Pontifícia Universidade Católica
do Rio de Janeiro

Marcos Lopes Muniz

Forecasting employment and unemployment in US. A comparison between models.

Dissertação de Mestrado

Dissertation presented to the Programa de Pós–graduação em Economia da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Economia .

Advisor: Prof. Marcelo Medeiros

Rio de Janeiro
April 2020

Pontifícia Universidade Católica
do Rio de Janeiro

Marcos Lopes Muniz

Forecasting employment and unemployment in US. A comparison between models.

Dissertation presented to the Programa de Pós–graduação em Economia da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Economia . Approved by the Examination Committee.

Prof. Marcelo Medeiros
Advisor
Departamento de Economia – PUC-Rio

Prof. Eduardo Zilberman
Departamento de Economia – PUC-Rio

Prof. Diogo Abry Guillen
Departamento de Economia – PUC Rio

Rio de Janeiro, April the 3rd, 2020

Marcos Lopes Muniz

B.A. in Economics, University of Sao Paulo (USP-SP), 2013.

Bibliographic data

Lopes Muniz, Marcos

Forecasting employment and unemployment in US. A comparison between models. / Marcos Lopes Muniz; advisor: Marcelo Medeiros. – Rio de janeiro: PUC-Rio, Departamento de Economia, 2020.

v., 47 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Economia.

Inclui bibliografia

1. Economia – Teses. 2. Econometria – Teses. 3. Mercado de Trabalho;. 4. Previsão;. 5. Taxa natural de desemprego;. 6. Random Forest;. 7. Long Short Term Memory;. I. Medeiros, Marcelo. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Economia. III. Título.

CDD: 620.11

Acknowledgments

First, I would like to thank my Mother Jurema Maria L. S. Muniz and my brothers Lucas and Thiago for the support, knowledge and comprehension that always pushed me to do better and go further.

Also I would like to thank my advisor Marcelo Medeiros for guidance and advices through this work. And last, would like to thank my classmates for the comradery during these two years.

Financial support from CNPq and CAPES is gratefully acknowledged.

Abstract

Lopes Muniz, Marcos; Medeiros, Marcelo (Advisor). **Forecasting employment and unemployment in US. A comparison between models.**. Rio de Janeiro, 2020. 47p. Dissertação de mestrado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

Forecasting employment and unemployment is of great importance for virtually all agents in the economy. Employment is one of the main variables analyzed as an economic indicator, and unemployment serves to policy makers as a guide to their actions. In this essay, I study what features of both series we can use on data treatment and methods used to add to the forecasting predictive power. Using an AR model as a benchmark, I compare machine (Random Forest and Adaptive Lasso) and deep (Long Short Term Memory) learning methods, seeking to capture non-linearities of both series dynamics. The results suggests that an AR model with a Random Forest on residuals (as a way to separate linear and non-linear part) is the best model for employment forecast, while Random Forest and AdaLasso with Random Forest on residuals were the best for unemployment forecast.

Keywords

Labor market; forecast; Natural rate of unemployment; Random Forest; Long Short Term Memory;

Resumo

Lopes Muniz, Marcos; Medeiros, Marcelo. **Prevendo emprego e desemprego nos EUA. Uma comparação entre modelos.**. Rio de Janeiro, 2020. 47p. Dissertação de Mestrado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

Prever emprego e desemprego é de grande importância para praticamente todos os agentes de uma economia. Emprego é uma das principais variáveis analisadas como indicador econômico, e desemprego serve para os policy makers como uma orientação às suas decisões. Neste trabalho, eu estudo quais características das duas séries podemos usar para auxiliar no tratamento dos dados e métodos empregados para auxiliar no poder preditivo das mesmas. Eu comparo modelos de machine (Random Forest e Lasso Adaptativo) e Deep (Long short Term memory) learning, procurando capturar as não linearidades e dinâmicas de ambas séries. Os resultados encontrados sugerem que o modelo AR com Random Forest aplicado nos resíduos, como uma maneira de separar parte linear e não linear, é o melhor modelo para previsão de emprego, enquanto Random Forest e AdaLasso com Random Forest aplicado nos resíduos são os melhores para o desemprego.

Palavras-chave

Mercado de Trabalho; Previsão; Taxa natural de desemprego; Random Forest; Long Short Term Memory;

Table of contents

List of figures

List of tables

List of Abreviations

NAIRU – Non-accelerating inflation rate of unemployment
SPF – *Survey of Prefessional Forecasters*
LSTM – Long Short Term Memory
TAR – Threshold Autoregressive Model
MSA – Markov Switching Autoregression
ARTM – Autoregressive trend model
LSTAR – Logistic smooth transition autoregressive model
ARDL – Autoregressive distributed lag
ML – Machine Learning
BLS – *Boureau of Labour Statistics*
ADF – Augmented Dickey-Fuller
RNN – Recurrent Neural Network

1
Introduction

Although closely related, there is a difference between the series analyzed. The unemployment rate refers to the unemployed as a share of the labour force, being labour force defined by people that have jobs plus people seeking for a job inside the working age. Employment growth is the number of people employed divided by the number of people employed in a precedent period. There could be a scenario where both series rise and move to the same direction. It happens if the number of people employed grows less than the labour force.

Unemployment is one of the main variables to measure the well being of a society. It shows the level of spare capacity of a country labour market, as well as can generate extra costs for the government, regarding unemployment insurance and more people using public services. As the Boureau of Labor Statistics (BLS) states, "Addressing the issue of unemployment... can be used by policymakers to determine whether measures should be taken to influence the future course of the economy or to aid those affected by joblessness."

Employment is one of the main variables to compose the coincident indicators of a business cycle created in the literature. The idea of coincident and leading variables was first suggested in Burns and Mitchell (1946), where he suggests four variables which have similar dynamics as the reference cycle, being employment, income, output and trade. The construction of a coincident indicator index was discussed deeply in Stock and Watson (1989), where they propose a way to construct an index (XCI) to explain the business cycles. While the concept of business cycles is hard to define, Stock and Watson (1989) suggest the index is to explain the "co-movements across several aggregate time series". And this co-mevements could be interpreted as reference business cycles. The employment in XCI, measured by employees on non agricultural payroll has the biggest weight of the the four variables in the index (48%). While income has weight around 22,5%, production of 15,5% and sales 14%. Issler and Vahid (2006) proposes a different Coincident Indices of economic activity, where the employment has an even more important role in the business cycle, being responsible for 84% of the variation in the index. The weight of employment in all of those indexes suggests it is one of the most, if not the most, important variable regarding economic activity. Hall et al. (2001) in a study of 2001 Recession, states "employment is probably the single most reliable indicator" of a recession.

The forecasting literature historically gave more attention to the unemployment forecast, giving less attention to the employment growth forecast. Although the former is the concept with more economic theory involved, being the one that goes in the Philips Curve and the Okun's law[1], the latter should be the variable to be analyzed if the goal of the forecaster is to use as a proxy for economic activity.

For those reasons, the forecast comparison was made with both series. I.e. unemployment forecast has more studies to base upon and employment forecast is more useful as an activity indicator.

1.1 Contribution

This study adds to the literature in a few ways. First, when it compares alternative models and elucidates the difference in predictive power between employment and unemployment series. Although they have an obvious high correlation, the performance in each series forecasts differs. The model that shows the best predictive power in longer horizons for employment, the AR with random forest in the residual, did not perform as well for unemployment.

Also, not only comparing models, but the attempt was to explore the nature of the data in order to choose what model to implement in order to deal with specific characteristics of the series. For example, You can find the NAIRU concept being used and discussed in the inflation forecast. Stock and Watson (1999) suggests Philip's curve can contribute to the inflation forecast, while Atkeson et al. (2001) argues simple random walks can consistently beat any Philip's curve prediction. There is no similar discussion in the unemployment forecast. In this study, I try to make a similar approach. So in order to use this concepts, I decompose the unemployment series in cycle and trend, or, in other words, in natural rate of unemployment and business cycle. This attempt didn't show good predictive power.

I used different models chosen from the winners of different unemployment forecast horse races in the literature. Coulombe et al. (2019) compares different machine learning methods for the forecast of different macroeconomic variables, and argues the Random Forest was the best model for the unemployment forecast. Similarly, Cook and Hall (2017) uses deep learning models and compare to the SPF, finding that Encoder-Decoder, a Recurrent Neural

[1]One possible reason for this imbalance in both forecasting literatures, is that there is not an established theory that tries to explain the population growth and the individual decision of participating in the labour force or not.

network based on a Long Short term Memory, is the best forecast model. In the models used in my study, LSTM and the Random Forest are included, among others, for comparison.

Lastly, this study contributes being one of the few studies to compare Machine Learning and Deep learning Models in the employment forecast. Doing a forecast comparison between models with employment as variable of interest contributes to a quite neglected variable in the forecasting literature if compared to other macroeconomic variables.

1.2 Literature

The literature discussed extensively how to deal with the asymmetry of the unemployment series. Mitchell (1927) noticed discrepancy in duration between expansion and recession periods. Neftci (1984) was one of the first to test the asymmetry in the business cycles, focusing in the unemployment rate of US. Montgomery et al. (1998) uses this asymmetry to propose non-linear models, like TAR or MSA, for forecast using quarterly data. Their findings suggest non linear models improve the predictive power compared to the linear models, but the ones used were still worst than the forecasts of SPF. Proietti (2003) extended this exercise with monthly data, and using non linear cyclical trends. He uses a kalman filter for estimating via maximum likelihood an ARTM model, being the forecast a direct result of this model. Differently than his approach, I use the kalman filter with the only goal to disentangle the cycle and trend, and do the forecast using a large dataset as explanatories for the cycle forecast. Proietti's results is not conclusive, but suggests structural time series model can be a very useful tool.

Skalin and Teräsvirta (2002) uses a LSTAR model to deal with the asymmetry, and includes the rate of unemployment as explanatory, with the first difference as variable of interest (and it's lags as explanatory) arguing he starts "from the realistic assumption that the unemployment rate is a stationary variable". Stock and Watson (2002b) [2] used diffusion indexes to forecast all four variables that compose the Index of Coincident Economic Indicator, being employment among those variables. After this work, Rapach and Strauss (2008) was one of the few to discuss the employment forecast. He uses ARDL models and focused mainly on gaining predictive power combining different forecasts.

[2] At the same year, Stock and Watson (2002a) uses a large dataset to forecast industrial production using principal components, but as it will be clearer in the next section their use of large datasets began before 2002

Recently, the forecasting literature as general seen a great advance in new forecasting models that demands more computational power. There are quite a few papers that compares different machine learning models as a forecasting competition. Hall (2018) uses a Elastic Net model to forecast unemployment and compare to Blue Chip unemployment forecast, a less common survey that collects US macroeconomic professional forecasts. Coulombe et al. (2019), mentioned earlier, used numerous methods to forecast different macroeconomic variables. Smeekes and Wijler (2018) also uses different ML methods to forecast numerous macroeconomic variables, including employment, but focus only on one month horizon forecasts. All of those, used the dataset constructed by McCracken and Ng (2016), also used in this study, and will be better explained on the next section.

1.3 Organization

Section 2 describes the dataset used and the transformation made for stationarity. Section 3 describes the general framework and how we approach the forecasting problem. Section 4 shows each model used for forecasting comparison. And sections 5 and 6 shows the results and concludes, respectively.

2
Data

2.1
Sources

Stock and Watson (1996) was their first work constructing a large dataset, initially with 76 variables. In subsequent works, as Stock and Watson (1998) and Stock and Watson (2002b), they improved this dataset adding more variables and getting what became to be known as the "Stock-Watson database". In an attempt to consolidate a dataset with the same characteristics and predictive power of Stock Watson‘s, McCracken and Ng (2016) builds their own database making it available in their website, and with the propelling feature of being updated in real time. Since then, their database is largely used in the forecasting literature, for the easy access, facilitating replication of methods and allowing a more meaningful comparison between studies.

For this study, I used McCracken and Ng (2016) database, available in McCracken's webpage[1].

As an attempt to gain predictive power disaggregating the employment in different sectors, I also used employment level by sectors. The disaggregated employment series were taken from BLS (Bureau of Labor Statistics) website. I used 11 series representing the eleven sectors of the economy that together fully compose the Non-farm payroll series. Those sectors are: Mining and Logging; Construction; Manufacturing; Trade, Transportation and utilities; Information; Financial Activities; Professional and Business Services; Education and health services; Leisure and hospitality; Other Services; and Government. The non farm payroll excludes part of workers (farmers, some government workers, private households, non profit employees and proprietors), since those sectors have high seasonal fluctuations and hard data collection, among other reasons.[2]

[1]https://research.stlouisfed.org/econ/mccracken/fred-databases/

[2]https://www.stlouisfed.org/open-vault/2019/july/nonfarm-payrolls-why-farmers-not-included

2.2 Period

The first observation of the McCracken's dataset monthly series is from January 1959. Since 6 of 128 variables have their first observation only in January 1960, I restricted the analysis beginning in January 1960. The last observation in the sample of this work is July 2019.

In McCracken's webpage, he separates the variables in groups and suggests different transformations according to characteristics of each group. Additionally to the transformation suggested (first differences), I also made the forecasting exercise using Seasonal Differences, as an attempt to capture the horizon dynamics of the economy. In ARRF there were gains in predictive power using seasonal differences. On the other hand, Adaptive Lasso was worst in longer horizons considering seasonal differences. For this reason, the results showed in this work are using first differences for AdaLasso and and AdaLasso with Random Forest in the residual, and seasonal differences in all other models.

Table 2.1: Transformation in explanatory variables

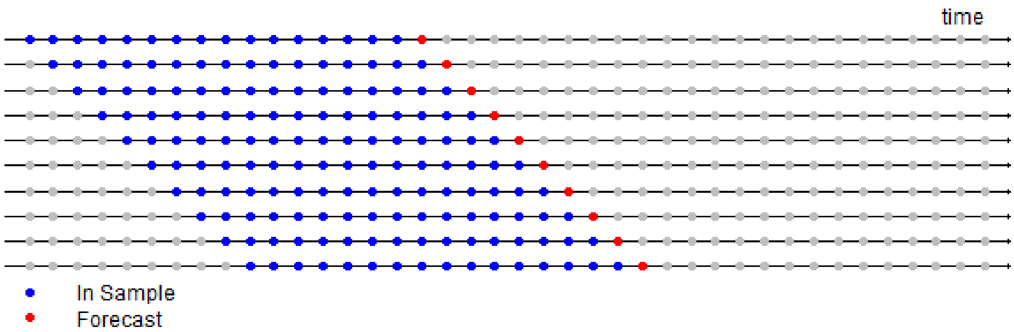
Number of Variables	1^{st} difference	Seasonal difference
11 variables	x_t	x_t
19 variables	Δx_t	$\Delta^h x_t$
10 variables	$log(x_t)$	$log(x_t)$
53 variables	$\Delta log(x_t)$	$\Delta^h log(x_t)$
34 variables	$(\Delta log(x_t))^2$	$(\Delta^h log(x_t))^2$

Once the main goal of this study is the comparison between models, I used only the vintage as of July 2019, without worrying what was the information set available at the time of each rolling window. In other words, I didn‘t considered issues regarding revised data or release dates of each variable.

3
Empirical Strategy

The sample used for estimating our model was a fixed length rolling window of 40 years, as scheme of Figure 3.1. The main reason for using a rolling window is that there are some models that could be benefited from an expanding window approach (most likely the neural network model). So doing an expanding window could potentially generate a scenario where one model be worst in the beginning (with less in-sample data) and better at the end (with more in-sample data) compared to the others. This would be hard to identify and make the comparison between models performance more problematic. The forecasts are made considering 5 different horizons, being 1, 3, 6, 12 and 24 months.

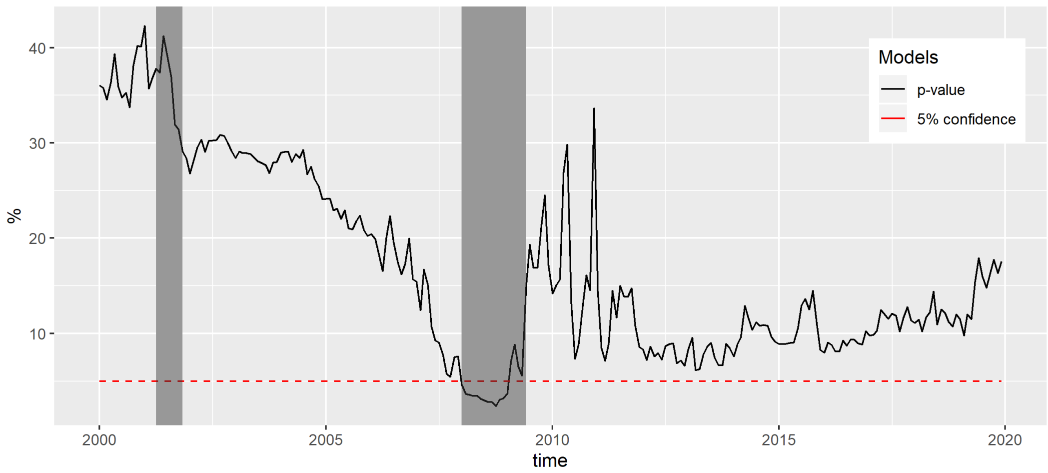
Figure 3.1: Rolling Window Approach



Unemployment (rate) and Employment (level) suffer different treatment. There is a vast literature regarding what treatment should be done in unemployment, since logics suggests we take the rate as it is, but hypothesis of stationarity in unemployment series normally are rejected following the standard tests. Roberts and Morin (1999) makes an analysis focused on US data confronting the hysteresis theory and the natural rate theory. Their results suggests there is no evidence for a unit root in unemployment series in US, i.e. there is no evidence supporting the idea of existing hysteresis in the unemployment rate. In figure 3.2, you can find simple ADF tests made on unemployment, considering each window used for estimation.

Following McCracken and Ng (2016), we take differences in unemployment rate and take differences on the log of the level of employment. Since we

Figure 3.2: P-value of US unemployment on each window



Notes: p-value of ADF test for the entire unemployment series (1939-2019) is 2,36%. Considering only the period analysed in our database (1960-2019), is 9,14%.

are interested in the forecast of the rate, for unemployment, and level, for employment, I take seasonal differences according to the horizon to be forecasted. So the variables of interest are:

$$\text{Unemployment:} \quad y_{t+h} = \Delta^h u_{t+h} = u_{t+h} - u_t$$
$$\text{Employment:} \quad y_{t+h} = \Delta^h ln(e_{t+h}) = ln(e_{t+h}) - ln(e_t)$$

The path of both series follow opposite tendencies and have high correlation between them, as expected. But the dynamics of them are a bit different, as the performance of the models are different. In figures 3.3 to 3.7, both series are put together in the same graph for comparison, with employment multiplied by minus one. It is clear to notice the employment series is more volatile than unemployment. Employment occasionally has it's peak in recessions greater, in absolute terms, than the unemployment series. Also, it maintains it's value below, in absolute terms, in almost all the expanding cycles.

Regarding the explanatory variables, I considered 3 lags for the autoregressive part, while for all other variables I considered only one lag. I started the exercise with four lags in all variables but it didn't contributed to the predictive power and were more computationally expensive [1]

Calling all explanatory variables, excluding the autoregressive variables, as x_t, the forecast equations are below. As you can see, the forecasts are made independently in each horizon, so the forecast for one month ahead doesn‘t have any influence in the forecast of 3 months ahead, and so on.

[1]With 1 lag in X_t, there were 128 variables as regressor. With 4 lags of X_t, there were 612 variables as regressor.

$$\Delta^h u_{t+h} = f(x_t, u_t, \Delta^h u_t, \Delta^h u_{t-1}, \Delta^h u_{t-2}) + \varepsilon_{t+h}$$

$$\Delta^h ln(e_{t+h}) = f(x_t, \Delta^h ln(e_t), \Delta^h ln(e_{t-1}), \Delta^h ln(e_{t-2})) + \varepsilon_{t+h}$$

As already mentioned, the dataset considered is from January 1960 to July 2019, hence having 715 observations. The 40 year rolling window corresponds to 480 monthly observations. So, the number of forecasts made depends on the horizon of the forecasts. Since the first window ends in December 1999, the first forecast for one month horizon is for January 2000, the first forecast for a 3 month horizon is March 2000, for a 6 month horizon is June 2000 and so on. So the number of forecasts made is equal to $715 - 480 - h$, being h the forecast horizon.

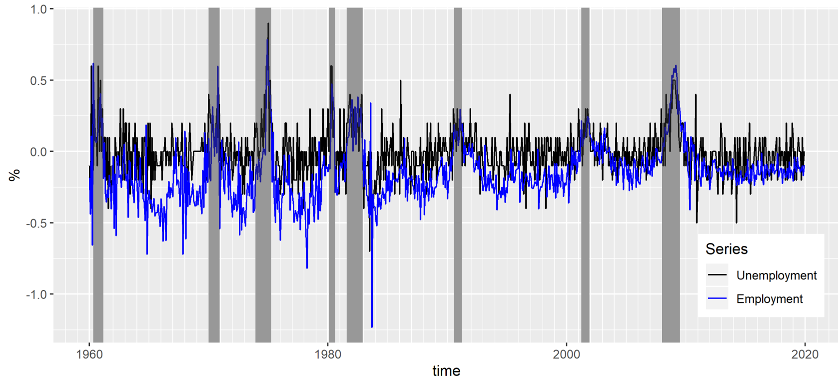


Figure 3.3: 1 month difference (correlation of -0,4655)

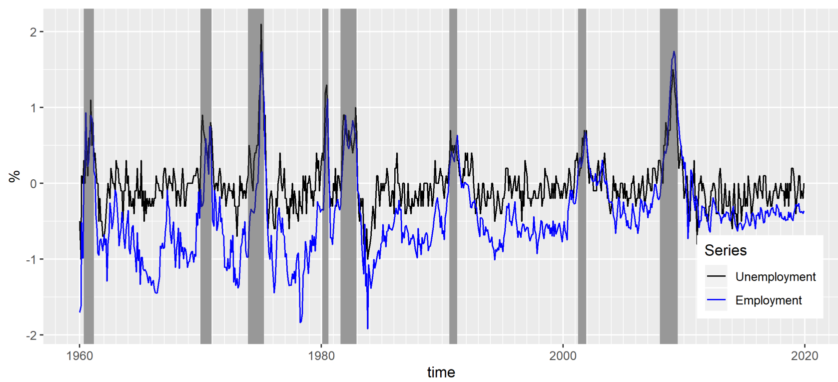


Figure 3.4: 3 months difference (correlation of -0,7417)

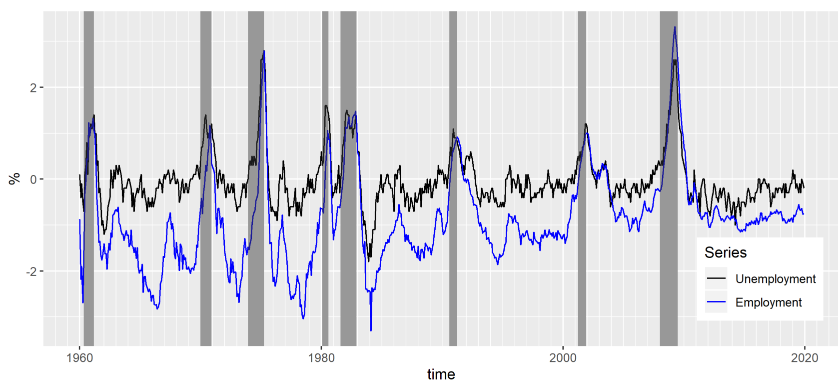


Figure 3.5: 6 months difference (correlation of -0,8195)

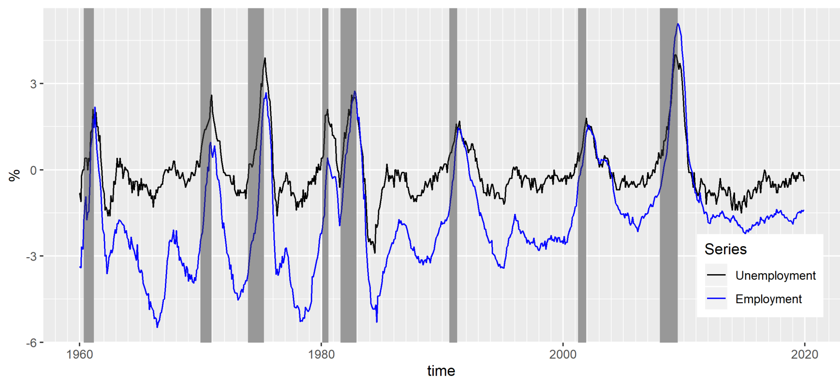


Figure 3.6: 12 months difference (correlation of -0,8291)

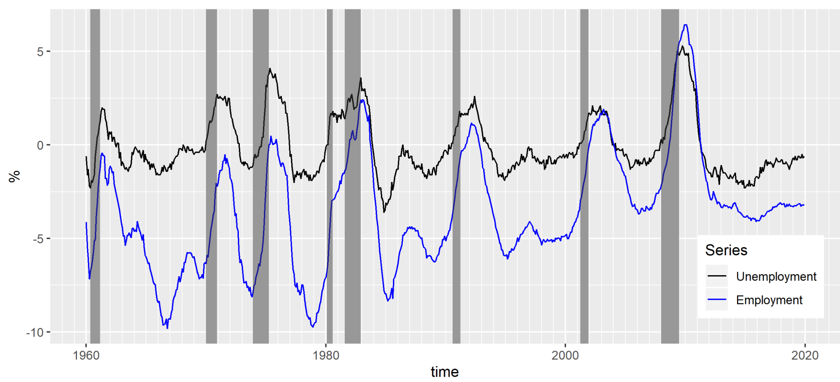


Figure 3.7: 24 months difference (correlation of -0,7961)

4
Models

4.1
Benchmark Model

The benchmark model used for both employment and unemployment is the AR-3 model. In the unemployment autoregressive model I included the rate of unemployment without differences as additional regressor, since it generates gains of performance, as suggested in Skalin and Teräsvirta (2002). So the forecasting equations are the following.

$$\text{Unemployment:} \quad \widehat{y_{t+h}} = \widehat{\Delta^h u_{t+h}} = \hat{\beta}_0 + \sum_{i=1}^{3} \hat{\beta}_i y_{t+1-i} + \hat{\beta}_4 u_t$$

$$\text{Employment:} \quad \widehat{y_{t+h}} = \Delta^h \widehat{ln(e_{t+h}}) = \hat{\beta}_0 + \sum_{i=1}^{3} \hat{\beta}_i y_{t+1-i}$$

4.2
Random Forest

Coulombe et al. (2019) tested different machine learning models forecasting different macroeconomic variables. For the unemployment forecast, the model that had the best performance was the Random Forest. For this reason we included this model as a candidate.

Random Forest is the average of numerous random regression trees, and one of the main benefits of its use is to capture non linear dynamics between the regressors and the variable of interest.

Each Random Forest estimation made used 500 regression trees, and each tree had to have a minimum of 10 observations in order for a nod to be created.

4.3
Long Short Term Memory

As an attempt to capture the business cycles of the economy I used a deep learning model. The family of Recurrent Neural Network (RNN) tends to be the best to use with time series, since it considers some time dependency in the estimation. All RNN consider hidden states between the regressors and the variable of interest.

A specific RNN model is the Long Short Term Memory (LSTM) proposed by Hochreiter and Schmidhuber (1997), which has a more robust structure. LSTM usage in time series is not very disseminated, being more known to be used in unstructured data application, as speech recognition. Following the arguments of Chen et al. (2019), LSTM should be a good fit to capture business cycles, since it is "designed to find hidden state processes allowing for lags of unknown and potentially long duration in the time series, which makes it well-suited to detect business cycles.". In appendix B, you can find the LSTM scheme that might give more intuition on this model.

4.4 Kalman Filter

As an attempt to use the concept of natural rate of unemployment, I tried to decompose the unemployment series in trend and cycle. The Kalman filter was made using an expanding window, so the estimation of trend and cycle could stay as close as possible in each iteration.

Most of the Kalman filters in unemployment rate in the literature uses quarterly or annual data. I followed Claar (2005), where he uses a state space structure made for annual frequency, and I tried to create an analogous monthly structure. The model used in this study is the following:

$$\begin{aligned} U_t &= U_t^{NAT} + \beta_t \\ U_t^{NAT} &= U_{t-12}^{NAT} + \varepsilon_t \\ \beta_t &= \rho\beta_{t-12} + \eta_t \end{aligned}$$

Since the Natural rate of unemployment follows a Random Walk, the forecast of the natural rate is equal the last observation of the time series.

The forecast of this method is the sum of the trend forecast and the cycle forecast. Since $\widehat{\Delta u_{t+h}^{NAT}} = 0$, the estimation of the change in unemployment rate is the forecast of the cycle minus the last observation of the cycle. The cycle forecast was made comparing all the previous methods, including X_t as explanatories. So, the forecast of this model was the following, where function G is each model showed in 4.1 to 4.3:

$$\beta_{t+h} = \hat{G}(X_t, \beta_t, \beta_{t-1}, \beta_{t-2}) + \varepsilon_{t+h}$$

$$\widehat{U_{t+h}} = \widehat{\beta_{t+h}} - \beta_t$$

In Figure 4.1 you can find the kalman filter made on the first window (first window ends in Dec/1999), where the red line is the trend, the blue line is the cycle and the black line is the actual unemployment. The best model was the AR, although all models applied to the trend performed poorly.

4.5 Adaptive Lasso

Adaptive Lasso is a shrinkage machine learning model with linear relation among dependent and independent variables. It is a two step estimator, where the first step is to choose the weight w_j that works as a penalty factor in the optimization below:

$$\widehat{\boldsymbol{\beta}}(\lambda) = \arg\min_{\beta \in \mathbb{R}^p} \sum_{t=1}^{T} \frac{1}{T} \left(y_t - \boldsymbol{\beta}' \boldsymbol{x}_t\right)^2 + \lambda \sum_{j=1}^{p} w_j \left|\beta_j\right|$$

The first step, i.e. initial estimator, in this study was a Lasso estimation that generated the $\tilde{\beta}_j$ parameter, using Bayesian Information Criterion ("bic"). Then, the penalty factor considered for the second step was $w_j = |\tilde{\beta}_j + \sqrt{T}|^{-1}$, where T is the quantity of independent variables in the regression.

4.6 AR with Random Forest in the Residuals

As already highlighted, the asymmetry and possible non linearities of the unemployment is well documented. In an attempt to disentangle the forecast in a linear and non linear term, I combined two methods (AR and Random Forest) so one could deal with the linear part, and the other with the non linear part. The forecast followed the steps below:

1. $$y_{t+h} = \hat{\alpha} + \hat{\beta}_0 y_t + \hat{\beta}_1 y_{t-1} + \hat{\beta}_2 y_{t-2} + \nu_{t+h}$$

2. $$\nu_{t+h} = \widehat{G}(X_t) + \varepsilon_{t+h}$$

3. $$\widehat{y_{t+h}} = \hat{\alpha} + \hat{\beta}_0 y_t + \hat{\beta}_1 y_{t-1} + \hat{\beta}_2 y_{t-2} + \widehat{G}(X_t)$$

Here, function G refers to Random Forest estimation. This model, as it will be better shown in the next section, had a good performance for employment forecast, but was not that good for unemployment forecast.

4.7 Adaptive Lasso with Random Forest in the Residuals

Similar to the model presented in section 4.6, I combined two machine learning models, now with Adaptive Lasso being the linear part of the estimation instead the autoregressive model. The idea was to separate a linear and a non linear part considering all the variables in the dataset, once the AR with RF in the Residuals considers only the autoregression as the linear part. Although more complex, it didn‘t showed an improve in performance if compared to the ARRF in employment forecast, and had very similar performance as the Random Forest in the unemployment forecast.

4.8 Employment Disaggregation

For the employment forecast I used a disaggregation to see if we could improve forecast accuracy. Once the ARRF was the model with most promising performance in the employment forecast, I forecasted each sector using ARRF. I didn’t analyzed the stationarity of each sector, once this could generate significant changes in predictive power treating each sector differently, and the comparison with all other models wouldn‘t be straightforward. For this reason, I treated each sector as the same way as the full employment (non stationary), taking the difference of the log level.

Each sector, considering the values of January 2000, has the following weight in the total employment (non farm payroll): mining and logging (0,45%); construction (5,12%); manufacturing (12,95%); trade, transportation and utilities (19,81%); information (2,79%); financial activities (5,91%); professional and business services (12,71%); education and health services (11,65%); leisure and hospitality (9,02%); other services(3,91%) and government (15,67%).

The aggregation was done considering the following process:

1. $$\Delta^h \widehat{ln(e_{t+h}^{sector}}) = \widehat{G}(\Delta^h ln(e_t^{sector}), \Delta^h ln(e_{t-1}^{sector}), \Delta^h ln(e_{t-2}^{sector}), X_t)$$

2. $$\widehat{e_{t+h}^{sector}} = exp(\Delta^h \widehat{ln(e_{t+h}^{sector}}))e_t^{sector}$$

3.
$$\widehat{y_t} = \Delta^h \widehat{ln(e_{t+h})} = ln\Big(\sum_{allsectors} \widehat{e_{t+h}^{sector}}\Big) - ln(e_t)$$

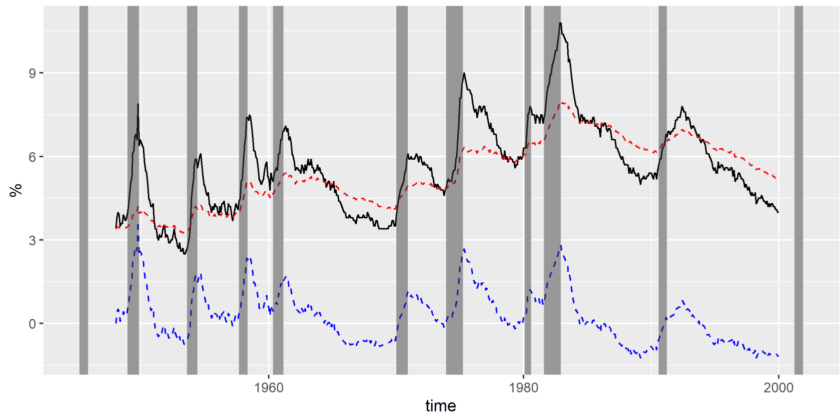


Figure 4.1: Kalman Filter to disentangle trend and cycle of the unemployment

5
Results

5.1
Overview

Following similar analysis made in Medeiros et al. (2019), the statistics for comparison were the mean squared error, mean absolute error and median absolute deviation (MSE, MAE and MAD, respectively). They are defined by:

$$MSE_{m,h} = \frac{1}{T - T_0 + 1} \sum_{t=T_0}^{T-h} \hat{e}^2_{t+h,m}$$

$$MAE_{m,h} = \frac{1}{T - T_0 + 1} \sum_{t=T_0}^{T-h} |\hat{e}_{t+h,m}|$$

$$MAD_{m,h} = median[|\hat{e}_{t+h,m} - median(\hat{e}_{t+h,m})|]$$

Where m refers to the model employed for the forecast, and h refers to the horizon to be forecasted (1, 3, 6, 12 or 24 months). MSE is the most common indicator for comparison in the forecasting literature. MAE and MAD are used so we could have different measures to validate the MSE analysis, and to have an idea of different variance of performance between the models.

For testing which model have better performance between the ones of this paper, I used the Diebold and Mariano (1995) test, using the MSE as the loss function. The test was applied using the $forecast$ package in R, and can only be done pairwise.[1]

In Tables 5.1 (unemployment) and 5.3 (employment), the models are compared relative to the AR benchmark, so the values shown are:

$$\frac{MSE_{m,h}}{MSE_{AR,h}}$$

Where m is the model mentioned in respective row, and h is the corresponding horizon mentioned in each column.

[1] In appendix B, the correlation between models in each forecast horizon is shown, in order to have an idea if there could be gains combining different forecasts.

5.2
Main Results - Unemployment

Autoregressive model with Random Forest in the Residual (ARRF) was the only model that had MSE and MAE lower than AR model in all horizons, although Random Forest (RF) had better performance than ARRF in those metrics in all but one horizon (12 months). If we go to Table 5.2 we see RF is better than ARRF with a confidence interval of 95% in 1 and 6 months of forecast horizon, and of 90% confidence for the 3 months horizon. This suggests RF is better than ARRF in shorter horizons. In longer horizons (12 and 24 months), we can argue it is equivalent, since ARRF is better at 12 months with a p-value of 83,4% (considering the alternative hypothesis RF is better than ARRF), and RF is better at 24 months with 98,6% of p-value. Adaptive Lasso (AdaLasso) had similar MSE values compared to RF, although only showed lower MSE than RF in the 12 month horizon. Adaptive Lasso with Random Forest in the residuals (AdaRF) had the lowest MSE for 3, 6 and 12 months horizon, but the Diebold Mariano test didn't showed any significance in favor of RF or AdaRF in any horizon. That means, both RF and AdaRF had similar performance considering all horizons.

The lower MSE in the 24 month horizon is from the Long Short Term Memory (LSTM) model. This is an example that analysing only MSE can be misleading, since the lowest MAE is from RF, and LSTM showed the highest MAD value in this horizon. If we check the Diebold-Mariano test, we see LSTM is not significantly better than any other model even if we consider 80% of confidence interval.

The kalman filter model does not stand out in any forecast horizon, so it didn't brought any benefit to the unemployment forecast exercise. Random Forest is significantly better than the kalman filter model, with a 90% confidence interval, in 4 of the 5 forecast horizons (12 month horizon is the exception).

5.3
Main Results - Employment

Differently than unemployment forecast, where RF and AdaRF where the models with best performance in almost all horizons, in employment forecast ARRF is the one that shows the best results overall.

In table 5.3 ARRF and ARRF made in each sector (referenced as "blocks" from now on.) are the only models to have lower MSE than AR in all horizons. RF fails to do so because of the 12 month forecast horizon, the same horizon that showed poor performance in the unemployment forecast. Between ARRF and disaggregated forecast, ARRF is better in longer horizons if we analyze the MSE performance. In MAE and MAD we see mix and quite similar results. ARRF had better MSE than RF in all horizons, although in MAE and MAD there was no explicit better model between both.

If we go to the Diebold Mariano test, in table 5.4, we see ARRF is better than RF with a 90% confidence in 2 horizons (3 and 12 months), and ARRF applied to each sector is better than RF with a 95% confidence in shorter horizons (1 and 3 months). When comparing to AdaRF, ARRF was better with a 90% confidence when forecasting one month ahead. This suggests ARRF (applied directly in employment, or disaggregated in 11 sectors) is the best model between the ones used to forecast employment growth.

LSTM didn‘t have a good performance in the employment forecast as well. It only had a p-value (considering alternative hypothesis LSTM is better forecaster than other model in a specific horizon prediction) greater than 50% in the 24 month horizon compared to the benchmark AR.

5.4 Main variables in the Machine Learning Models

The interpretation of each variable and its influence in the forecasting power in large datasets and machine learning models are not always straightforward. In order to find which variables are driving the results in the ML Models, namely, Random Forest and AdaLasso, the following strategies are taken. For the former, the importance is calculated randomly shufling the values of a variable of the out-of-bag sample, and calculating the decrease in accuracy for this specific variable. For the latter, since each series are standardized in the rolling windows, I simply compare the value of the parameters. In the models that uses Random Forest in the Residuals (ARRF and AdaRF) the analysis in the Random Forest part is the predictive power forecasting the residual. So, in the model AdaRF, the importance of each variable was separated in two analysis, being the AdaLasso part the same analysis as the the Model AdaLasso alone (Figures 5.5 amd 5.6), and the Random Forest importance considering the additional forecasting power on the residual (Figures 5.3 and 5.4 for the ARRF, and 5.7 and 5.8 for the AdaRF).

Once there are 128 variables, I used the classification suggested by McCracken and Ng (2016) to have a better interpretation of the results. In their paper they classify each variable within 8 different categories. The 8 categories are (1) Output and Income, (2) Labor Market, (3) Housing, (4) Consumption, (5) Orders and Inventories, (6) Money and Credit and (8) Stock market. Additionally to this groups, I added a group for the autoregressive part, following Medeiros et al. (2019). In the unemployment analysis, I considered the level of unemployment as part of the autoregressive part.

In almost all models, the results showed the longer the horizon, more important variables in the Interest Group where. The same thing happened with the Output group, but in a minor scale. This happened both in the Unemployment and Employment forecasts. ARRF forecasting employment is the only exception, where variables classified as Interest had an important role in shorter horizon forecasts.

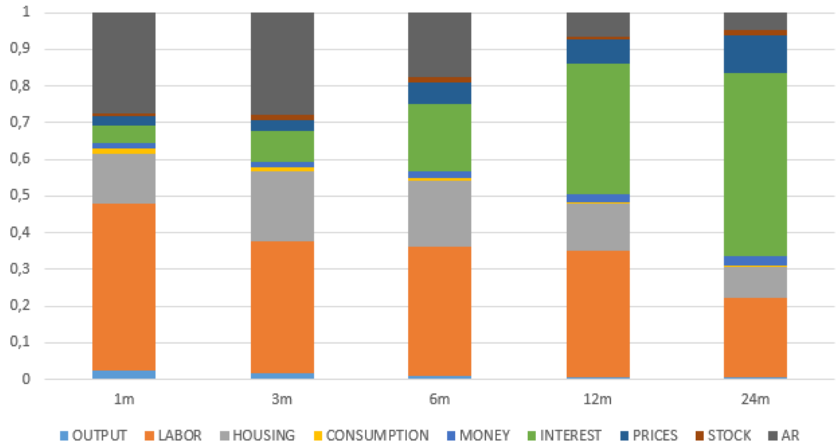


Figure 5.1: Variable Importance: Random Forest (Employment)

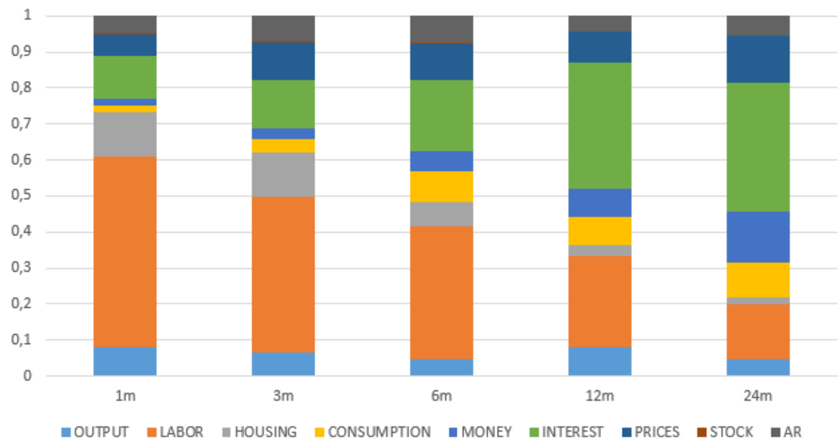


Figure 5.2: Variable Importance: Random Forest (Unemployment)

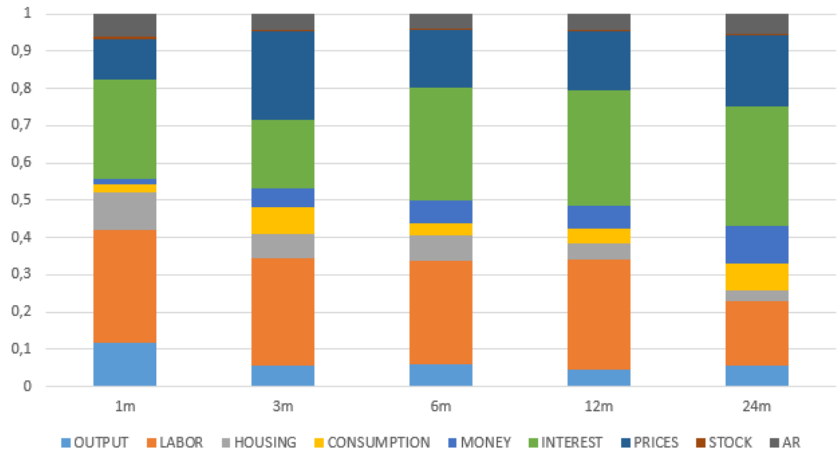


Figure 5.3: Variable Importance: ARRF (Employment)

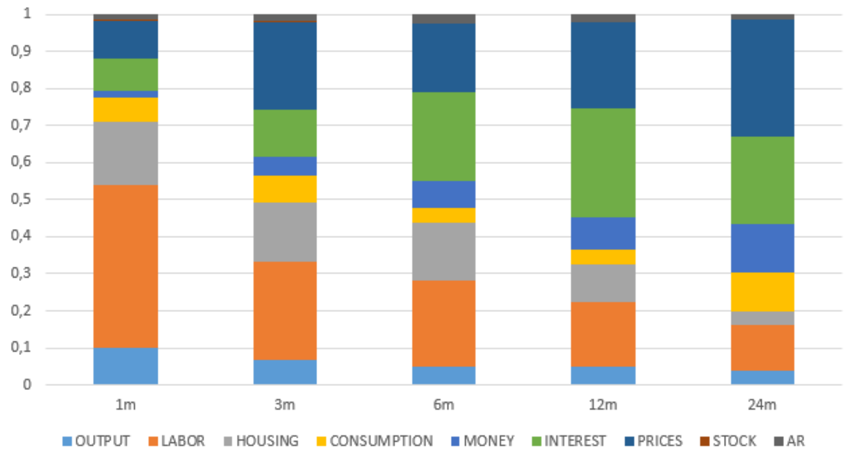


Figure 5.4: Variable Importance: ARRF (Unemployment)

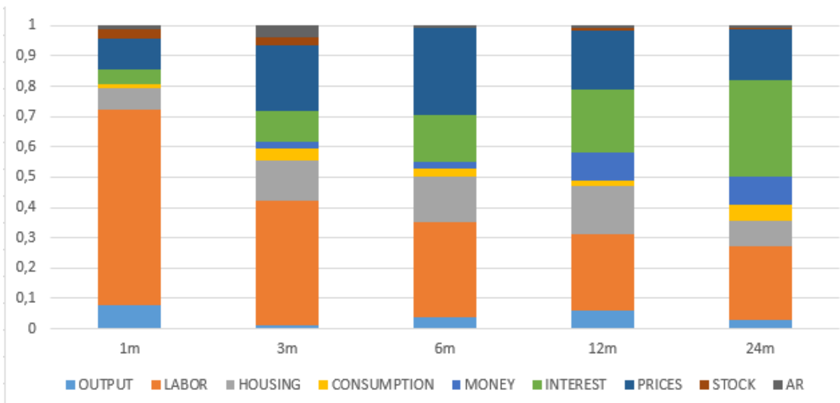


Figure 5.5: Variable Importance in AdaLasso (Unemployment)

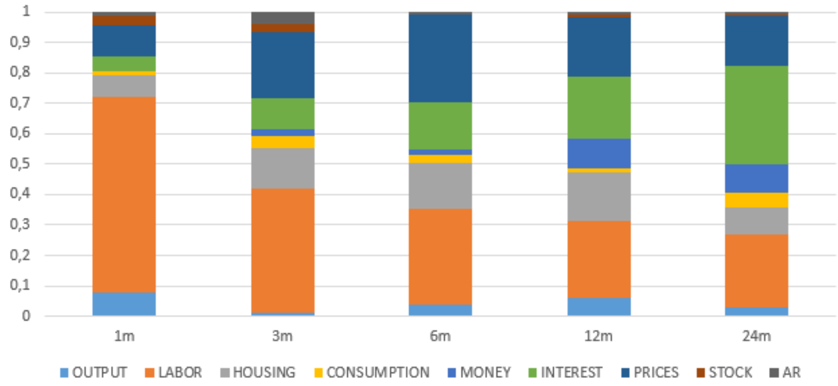


Figure 5.6: Variable Importance in AdaLasso (Employment)

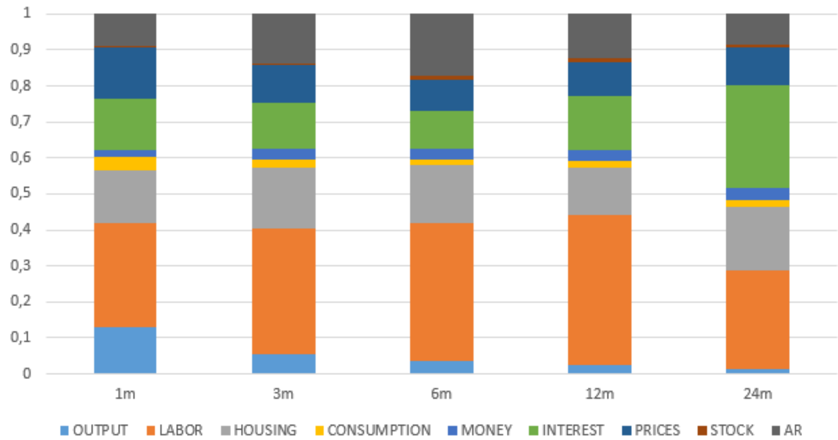


Figure 5.7: Variable Importance in AdaRF - Random Forest (Employment)

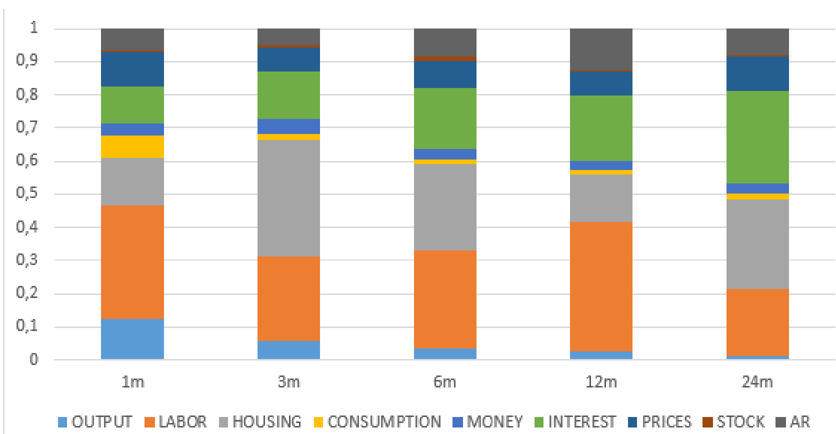


Figure 5.8: Variable Importance in AdaRF - Random Forest (Unemployment)

Table 5.1: Comparison between Models - Unemployment

MSE comparison					
	forecast horizon (in months)				
Models	1	3	6	12	24
AR	1	1	1	1	1
RF	0.823	0.734	0.721	1.097	0.834
LSTM	1.167	1.535	1.54	1.272	0.778
ARRF	0.95	0.818	0.846	0.897	0.886
AR kalman	1.048	1.115	1.219	1.127	1.162
Adalasso	0.853	0.734	0.73	0.872	1.126
AdaRF	0.839	0.678	0.651	0.837	1.08
MAE comparison					
	forecast horizon (in months)				
Models	1	3	6	12	24
AR	1	1	1	1	1
RF	0.9	0.857	0.793	0.998	0.797
LSTM	1.072	1.143	1.208	1.114	0.86
ARRF	0.952	0.912	0.934	0.944	0.92
AR kalman	1.01	1.014	1.09	1.023	1.022
Adalasso	0.93	0.831	0.893	1.003	1.065
AdaRF	0.908	0.797	0.84	0.948	1.062
MAD comparison					
	forecast horizon (in months)				
Models	1	3	6	12	24
AR	1	1	1	1	1
RF	0.778	0.999	1.098	1.186	0.814
LSTM	0.809	0.981	1.235	1.355	1.689
ARRF	0.9	0.763	1.111	0.771	0.993
AR kalman	0.943	0.858	0.885	0.83	0.93
Adalasso	0.786	0.971	0.791	0.667	0.834
AdaRF	0.754	0.931	0.847	0.588	0.813

Table 5.2: Diebold Mariano test between Models - Unemployment

Diebold Mariano test (1 month)							
Models	AR	RF	LSTM	ARRF	AR-Kalman	AdaLasso	AdaRF
AR	-	0	0.989	0.181	0.959	0.003	0.005
RF	1	-	1	1	0.999	0.869	0.803
LSTM	0.011	0	-	0.007	0.049	0	0
ARRF	0.819	0	0.993	-	0.917	0.007	0.001
AR-Kalman	0.041	0.001	0.951	0.083	-	0.001	0.003
Adalasso	0.997	0.131	1	0.993	0.999	-	0.284
AdaRF	0.995	0.197	1	0.999	0.997	0.716	-
Diebold Mariano test (3 month)							
Models	AR	RF	LSTM	ARRF	AR-Kalman	AdaLasso	AdaRF
AR	-	0.02	0.989	0.114	0.864	0.08	0.074
RF	0.98	-	0.995	0.914	0.961	0.54	0.333
LSTM	0.011	0.005	-	0.017	0.006	0.013	0.014
ARRF	0.886	0.086	0.983	-	0.888	0.145	0.072
AR-Kalman	0.136	0.039	0.994	0.112	-	0.081	0.08
Adalasso	0.92	0.46	0.987	0.855	0.919	-	0.109
AdaRF	0.926	0.667	0.986	0.928	0.92	0.891	-
Diebold Mariano test (6 month)							
Models	AR	RF	LSTM	ARRF	AR-Kalman	AdaLasso	AdaRF
AR	-	0.007	0.955	0.196	0.933	0.169	0.129
RF	0.993	-	0.98	0.953	0.978	0.543	0.381
LSTM	0.045	0.02	-	0.052	0.072	0.077	0.067
ARRF	0.804	0.047	0.948	-	0.89	0.218	0.121
AR-Kalman	0.067	0.022	0.928	0.11	-	0.118	0.1
Adalasso	0.831	0.457	0.923	0.782	0.882	-	0.072
AdaRF	0.871	0.619	0.933	0.879	0.9	0.928	-
Diebold Mariano test (12 month)							
Models	AR	RF	LSTM	ARRF	AR-Kalman	AdaLasso	AdaRF
AR	-	0.712	0.888	0.315	0.81	0.376	0.336
RF	0.288	-	0.971	0.166	0.557	0.257	0.207
LSTM	0.112	0.029	-	0.041	0.295	0.092	0.064
ARRF	0.685	0.834	0.959	-	0.794	0.446	0.385
AR-Kalman	0.19	0.443	0.705	0.206	-	0.298	0.276
Adalasso	0.624	0.743	0.908	0.554	0.702	-	0.312
AdaRF	0.664	0.793	0.936	0.615	0.724	0.688	-
Diebold Mariano test (24 month)							
Models	AR	RF	LSTM	ARRF	AR-Kalman	AdaLasso	AdaRF
AR	-	0.154	0.242	0.345	0.806	0.896	0.777
RF	0.846	-	0.428	0.876	0.936	0.962	0.896
LSTM	0.758	0.572	-	0.744	0.799	0.913	0.934
ARRF	0.655	0.124	0.256	-	0.811	0.928	0.842
AR-Kalman	0.194	0.064	0.201	0.189	-	0.621	0.521
Adalasso	0.104	0.038	0.087	0.072	0.379	-	0.33
AdaRF	0.223	0.104	0.066	0.158	0.479	0.67	-

Notes: The results were rounded to nearest thousandth, so p-values that are greater than 0,9995 appears as 1. The p-value is considering the alternative hypothesis the corresponding model in column is better than the corresponding model in row.

Table 5.3: Comparison between Models - Employment

MSE comparison					
	forecast horizon (in months)				
Models	1	3	6	12	24
AR	1	1	1	1	1
RF	0.889	0.971	0.9	1.163	0.764
LSTM	2.778	3.159	2.447	1.372	0.618
ARRF	0.889	0.754	0.842	0.721	0.569
blocks	0.778	0.754	0.963	0.881	0.807
AdaLasso	1.111	1.072	1.069	0.895	0.636
AdaRF	0.889	0.884	0.857	0.763	0.598
MAE comparison					
	forecast horizon (in months)				
Models	1	3	6	12	24
AR	1	1	1	1	1
RF	0.972	0.994	1.016	1.114	0.796
LSTM	1.486	1.606	1.555	1.27	0.737
ARRF	0.931	0.9	0.967	1.02	0.852
blocks	0.917	0.906	1.011	0.949	0.845
AdaLasso	1.083	1.122	1.235	1.119	0.782
AdaRF	1	1.011	1.101	1.023	0.774
MAD comparison					
	forecast horizon (in months)				
Models	1	3	6	12	24
AR	1	1	1	1	1
RF	0.982	0.832	1.037	0.961	0.398
LSTM	1.4	1.252	1.378	1.483	0.605
ARRF	0.927	0.863	0.931	1.192	0.973
blocks	0.964	0.802	0.986	0.823	0.7
AdaLasso	1.091	1.031	1.493	1.301	0.777
AdaRF	0.982	0.954	1.447	1.377	0.756

Table 5.4: Diebold Mariano test between Models - Employment

Diebold Mariano test (1 month)							
Models	AR	RF	LSTM	ARRF	blocks	AdaLasso	AdaRF
AR	-	0.231	1	0.009	0.016	0.913	0.324
RF	0.769	-	1	0.112	0.038	0.986	0.536
LSTM	0	0	-	0	0	0.001	0
ARRF	0.991	0.888	1	-	0.383	0.999	0.914
blocks	0.984	0.962	1	0.617	-	1	0.962
Adalasso	0.087	0.014	0.999	0.001	0	-	0
AdaRF	0.676	0.464	1	0.086	0.038	1	-
Diebold Mariano test (3 month)							
Models	AR	RF	LSTM	ARRF	blocks	AdaLasso	AdaRF
AR	-	0.43	0.967	0.05	0.016	0.615	0.336
RF	0.57	-	0.966	0.059	0.045	0.673	0.344
LSTM	0.033	0.034	-	0.03	0.027	0.057	0.046
ARRF	0.95	0.941	0.97	-	0.496	0.98	0.807
blocks	0.984	0.955	0.973	0.504	-	0.978	0.777
Adalasso	0.385	0.327	0.943	0.02	0.022	-	0.003
AdaRF	0.664	0.656	0.954	0.193	0.223	0.997	-
Diebold Mariano test (6 month)							
Models	AR	RF	LSTM	ARRF	blocks	AdaLasso	AdaRF
AR	-	0.27	0.947	0.157	0.34	0.584	0.352
RF	0.73	-	0.948	0.224	0.767	0.769	0.445
LSTM	0.053	0.052	-	0.048	0.051	0.114	0.096
ARRF	0.843	0.776	0.952	-	0.86	0.845	0.523
blocks	0.66	0.233	0.949	0.14	-	0.651	0.38
Adalasso	0.416	0.231	0.886	0.155	0.349	-	0.018
AdaRF	0.648	0.555	0.904	0.477	0.62	0.982	-
Diebold Mariano test (12 month)							
Models	AR	RF	LSTM	ARRF	blocks	AdaLasso	AdaRF
AR	-	0.773	0.87	0.177	0.251	0.397	0.284
RF	0.227	-	0.761	0.096	0.07	0.244	0.173
LSTM	0.13	0.239	-	0.098	0.121	0.225	0.173
ARRF	0.823	0.904	0.902	-	0.793	0.795	0.592
blocks	0.749	0.93	0.879	0.207	-	0.523	0.332
Adalasso	0.603	0.756	0.775	0.205	0.477	-	0.049
AdaRF	0.716	0.827	0.827	0.408	0.668	0.951	-
Diebold Mariano test (24 month)							
Models	AR	RF	LSTM	ARRF	blocks	AdaLasso	AdaRF
AR	-	0.208	0.051	0.142	0.235	0.102	0.123
RF	0.792	-	0.318	0.256	0.975	0	0.142
LSTM	0.949	0.682	-	0.414	0.74	0.534	0.47
ARRF	0.858	0.744	0.586	-	0.788	0.647	0.572
blocks	0.765	0.025	0.26	0.212	-	0.101	0.099
Adalasso	0.898	0.825	0.466	0.353	0.899	-	0.269
AdaRF	0.877	0.858	0.53	0.428	0.901	0.731	-

Notes: The results were rounded to nearest thousandth, so p-values that are greater than 0,9995 appears as 1. The p-value is considering the alternative hypothesis the corresponding model in column is better than the corresponding model in row.

6
Conclusion

This study suggests the analysis made in the unemployment rate forecast shouldn‘t be directly applied to the employment growth forecast. The results found in unemployment ratifies the results found in Coulombe et al. (2019), where Random Forest is among the best models. A combination of methods in order to have a linear and non linear estimation (AR or AdaLasso as the linear part with Random forest applied in residuals), to the best of my knowledge, was never proposed in the forecasting literature and its performance in the employment forecast is promising.

In this study, LSTM showed poor performance in both variables predictions. Deep learning models are very unstable and sensitive to hyperparameters, and this poor performance doesn‘t mean it is not useful. More complex neural networks could generate better results, being LSTM only a part (one layer of multiple layers) of the estimation, for example.

Bibliography

Atkeson, A., Ohanian, L. E., et al. (2001). Are phillips curves useful for forecasting inflation? *Federal Reserve bank of Minneapolis quarterly review*, 25(1):2–11.

Burns, A. F. and Mitchell, W. C. (1946). *The Basic Measures of Cyclical Behavior*, pages 115–202. NBER.

Chen, L., Pelger, M., and Zhu, J. (2019). Deep learning in asset pricing. *Available at SSRN 3350138*.

Chollet, F. and Allaire, J. J. (2018). *Deep Learning with R*. Manning Publications Co., USA, 1st edition.

Claar, V. V. (2005). A kalman-filter approach to estimating the natural rate of unemployment. *Zbornik radova Ekonomskog fakulteta u Rijeci, časopis za ekonomsku teoriju i praksu-Proceedings of Rijeka Faculty of Economics, Journal of Economics and Business*, 23(1):1–24.

Cook, T. and Hall, A. S. (2017). Macroeconomic indicator forecasting with deep neural networks. *Federal Reserve Bank of Kansas City, Research Working Paper*, (17-11).

Coulombe, P. G., Leroux, M., Stevanovic, D., Surprenant, S., et al. (2019). How is machine learning useful for macroeconomic forecasting? Technical report, CIRANO.

Diebold, F. and Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–63.

Hall, A. S. (2018). Machine learning approaches to macroeconomic forecasting. *The Federal Reserve Bank of Kansas City Economic Review*, 103(63):2.

Hall, R., Feldstein, M., Bernanke, B., Frankel, J., Gordon, R., and Zarnowitz, V. (2001). The business-cycle peak of march 2001. *Business Cycle Dating Committee, National Bureau of Economic Research*.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

Issler, J. V. and Vahid, F. (2006). The missing link: Using the nber recession indicator to construct coincident and leading indices of economic activity. *Journal of Econometrics*, 132(1):281–303.

McCracken, M. W. and Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.

Medeiros, M. C., Vasconcelos, G. F., Veiga, Á., and Zilberman, E. (2019). Forecasting inflation in a data-rich environment: the benefits of machine learning methods. *Journal of Business & Economic Statistics*, pages 1–22.

Mitchell, W. C. (1927). *The Contribution of Statistics*, pages 189–360. NBER.

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

Neftci, S. N. (1984). Are economic time series asymmetric over the business cycle? *Journal of Political Economy*, 92(2):307–328.

Proietti, T. (2003). Forecasting the us unemployment rate. *Computational Statistics & Data Analysis*, 42(3):451–476.

Rapach, D. E. and Strauss, J. K. (2008). Forecasting us employment growth using forecast combining methods. *Journal of Forecasting*, 27(1):75–93.

Roberts, J. M. and Morin, N. J. (1999). Is hysteresis important for us unemployment?

Skalin, J. and Teräsvirta, T. (2002). Modeling asymmetries and moving equilibria in unemployment rates. *Macroeconomic Dynamics*, 6(2):202–241.

Smeekes, S. and Wijler, E. (2018). Macroeconomic forecasting using penalized regression methods. *International journal of forecasting*, 34(3):408–430.

Stock, J. H. and Watson, M. W. (1989). New indexes of coincident and leading economic indicators. *NBER macroeconomics annual*, 4:351–394.

Stock, J. H. and Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics*, 14(1):11–30.

Stock, J. H. and Watson, M. W. (1998). Diffusion indexes. Technical report, National bureau of economic research.

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

Stock, J. H. and Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460):1167–1179.

Stock, J. H. and Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

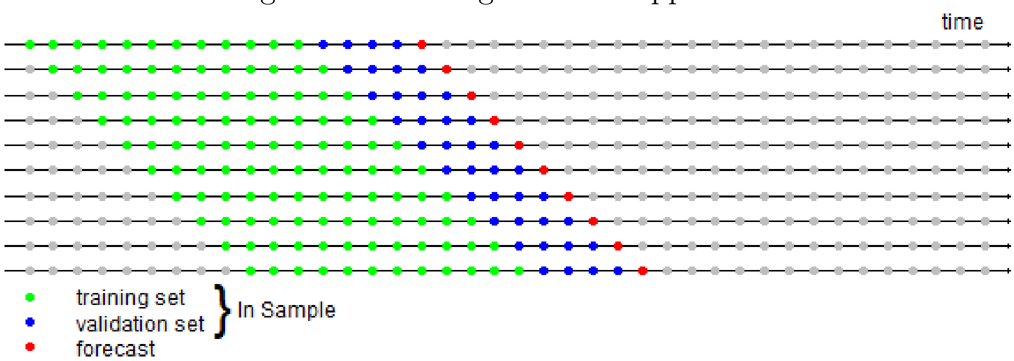
A
Long Short Term Memory

The LSTM model is estimated using a training set, a validation set and a test set. The test set in the case of this study is the 5 forecast horizons (1, 3, 6, 12 and 24 months). The training set and validation set has to be within the in sample set. There is no optimum size (or size ratio) between the validation and training set. Chollet and Allaire (2018) suggests the paddern is 80% of the in Sample for training and 20% for the validation set.

I used in the LSTM model 25% of the In Sample for the validation set and 75% for the training set. This should be around the upper bound size of the validation set, without drastically compromising the estimation of the parameters. Other important issue is that both validation and training set should cover at least one entire business cycle. if we did a 30 year rolling window, the validation set would have only 6 years window, and the validation would be problematic. As an idea, the gap between the peak and the valley in the unemployment rate in the business cycle of the end of the century, is 7 years and 10 months (7,8% in June 1992 and 3,8% in april 2000), the gap between the beginnings of the 1991 NBER recession and 2001 NBER recession is 10 years and 8 months.

For this reason, I used a 40 year rolling window, with a 10 year validation block, so it could cover more than just the expansion of the economy and, most of the time, getting a full business cycle. The rolling window scheme for the Neural Network estimation follows the dynamics of the graph below.

Figure A.1: Rolling Window Approach



Below is the structure of the LSTM, where uppercase W's are the weight matrixes, the lowercase w's are the bias (similar to an intercept in simple regressions), σ refers to the sigmoid function, commonly used in neural networks systems, and tanh refers to hyperbolic tangent:

$$
\begin{aligned}
\text{input}_t &= \sigma\left(\underset{(4x4)}{W_h^{(i)}} h_{t-1} + W_x^{(i)} x_t + \underset{(4x1)}{w_0^{(i)}}\right)\\
\text{forget}_t &= \sigma\left(W_h^{(f)} h_{t-1} + W_x^{(f)} x_t + w_0^{(f)}\right)\\
\text{out}_t &= \sigma\left(W_h^{(o)} h_{t-1} + W_x^{(o)} x_t + w_0^{(o)}\right)\\
\tilde{c}_t &= \tanh\left(W_h^{(c)} h_{t-1} + W_x^{(c)} x_t + w_0^{(c)}\right)\\
c_t &= \text{ forget }_t \circ c_{t-1} + \text{ input }_t \circ \tilde{c}_t\\
h_t &= \text{ out }_t \circ \tanh\left(c_t\right)\\
y_{t+h} &= W_h^{(y)} h_t + w_0^{(y)}
\end{aligned}
\tag{A-1}
$$

You can see in the last equation of the system that the variable of interest y_{t+h} is a linear function of the hidden states of the economy. The idea is that LSTM can reduce the dimensionality of the explaining variables, serving as a factor model with the feature of not having a fixed temporal dependency. Also we include only one lag of the explanatory variables, since the effect of lagged variables should be captured in the cells.

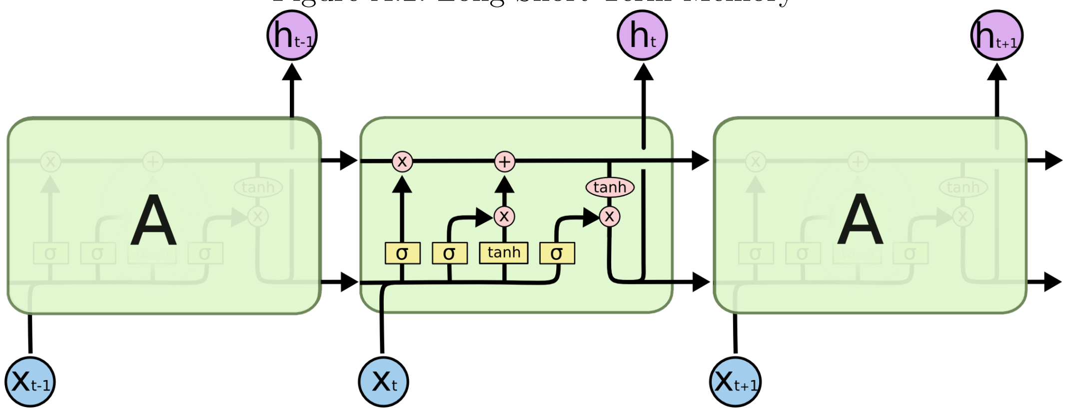
Taking first or seasonal differences here shouldn't matter, since the optimization process taking different temporal dependencies should consider in the weight matrixes and hidden cells the impact of longer horizons taking care of the seasonal dynamics of the economy.

The figure below [1] is a scheme that should help understand the LSTM network. Each block A is a period in time. Within each period in time, there are gates that transform the inputs into h_t and c_t. There are two arrows from one period representing the hidden states h_t, and information that was not used into the construction of the hidden states c_t.

The inputs in each period are the explanatory variables x_t (here the autoregressive part is considered in x_t), the hidden state h_{t-1} from the previous period, and the the result of the forget gate c_t.

[1]taken from http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Figure A.2: Long Short Term Memory



B
Correlograms

Below you can find the correlograms between the forecast models, for unemployment and employment.

In longer horizons, the correlation bewteen models were weaker than shorter horizons, so combining forecasts could improve the predictive power.

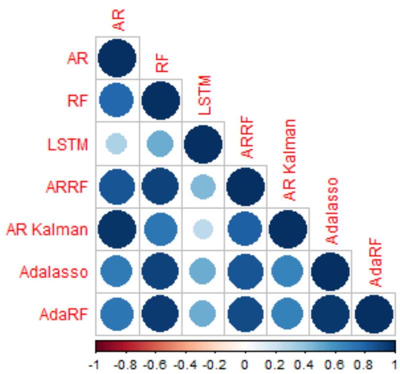


Figure B.1: 1 month horizon - Correlogram Unemployment Forecast

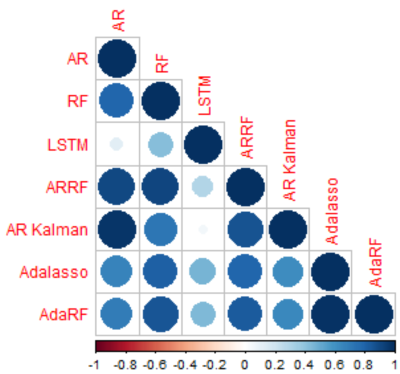


Figure B.2: 3 month horizon - Correlogram Unemployment Forecast

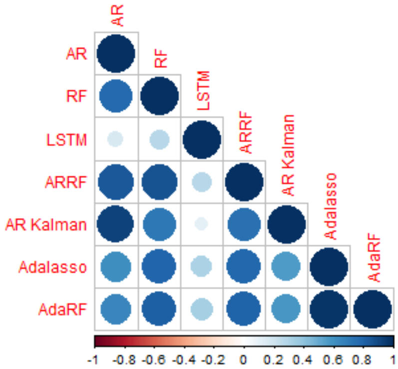


Figure B.3: 6 month horizon - Correlogram Unemployment Forecast

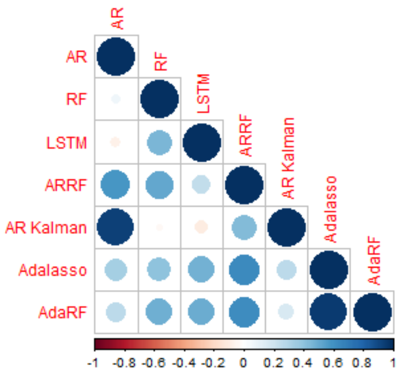


Figure B.4: 12 month horizon - Correlogram Unemployment Forecast

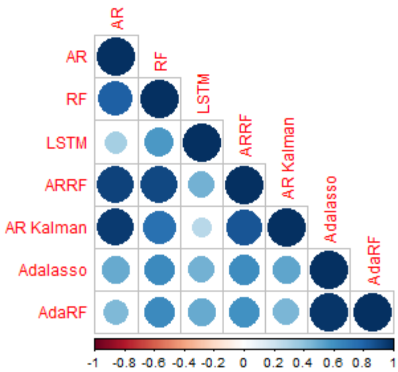


Figure B.5: 24 month horizon - Correlogram Unemployment Forecast

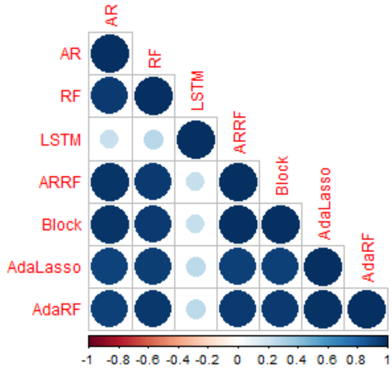


Figure B.6: 1 month horizon - Correlogram Employment Forecast

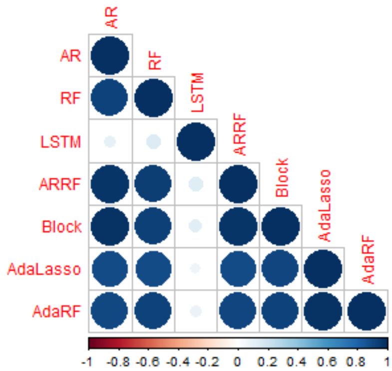


Figure B.7: 3 month horizon - Correlogram Employment Forecast

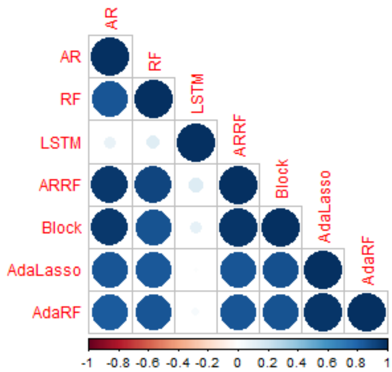


Figure B.8: 6 month horizon - Correlogram Employment Forecast

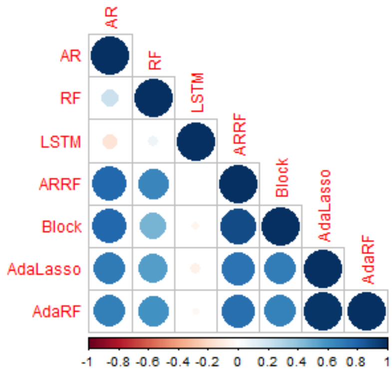


Figure B.9: 12 month horizon - Correlogram Employment Forecast

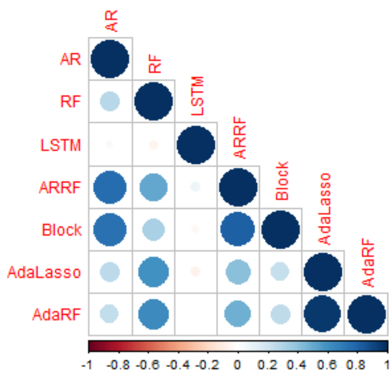


Figure B.10: 24 month horizon - Correlogram Employment Forecast