Search Variable Selection (SSVS) VAR. As seen in (Jochmann et al., 2010), SSVS consist of a hierarchical prior where each of the parameters in the VAR is drawn from one of two normal distributions. The first of these has a zero mean and a variance very close to zero. The second also has a zero mean but has a large variance, implying a relatively non-informative prior. This process helps deal with the overparameterization problem presented by a basic VAR since it allows for the shrinkage of parameters.

We apply a "default semi-automatic approach", defined in (George et al., 2008), to select the prior hyperparameters. The pre-selected constants c_0 and c_1 that control prior significance are set to 0.1 and 10 respectively, while the likelihood of each coefficient being a priori equally likely to be included is set to 0.5.

We then further simulate the model by selecting very tight priors for irrelevant variables and relatively uninformative priors for relevant parameters. We do this by considering coefficients relevant if their posterior inclusion probability is above a defined threshold of 50 percent (Barbieri & Berger, 2004) and in doing so, construct a highly probable model. The BVAR models will be defined as $BVAR_1$ and $BVAR_2$ for the analysis without the impact of COVID-19 and with it, respectively.

4. RESULTS

We begin by testing the stationarity of the series, applying an Augmented Dickey-Fuller Test (ADF Test), which is used to test for unit-root non-stationarity in time series data. We confirm that most of our time series are not stationary since the null hypothesis of the test is not rejected for a significance level of 0.05. Therefore, we apply first differences to all macroeconomic time series to make the data stationary and preserve the meaningful relations the variables have between them, obtaining a training set of 171 observations, from February 2007 to April 2021.

Table 3. ADF Test results

Note: The unit root test is applied with an intercept and linear trend regressors. The maximum lag order is 5, based on the data size of the time series.

H0: The time series has a unit root present.

Series	**ADF Test p-value**
Unemployment rate	0.59
Number of unemployed people registered with the IEFP	0.04
Electricity consumption	0.07
Electricity production	0.11
Inflation rate	0.51
Total industrial production	0.07
Total construction	0.02
Imports in goods	0.01
Consumer Price Indices of Alcoholic beverages, tobacco and narcotics	0.09
Consumer Price Indices Clothing and footwear	0.82
Consumer Price Indices Housing, water, electricity, gas and other fuels	0.79
Consumer price indices of Health	0.01
Consumer price indices of Communication	0.28
Consumer price indices of Education	0.18
Consumer price indices of Restaurants and hotels	0.11
Total rail passenger transport	0.01

First registration of brand-new passenger cars	0.01
Service Payment Index - Wholesale	0.11
Consumption of goods per household	0.06
Total assets of investment funds	0.01
Investment funds in Cash and deposits	0.01
Total registered cases	0.05
Total number of patients admitted into hospitals	0.01
Total number of administered vaccinations	0.01

We then test the Granger causality between the endogenous information. The results go mostly in line with the literature review. However, we cannot reject the null hypothesis that inflation and imports do not Granger cause unemployment. With this in mind, we will make use of the CPI of alcoholic beverages, health and education as good indicators for inflation, since we can reject the null hypothesis that these variables do not Granger cause the unemployment rate and will remove imports from the VAR, BVAR and Artificial Neural network models. Finally, regarding the COVID-19 data, we can reject the null hypothesis that these variables do not Granger cause the unemployment rate, meaning they could prove relevant to forecast unemployment.

Table 4. Causality test between unemployment and endogenous data

Causality test	Number of lags	F-Statistic
The number of unemployed people registered with the IEFP does not Granger cause the unemployment rate	3	33.06***
Inflation rate does not Granger cause the unemployment rate	3	1.24
Electricity consumption does not Granger cause the unemployment rate	3	0.91
Electricity production does not Granger cause the unemployment rate	3	0.95
Industrial production does not Granger cause the unemployment rate	3	16.88***
Construction does not Granger cause the unemployment rate	3	61.89***
Imports do not Granger cause the unemployment rate	3	2.07
Consumer Price Indices of Alcoholic beverages, tobacco and narcotics do not Granger cause the unemployment rate	3	20.02***

Consumer Price Indices Housing, water, electricity, gas and other fuels do not Granger cause the unemployment rate	3	0.23
Consumer price indices of Health do not Granger cause the unemployment rate	3	4.35***
Consumer price indices of Communication do not Granger cause the unemployment rate	3	0.27
Consumer price indices of Education do not Granger cause the unemployment rate	3	11.91***
Consumer price indices of Restaurants and hotels do not Granger cause the unemployment rate	3	0.95
Total rail passenger transport does not Granger cause the unemployment rate	3	38.46***
First registration of brand-new passenger cars does not Granger cause the unemployment rate	3	35.05***
Service Payment index - Wholesale do not Granger cause the unemployment rate	3	14.22***
Consumption of goods per household does not Granger cause the unemployment rate	3	75.90***
Total assets of investment funds do not Granger cause the unemployment rate	3	8.72***
Investment funds in Cash and deposits do not Granger cause the unemployment rate	3	0.54
COVID-19 total registered cases do not Granger cause the unemployment rate	3	5.83***
COVID-19 total number of patients admitted into hospitals does not Granger cause the unemployment rate	3	14.76***
COVID-19 total number of vaccinations does not Granger cause the unemployment rate	3	29.23***

* Denotes significance at 10%; ** Denotes significance at 5%; *** Denotes significance at 1%

With the tests applied, the variables chosen for the VAR and Artificial Neural Network models are the following[1]: *unemployment rate; number of unemployed people registered with the IEFP; industrial production; construction; CPI of Alcoholic beverages, tobacco and narcotics; CPI of Health; CPI of Education; number of newly registered passenger vehicles; wholesale service payment index, household consumption; Total assets of investment funds.* The ARIMA models will make use of the entire set.

---

[1] COVID-19 indicators remain the same for all models.

Regarding the benchmark models, for the ARIMAs, we choose the order of the analysis based on the ACF and PACF graphs of the differentiated unemployment series, presented in figure 5. Based on their analysis, we make use of an $ARIMA(2,1,1)$ for all ARIMA benchmark estimations. The neural networks make use of 2 hidden layers, as seen in section 3.

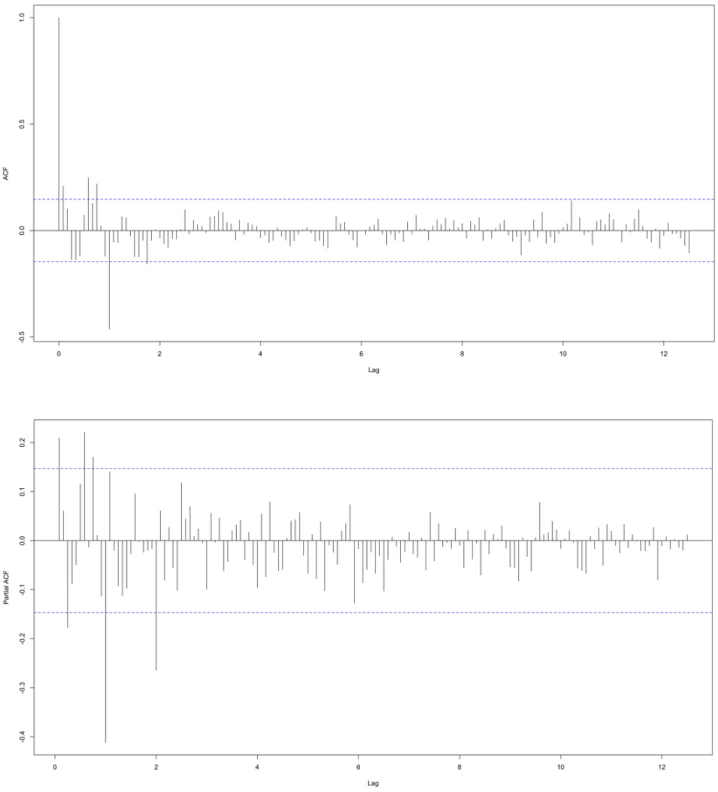


Fig 5. ACF and PACF graphs of the unemployment rate.
Note: The unemployment rate has been differentiated for the analysis.

For the VAR and BVAR models, we explain both the lag decision for the models and the choice of hyperparameters for the BVAR in section 3 of this paper. Table 5 shows the results of fitting the VAR and BVAR models. We obtain very positive results for the multiple and adjusted R^2 of the VAR models, showing that the variables chosen can explain almost all the variance present in the unemployment rate.

Table 5. VAR and BVAR model coefficients for the unemployment rate.
Note: In parenthesis are presented the s.e. of the coefficient estimations.

	VAR_1	VAR_2	$BVAR_1$	$BVAR_2$
Unemployment (-1)	0.12 (0.08)	0.09 (0.08)	0.00 (0.00)	0.00 (0.00)
IEFP registrations (-1)	0.02 (0.11)	0.04 (0.11)	0.00 (0.00)	0.00 (0.00)
Industrial Production (-1)	0.03 (0.06)	0.01 (0.06)	0.00 (0.00)	0.00 (0.00)
Construction (-1)	0.03 (0.06)	-0.02 (0.06)	0.00 (0.00)	0.00 (0.00)
CPI Alchool (-1)	0.10 (0.19)	0.01 (0.19)	0.00 (0.00)	0.00 (0.00)

CPI Health (-1)	0.07 (0.22)	0.04 (0.22)	0.00 (0.00)	0.00 (0.00)
CPI Education (-1)	-0.09 (0.51)	-0.15 (0.49)	0.00 (0.00)	0.00 (0.00)
New registered vehicles (-1)	0.00 (0.01)	0.01 (0.01)	0.00 (0.00)	0.00 (0.00)
Service Payment Index (-1)	0.05 (0.05)	0.01 (0.05)	0.00 (0.00)	0.00 (0.00)
Investments (-1)	0.10 (0.11)	0.07 (0.11)	0.00 (0.00)	0.00 (0.00)
Household consumption (-1)	0.19 (0.16)	0.11 (0.16)	0.42 (0.08)	0.00 (0.00)
Unemployment (-2)	0.22 (0.08)	0.17 (0.08)	0.17 (0.06)	0.00 (0.00)
IEFP registrations (-2)	0.05 (0.12)	0.01 (0.12)	0.00 (0.00)	0.00 (0.00)
Industrial Production (-2)	0.06 (0.07)	0.08 (0.07)	0.00 (0.00)	0.00 (0.00)
Construction (-2)	-0.17 (0.06)	-0.17 (0.06)	-0.23 (0.03)	0.00 (0.00)
CPI Alchool (-2)	-0.36 (0.19)	-0.44 (0.19)	0.00 (0.00)	0.00 (0.00)
CPI Health (-2)	0.07 (0.23)	0.08 (0.22)	0.00 (0.00)	0.00 (0.00)
CPI Education (-2)	-0.07 (0.50)	-0.53 (0.50)	0.00 (0.00)	0.00 (0.00)
New registered vehicles (-2)	0.00 (0.01)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)
Service Payment Index (-2)	-0.09 (0.05)	-0.07 (0.05)	0.00 (0.00)	0.00 (0.00)
Investments (-2)	0.03 (0.11)	0.01 (0.11)	0.00 (0.00)	0.00 (0.00)
Household consumption (-2)	-0.16 (0.18)	-0.19 (0.17)	0.00 (0.00)	-0.66 (0.08)
Unemployment (-3)	-0.04 (0.08)	-0.04 (0.08)	0.00 (0.00)	0.00 (0.00)
IEFP registrations (-3)	0.21 (0.11)	0.33 (0.11)	0.47 (0.07)	0.40 (0.06)
Industrial Production (-3)	-0.16 (0.07)	-0.14 (0.07)	-0.20 (0.05)	-0.21 (0.05)
Construction (-3)	-0.15 (0.06)	-0.12 (0.06)	-0.22 (0.04)	0.00 (0.00)
CPI Alchool (-3)	0.43 (0.20)	0.44 (0.19)	0.30 (0.15)	0.00 (0.00)
CPI Health (-3)	0.05 (0.22)	-0.06 (0.22)	0.00 (0.00)	0.00 (0.00)
CPI Education (-3)	0.29 (0.40)	-0.25 (0.41)	0.00 (0.00)	0.00 (0.00)
New registered vehicles (-3)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.00 (0.00)
Service Payment Index (-3)	-0.10 (0.05)	-0.04 (0.06)	0.00 (0.00)	0.00 (0.00)
Investments (-3)	-0.24 (0.11)	-0.24 (0.11)	-0.26 (0.08)	0.00 (0.00)
Household consumption (-3)	-0.05 (0.15)	-0.16 (0.15)	0.00 (0.00)	0.00 (0.00)
const	-0.02 (0.17)	-0.02 (0.17)	0.00 (0.00)	0.00 (0.00)
Total cases	-	-0.02 (0.01)	-	0.04 (0.01)

Total hospitalized patients	-	-0.03 (0.01)	-	-0.05 (0.01)
Total vaccinations	-	0.01 (0.00)	-	0.00 (0.00)
Model Evaluation				
Multiple R^2	0.74	0.77	-	-
Adjusted R^2	0.68	0.70	-	-
F-Statistic	12.11 on 33 and 141 df	12.55 on 36 and 138 df	-	-

We tested the out-sample forecasting performance of the models by comparing it to the set of data from May 2021 to November 2021. We analyzed the quality of predictions in all rolling windows and evaluated the changes considering different forecasting horizons, from 1 to 7 months. The forecasting results can be seen below in figure 6. We then bootstrapped the confidence intervals of the VAR models, to confirm if the results present in the standard VAR models can still be accurate for longer forecasting horizons. We make use of 1000 resamples to obtain the 95% confidence bands for the forecasts of the VAR models. Figure 7 shows the results.

We can conclude that the BVAR models provide a closer representation of the real data, when compared to the benchmark models and the VAR models, especially for longer forecasting horizons. The standard VARs provide more accurate results for shorter horizons than the benchmark models however, the predictions seem to get increasingly worse for longer forecasting horizons. These results are backed by figure 8. By bootstrapping the VAR model confidence bands, we see that the forecasts fall inside the 95% confidence bands, apart from the 7-step-ahead forecasts of VAR_2. This can be a result of the overparameterization present in this model. VAR models can prove unreliable when dealing with a high number of dimensions, which combined with the large forecasting horizon, leads to inaccurate estimations. The Baysen method can deal with this overparameterization and therefore the BVAR model estimations remain accurate for longer forecasting horizons.

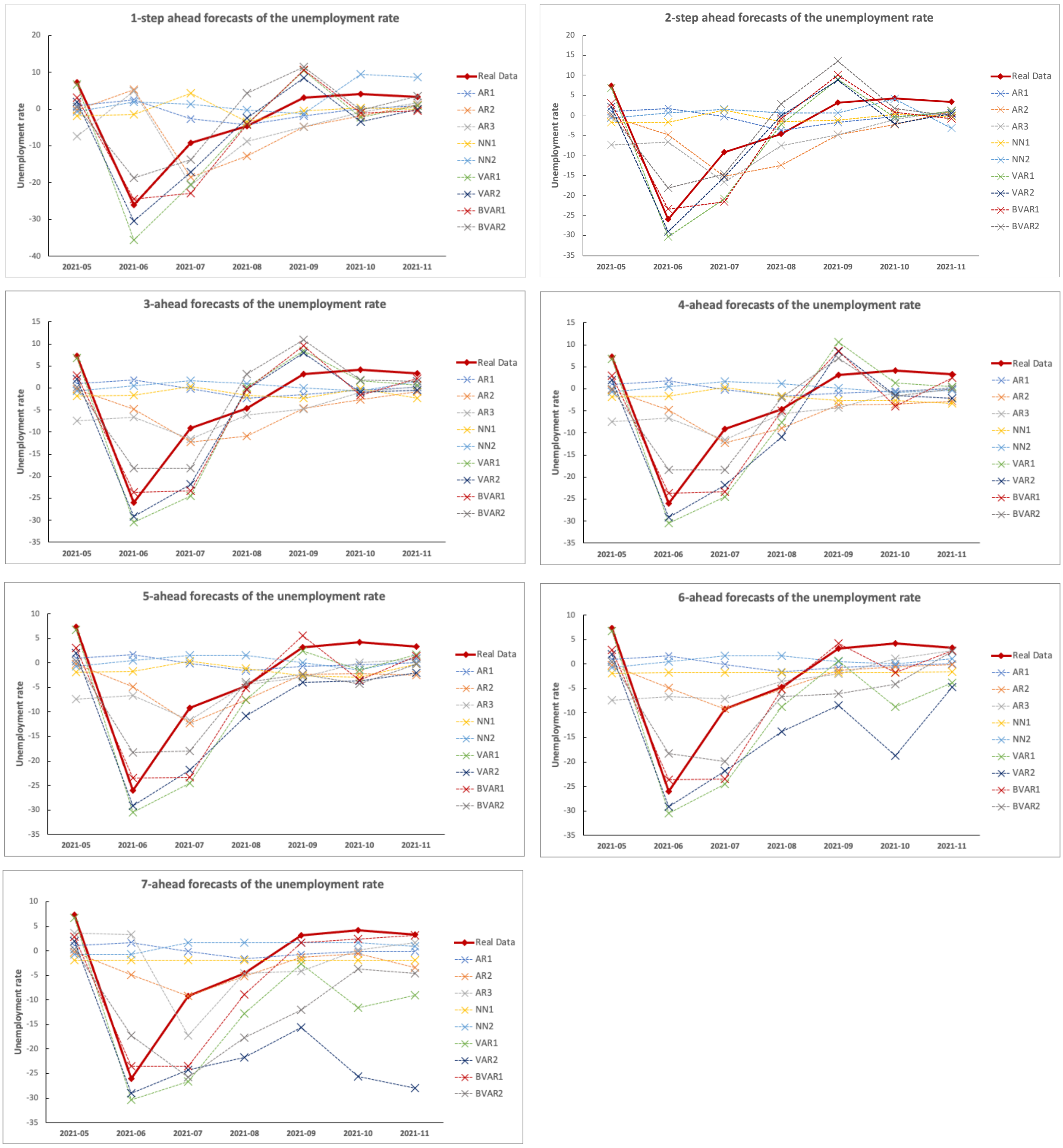
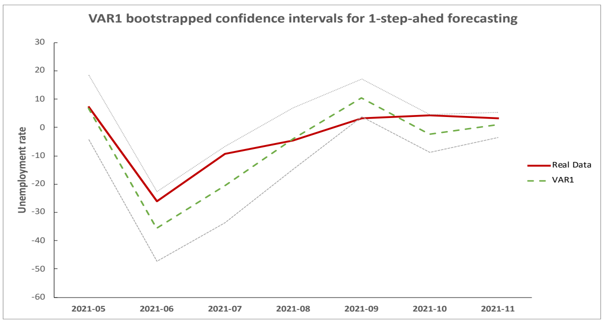
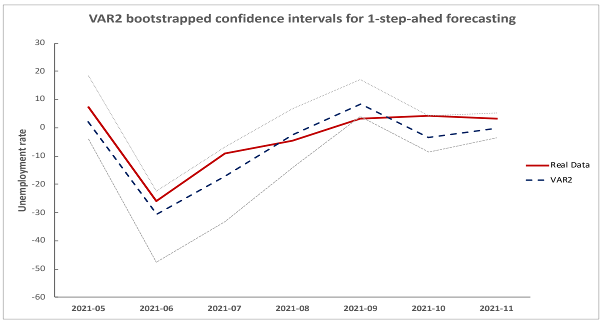
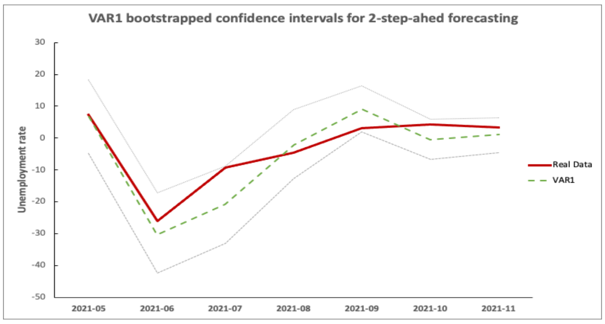
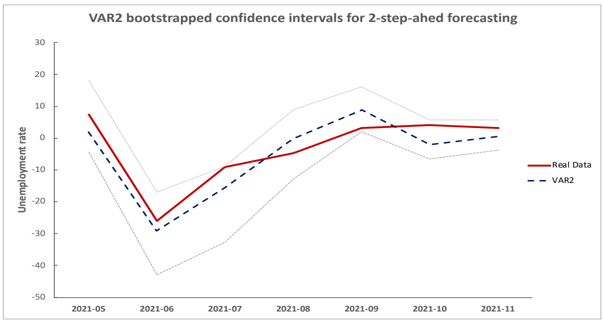


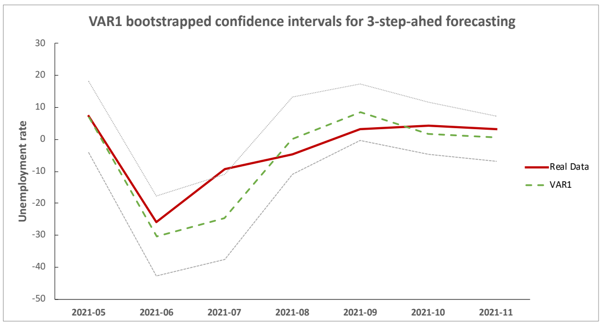
Fig 6. 1 through 7-step-ahead out-of-sample forecasts for the unemployment rate.
Note: We compare the forecasted out-of-sample results from the different models with the real data from the same period. The predictions are compared to the stationary data of the unemployment rate. The models were calculated using multiple rolling windows for the different horizons being studied. The predicted values are averaged for each rolling window and horizon.

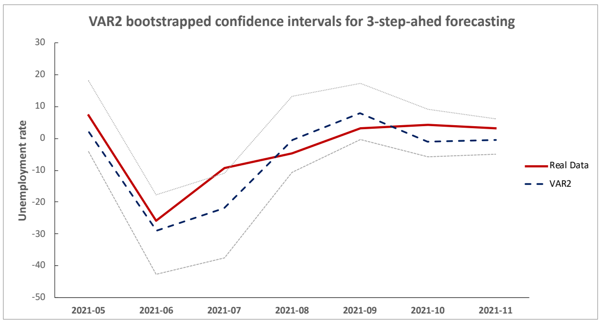

VAR1 bootstrapped confidence intervals for 1-step-ahed forecasting
Unemployment rate
30
20
10
0
-10
-20
-30
-40
-50
-60
2021-05
2021-06
2021-07
2021-08
2021-09
2021-10
2021-11
Real Data
VAR1

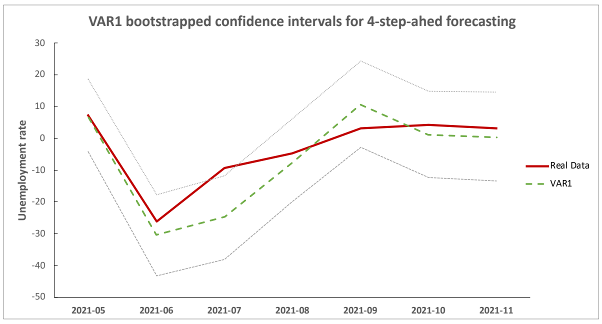

VAR2 bootstrapped confidence intervals for 1-step-ahed forecasting
Unemployment rate
30
20
10
0
-10
-20
-30
-40
-50
-60
2021-05
2021-06
2021-07
2021-08
2021-09
2021-10
2021-11
Real Data
VAR2

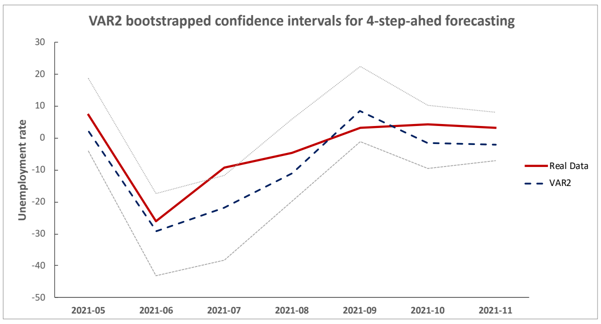

VAR1 bootstrapped confidence intervals for 2-step-ahed forecasting
Unemployment rate
30
20
10
0
-10
-20
-30
-40
-50
2021-05
2021-06
2021-07
2021-08
2021-09
2021-10
2021-11
Real Data
VAR1

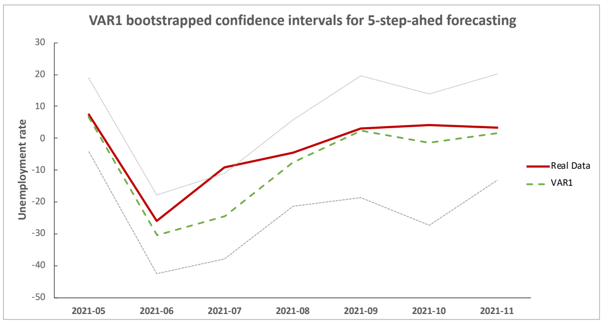

VAR2 bootstrapped confidence intervals for 2-step-ahed forecasting
Unemployment rate
30
20
10
0
-10
-20
-30
-40
-50
2021-05
2021-06
2021-07
2021-08
2021-09
2021-10
2021-11
Real Data
VAR2

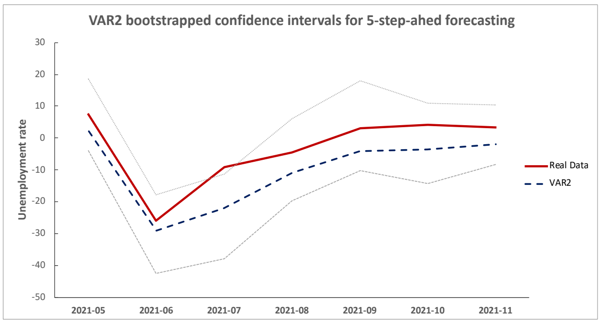

VAR1 bootstrapped confidence intervals for 3-step-ahed forecasting
Unemployment rate
30
20
10
0
-10
-20
-30
-40
-50
2021-05
2021-06
2021-07
2021-08
2021-09
2021-10
2021-11
Real Data
VAR1


VAR2 bootstrapped confidence intervals for 3-step-ahed forecasting
Unemployment rate
30
20
10
0
-10
-20
-30
-40
-50
2021-05
2021-06
2021-07
2021-08
2021-09
2021-10
2021-11
Real Data
VAR2


VAR1 bootstrapped confidence intervals for 4-step-ahed forecasting
Unemployment rate
30
20
10
0
-10
-20
-30
-40
-50
2021-05
2021-06
2021-07
2021-08
2021-09
2021-10
2021-11
Real Data
VAR1


VAR2 bootstrapped confidence intervals for 4-step-ahed forecasting
Unemployment rate
30
20
10
0
-10
-20
-30
-40
-50
2021-05
2021-06
2021-07
2021-08
2021-09
2021-10
2021-11
Real Data
VAR2


VAR1 bootstrapped confidence intervals for 5-step-ahed forecasting
Unemployment rate
30
20
10
0
-10
-20
-30
-40
-50
2021-05
2021-06
2021-07
2021-08
2021-09
2021-10
2021-11
Real Data
VAR1


VAR2 bootstrapped confidence intervals for 5-step-ahed forecasting
Unemployment rate
30
20
10
0
-10
-20
-30
-40
-50
2021-05
2021-06
2021-07
2021-08
2021-09
2021-10
2021-11
Real Data
VAR2

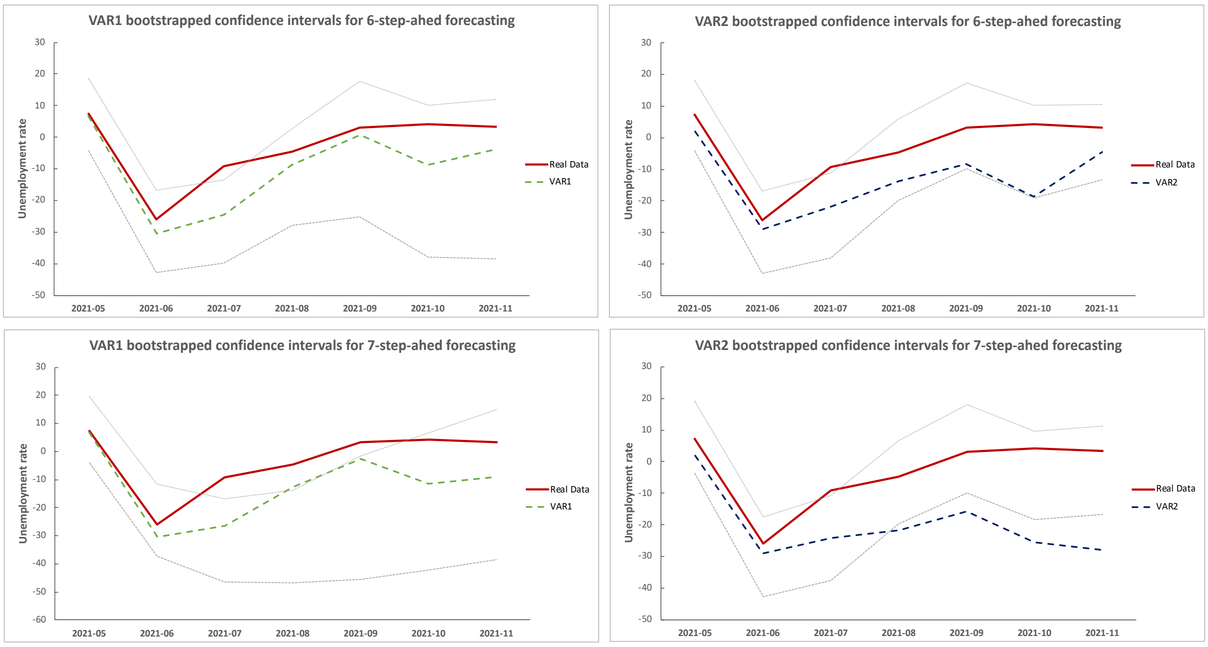


Fig 7. VAR_1 and VAR_2 bootstrapped confidence intervals for the 1 through 7-step-ahead forecasts estimated. *Note:* The confidence intervals are generated by resampling the data 1000 times. All figures present the 95% confidence intervals, represented by the grey dotted lines. The dashed green line presents the forecasted values for the i-step-ahead horizon of VAR_1 while the dashed blue line presents the results of VAR_2.

To analyse the forecasting models, we tested the presence of heteroskedasticity in the residuals through a Box-Pierce test. We can assume no autocorrelation and heteroskedasticity of residuals at a 95% confidence level for all ARIMA and VAR models.

Table 6. Autocorrelation test for ARIMA and VAR model residauls

	AR_1	AR_2	AR_3	VAR_1	VAR_2
Box-Pierce test p-value	0.91	0.71	0.95	0.53	0.67

We then evaluate the RMSE, MSE and MAE of the out-of-sample forecasts for the different models and horizons tested. Figure 8 shows the average out-of-sample forecasting performance of the residuals, for the various models. As expected, lower forecasting horizons provide lower prediction errors. The predicted performances of all VAR models were superior to the benchmark models only for lower horizons and the BVAR models provided the lowest RMSE, MSE and MAE across all models. This is also expected as the Bayesian approach can deal with the overparameterization problem of normal VAR models.

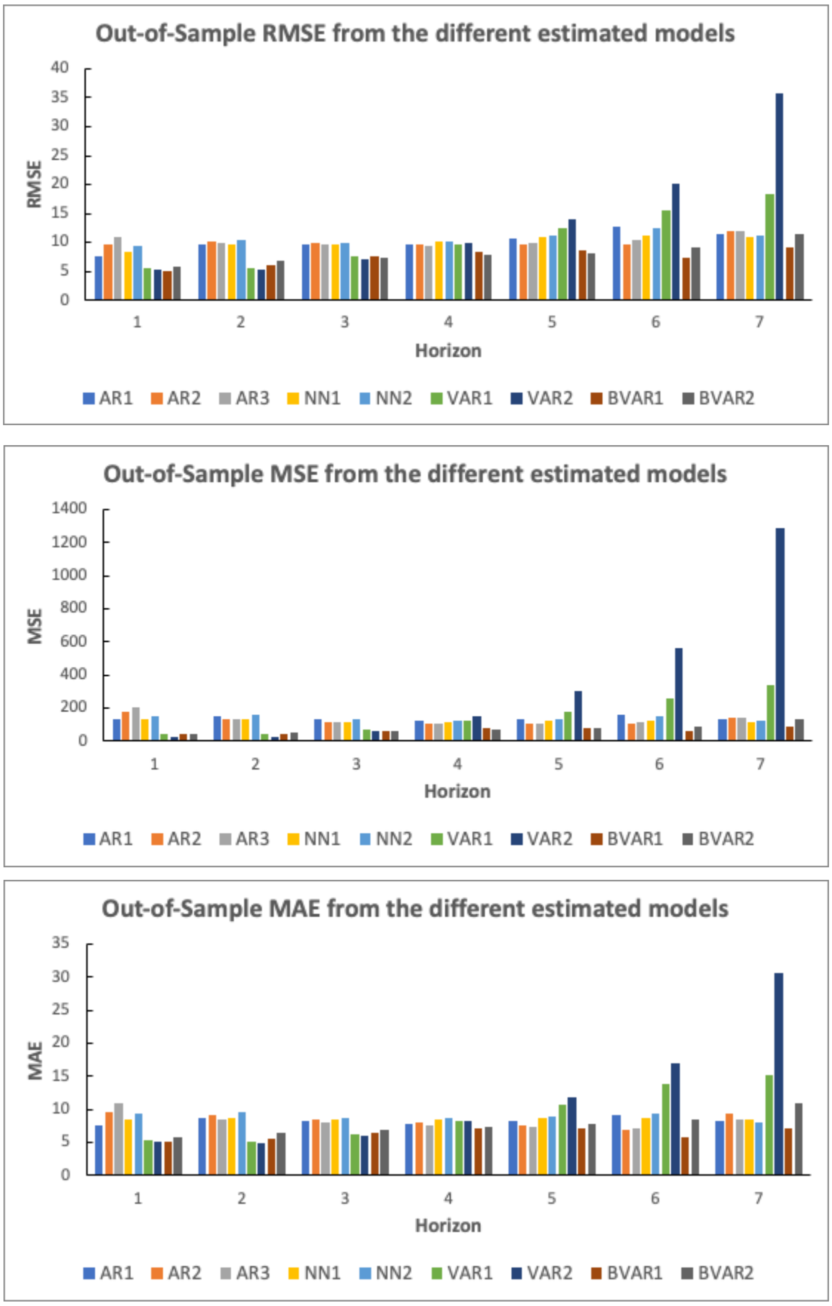


Fig 8. Out-of-sample RMSE, MSE and MAE for the different estimated models.
Note: The graphs show the average RMSE, MSE and MAE over the different rolling windows calculated for the various horizons and models. The calculations are made by comparing the real data from May 2021 to November 2021 with the predicted values.

To conclude, we apply the Diebold-Mariano test to the forecasting results, to understand if the models do indeed provide different accuracies. For simplicity, we calculate the test for the 1-step-ahead and 7-step-ahead forecasts, to confirm if longer horizons present differences in the forecasting results. Results can be seen in tables 7 and 8. The test indicates that apart from AR_1, the models provide the same accuracy for a 1-step-ahead prediction, as we are not able to reject the null hypothesis that the forecasts have different accuracies for a significance value of 0.05. However, for a longer forecasting horizon, we reject the null hypothesis when comparing the VAR and BVAR models. We can reject the null hypothesis that VAR_2 has the same accuracy as the remaining VAR and BVAR model forecasts and reject the null hypothesis that VAR_1 has the same accuracy as $BVAR_1$, for a significance level of 0.05, something that was expected from the previous results.

It is important to note that, although we can not assume the models have different accuracies for the most part, a low p-value does not necessarily mean that one model is better than the other and only means that the forecast errors of the two models are different. Therefore, it is important to consider the remaining metrics presented in this paper to assess the quality of the models.

Table 7. Diebold-Mariano test p-value results for 1-step-ahead forecasts.
H0*:* The forecasts from model 1 and model 2 have the same accuracy.

	AR_1	AR_2	AR_3	NN_1	NN_2	VAR_1	VAR_2	$BVAR_1$	$BVAR_2$
AR_1									
AR_2	0.012								
AR_3	0.006	0.483							
NN_1	0.890	0.295	0.131						
NN_2	0.159	0.202	0.065	0.449					
VAR_1	0.366	0.231	0.166	0.213	0.251				
VAR_2	0.327	0.215	0.155	0.196	0.228	0.274			
$BVAR_1$	0.415	0.275	0.206	0.277	0.305	0.866	0.491		
$BVAR_2$	0.367	0.236	0.173	0.245	0.263	0.841	0.478	0.952	

Table 8. Diebold-Mariano test p-value results for 7-step-ahead forecasts
H0*:* The forecasts from model 1 and model 2 have the same accuracy.

	AR_1	AR_2	AR_3	NN_1	NN_2	VAR_1	VAR_2	$BVAR_1$	$BVAR_2$
AR_1									
AR_2	0.252								
AR_3	0.536	0.289							
NN_1	0.493	0.075	0.491						
NN_2	0.612	0.176	0.573	0.559					
VAR_1	0.883	0.691	0.842	0.992	0.943				
VAR_2	0.197	0.063	0.232	0.126	0.158	0.020			
$BVAR_1$	0.343	0.475	0.356	0.319	0.299	0.011	0.012		
$BVAR_2$	0.980	0.523	0.926	0.858	0.936	0.742	0.074	0.000	

5. CONCLUSIONS

In this paper, we analyze the possibility of the VAR approach being a good prediction method for the Portuguese unemployment rate, while also confirming if the inclusion of COVID-19 related data improves the prediction results. We came to the following conclusions.

5.1 THEORETICAL IMPLICATIONS

The VAR models do in general predict better than other approaches, such as ARIMA and Neural Networks, when estimated for shorter horizons. When estimated for longer horizons, the combination of overparameterization and long forecasting horizons leads to more inaccurate results than the benchmark models. The Bayesian approach for VAR models provides the best results as it can more accurately deal with the overparameterization problem of VARs.

The results here are in line with the ones found in (Sugita, 2022) which also proves that the SSVS prior can provide lower MSE and better forecasts for longer horizons than traditional estimators. We also conclude that, as the forecast horizon increases, the quality of the models decreases. VAR model forecasts also worsen substantially more as the forecasting horizon increases, when compared to Neural Networks, something that is consistent with the literature review presented.

5.2 PRACTICAL IMPLICATIONS

The inclusion of exogenous COVID-19 data provides no significant improvement to the forecasting ability of the models. Although COVID-19 caused an enormous economic crisis, its impact on unemployment seems to have been quickly mitigated, possibly due to quick government and companies' actions. The impact of the pandemic was felt more in some sectors than others, as some companies could easily transition to remote work while others could not. It is likely that the ones that were able to do so were also larger businesses than the ones that could not. This is relevant because, in this study, we analysed the global unemployment rate of Portugal across all sectors, disregarding this nuance. It might prove relevant to study the effects of the pandemic across independent sectors and analyse if there were indeed differences in the way COVID-19 affected unemployment.

5.3 LIMITATIONS AND FUTURE WORK

Due to the inclusion of exogenous information regarding vaccinations, the models have a smaller test set than intended. With the progression of the pandemic and the rollout of vaccinations, it might be relevant to apply this methodology to larger forecasting horizons, making use of more recent information.

REFERENCES

Aiken, M. (1996). A neural network to predict civilian unemployment rates. *Journal of International Information Management*, *5*(1). https://scholarworks.lib.csusb.edu/jiim/vol5/iss1/3

Artusi, R., Verderio, P., & Marubini, E. (2002). Bravais-Pearson and Spearman Correlation Coefficients: Meaning, Test of Hypothesis and Confidence Interval. *The International Journal of Biological Markers*, *17*(2), 148–151. https://doi.org/10.1177/172460080201700213

Balcerzak, A. P., & Zurek, M. (2011). Foreign direct investment and unemployment: VAR analysis for Poland in the years 1995-2009. *European Research Studies*, *XIV*(1), 2011. https://www.um.edu.mt/library/oar/handle/123456789/31401

Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, *25*(1), 71–92. https://doi.org/10.1002/jae.1137

Barbieri, M. M., & Berger, J. O. (2004). Optimal predictive model selection. *The Annals of Statistics*, *32*(3), 870–897. https://doi.org/10.1214/009053604000000238

Barro, R. J., Ursúa, J. F., Weng, J., Ursúa Dodge, J. F., & Weng EverLife, J. (2020). *The Coronavirus and the Great Influenza Pandemic: Lessons from the "Spanish Flu" for the Coronavirus's Potential Effects on Mortality and Economic Activity*. https://www.nber.org/papers/w26866

Benkwitz, A., Neumann, M. H., & Lütekpohl, H. (2000). Problems related to confidence intervals for impulse responses of autoregressive processes. *Econometric Reviews*, *19*(1), 69–103. https://doi.org/10.1080/07474930008800460

Blustein, D. L., Duffy, R., Ferreira, J. A., Cohen-Scali, V., Cinamon, R. G., & Allan, B. A. (2020). Unemployment in the time of COVID-19: A research agenda. *Journal of Vocational Behavior*, *119*. https://doi.org/10.1016/J.JVB.2020.103436

Claveria, O. (2019). Forecasting the unemployment rate using the degree of agreement in consumer unemployment expectations. *Journal for Labour Market Research*, *53*(1), 1–10. https://doi.org/10.1186/S12651-019-0253-4/FIGURES/5

Coelho, J. M. (2003). *Duração do Desemprego e das Vagas de Emprego no Mercado de Trabalho Português*. https://www.bportugal.pt/sites/default/files/anexos/papers/ab200302_p.pdf

Cuestas, J. C., & Ordóñez, J. (2018). Oil prices and unemployment in the UK before and after the crisis: A Bayesian VAR approach. A note. *Physica A: Statistical Mechanics and Its Applications*, *510*, 200–207. https://doi.org/10.1016/J.PHYSA.2018.06.114

Damásio, B., Louçã, F., & Nicolau, J. (2018). The changing economic regimes and expected time to recover of the peripheral countries under the euro: A nonparametric approach. *Physica A: Statistical Mechanics and Its Applications*, *507*, 524–533. https://doi.org/10.1016/J.PHYSA.2018.05.089

Damásio, B., & Mendonça, S. (2019). Modelling insurgent-incumbent dynamics: Vector autoregressions, multivariate Markov chains, and the nature of technological competition. *Applied Economics Letters*, *26*(10), 843–849. https://doi.org/10.1080/13504851.2018.1502863

Damásio, B., & Mendonça, S. (2022). Leader-follower dynamics in real historical time: a Markovian test of non-linear causality between sail and steam (co-)development. *Applied Economics*. https://doi.org/10.1080/00036846.2022.2100868

Damásio, B., & Nicolau, J. (2014). Combining a regression model with a multivariate Markov chain in a forecasting problem. *Statistics & Probability Letters*, *90*(1), 108–113. https://doi.org/10.1016/J.SPL.2014.03.026

Darko, S. (2017). Inflation, Unemployment and Economic Growth: Evidence from the VAR Model Approach for the Economy of Iraq. *International Journal of Developing and Emerging Economies*, *5*(1), 26–39. https://www.researchgate.net/publication/329625161

Dolado, J. J., & Jimeno, J. F. (1997). The causes of Spanish unemployment: A structural VAR approach. *European Economic Review*, *41*(7), 1281–1307. https://doi.org/10.1016/S0014-2921(97)00058-5

Duarte, V. (2020). *The Portuguese labour market in times of the pandemic*. https://www.caixabankresearch.com/en/economics-markets/labour-market-demographics/portuguese-labour-market-times-pandemic

Efron, B., & Tibshirani, R. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science*, *1*(1), 54–75. https://doi.org/10.1214/SS/1177013815

Ègner, L. S. (2001). Okun's Law: Does the Austrian unemployment-GDP relationship exhibit structural breaks? *Empirical Economics*, *26*, 553–564. https://link.springer.com/content/pdf/10.1007/s001810000070.pdf

George, E. I., Sun, D., & Ni, S. (2008). Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics*, *142*(1), 553–580. https://doi.org/10.1016/J.JECONOM.2007.08.017

Giles, J., Park, A., & Zhang, J. (2005). What is China's true unemployment rate? *China Economic Review*, *16*(2), 149–170. https://doi.org/10.1016/J.CHIECO.2004.11.002

Gunduz, M. (2020). The Link between Unemployment and Industrial Production: The Fourier Approach with Structural Breaks. *FOREIGN EXPERIENCE*, *13*(3), 228–240. https://doi.org/10.15838/esc.2020.3.69.15

Hatemi-J, A. (2004). Multivariate tests for autocorrelation in the stable and unstable VAR models. *Economic Modelling*, *21*, 661–683. https://doi.org/10.1016/J.ECONMOD.2003.09.005

He, M. M. (2016). Driving through the Great Recession: Why does motor vehicle fatality decrease when the economy slows down? *Social Science & Medicine*, *155*, 1–11. https://doi.org/10.1016/J.SOCSCIMED.2016.02.016

INE. (2021). *Estimativas Mensais Emprego e Desemprego*. https://www.gee.gov.pt/pt/indicadores-diarios/ultimos-indicadores/32650-estimativas-mensais-de-emprego-e-desemprego-ine-15

Jochmann, M., Koop, G., & Strachan, R. W. (2010). Bayesian forecasting using stochastic search variable selection in a VAR subject to breaks. *International Journal of Forecasting*, *26*(2), 326–347. https://doi.org/10.1016/J.IJFORECAST.2009.11.002

Katris, C. (2020). Prediction of Unemployment Rates with Time Series and Machine Learning Techniques. *Computational Economics*, *55*(2), 673–706. https://doi.org/10.1007/S10614-019-09908-9/TABLES/19

Lai, H., Khan, Y. A., Thaljaoui, A., Chammam, W., & Abbas, S. Z. (2021). COVID-19 pandemic and unemployment rate: A hybrid unemployment rate prediction approach for developed and developing countries of Asia. *Soft Computing*. https://doi.org/10.1007/s00500-021-05871-6

Litterman, R. B. (1986). Forecasting with Bayesian Vector Autoregressions: Five Years of Experience. *Journal of Business & Economic Statistics*, *4*(1), 25. https://doi.org/10.2307/1391384

Liu, X., Song, H., & Romilly, P. (1997). An empirical investigation of the causal relationship between openness and economic growth in China. *Applied Economics*, *29*, 1679–1686. https://doi.org/10.1080/00036849700000043

Lopes, A. S., Sargento, A., & Carreira, P. (2021). Vulnerability to COVID-19 unemployment in the Portuguese tourism and hospitality industry. *International Journal of Contemporary Hospitality Management*, *33*(5), 1850–1869. https://doi.org/10.1108/IJCHM-11-2020-1345

Marcellino, M., Stock, J. H., Watson, M. W., & Nber, the. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, *135*, 499–526. https://doi.org/10.1016/j.jeconom.2005.07.020

Martins, D., & Damásio, B. (2020). One Troika fits all? Job crash, pro-market structural reform and austerity-driven therapy in Portugal. *Empirica*, *47*(3), 495–521. https://doi.org/10.1007/S10663-019-09433-W/FIGURES/1

Mondal, P., Shit, L., Goswami, S., & Professor, A. (2014). Study of Effectiveness of Time Series Modeling (Arima) in Forecasting Stock Prices. *International Journal of Computer Science Engineering and Applications*, *4*(2). https://doi.org/10.5121/ijcsea.2014.4202

Mulaudzi, R., & Ajoodha, R. (2020). Application of Deep Learning to Forecast the South African Unemployment Rate: A Multivariate Approach. *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering*. https://doi.org/10.1109/CSDE50874.2020.9411581

Naccarato, A., Falorsi, S., Loriga, S., & Pierini, A. (2018). Combining official and Google Trends data to forecast the Italian youth unemployment rate. *Technological Forecasting and Social Change*, *130*, 114–122. https://doi.org/10.1016/J.TECHFORE.2017.11.022

Nigam, R., Tripathi, G., Priya, T., Luis, A. J., Vaz, E., Kumar, S., Shakya, A., Damásio, B., & Kotha, M. (2022). Did Covid-19 lockdown positively affect the urban environment and UN- Sustainable Development Goals? *PLOS ONE*, *17*(9), e0274621. https://doi.org/10.1371/JOURNAL.PONE.0274621

Olmedo, E. (2014). Forecasting Spanish Unemployment Using Near Neighbour and Neural Net Techniques. *Computational Economics*, *43*(2), 183–197. https://doi.org/10.1007/S10614-013-9371-1/TABLES/6

Ozili, P. (2020). *Spillover of COVID-19: impact on the Global Economy*. https://ssrn.com/abstract=3562570

Pfaff, B. (2008). VAR, SVAR and SVEC Models: Implementation Within R Package vars. *Journal of Statistical Software*, *27*(4). http://www.jstatsoft.org/

Phillips, A. W. (1958). The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861-1957. *Economica*, *25*(100), 283. https://doi.org/10.2307/2550759

Reinsdorf, M. (2020). *WP/20/224 COVID-19 and the CPI: Is Inflation Underestimated?* https://ssrn.com/abstract=3758057

Rossi, B. (2013). Advances in forecasting under instability. *Handbook of Economic Forecasting*, *2*, 1203–1324. https://doi.org/10.1016/B978-0-444-62731-5.00021-X

Şahin, A., Tasci, M., & Yan, J. (2020). The Unemployment Cost of COVID-19: How High and How Long? *Economic Commentary (Federal Reserve Bank of Cleveland)*, 1–7. https://doi.org/10.26509/frbc-ec-202009

Sims, C. A. (1980). Macroeconomics and Reality. *The Econometric Society*, *48*(1), 1–48. https://www.jstor.org/stable/1912017

Sims, C. A., & Zha, T. (1998). Bayesian Methods for Dynamic Multivariate Models. *International Economic Review*, *39*(4), 949–968. https://doi.org/10.2307/2527347

Sivak, M., & Schoettle, B. (2009). *Economic Indicators as Predictors of the Number and Fuel Economy of Purchased New Vehicles*. http://www.umich.edu/~umtriswt

Stock, J. H., & Watson, M. W. (1999). Forecasting inflation. *Journal of Monetary Economics*, *44*, 293–335. https://www.princeton.edu/~mwatson/papers/Stock_Watson_JME_1999.pdf

Stock, J. H., & Watson, M. W. (2002). Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association*, *97*(460), 1167–1179. https://www.jstor.org/stable/3085839

Sugita, K. (2022). Forecasting with Bayesian vector autoregressive models: comparison of direct and iterated multistep methods. *Asian Journal of Economics and Banking*, *6*(2), 142–154. https://doi.org/10.1108/AJEB-04-2022-0044

Tanty, R., & Desmukh, T. S. (2015). Application of Artificial Neural Network in Hydrology- A Review. *International Journal of Engineering Research and Technology*, *V4*(06). https://doi.org/10.17577/IJERTV4IS060247

Thomakos, D. D., & Guerard, J. B. (2004). Naïve, ARIMA, nonparametric, transfer function and VAR models: A comparison of forecasting performance. *International Journal of Forecasting*, *20*(1), 53–67. https://doi.org/10.1016/S0169-2070(03)00010-4

Vaz, E., Bação, F., Damásio, B., Haynes, M., & Penfound, E. (2021). Machine learning for analysis of wealth in cities: A spatial-empirical examination of wealth in Toronto. *Habitat International*, *108*, 102319. https://doi.org/10.1016/J.HABITATINT.2021.102319

Vieira, D. A., & Meirinhos, V. (2021). COVID-19 Lockdown in Portugal: Challenges, Strategies and Effects on Mental Health. *Trends in Psychology*, *29*(2), 354–374. https://doi.org/10.1007/s43076-021-00066-2

Wang, G., & Zheng, X. (2009). The Unemployment rate forecast model basing on BP neural network. *Proceedings - 2009 International Conference on Electronic Computer Technology, ICECT 2009*, 475–478. https://doi.org/10.1109/ICECT.2009.58

Zivot, E., & Wang, J. (2006). *MODELING FINANCIAL TIME SERIES WITH S-PLUS* (Issue 2).