Using Mixed Frequency Data to Forecast Recessions and GDP

Ryan Mitchell

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:
Yu-chin Chen, Chair
Eric Zivot
Philip Brock

Program Authorized to Offer Degree:
Economics

University of Washington

Abstract

Using Mixed Frequency Data to Forecast Recessions and GDP

Ryan Mitchell

Chair of the Supervisory Committee:
Professor Yu-chin Chen
Department of Economics

This dissertation studies how and if mixed frequency time series data should be used to forecast recessions in the US and Canada, as well as GDP in a selection of OECD countries.

The first chapter combines daily and weekly financial data with monthly macroeconomic indicators in a mixed frequency probit (MFP) regression to forecast and nowcast US and Canadian recessions. This chapter adds to the existing literature in multiple ways, including; developing a mixed frequency binary model that could be helpful for topics outside of recession forecasting, examining how higher frequency data should be weighted in the context of recession prediction, as well as a methodology that can nowcast current economic conditions. Overall I find significant improvements in the forecasting and nowcasting accuracy of recessions when using mixed frequency data, compared to a benchmark model that aggregates data into the same frequency.

The second chapter extends from the first to apply machine learning techniques to the same problem. I add to the existing literature by incorporating mixed frequency data directly into a classification artificial neural network (MF-ANN) as well as using novel cross validation methods to tune hyperparameters and carry out feature selection with time series data. Overall when comparing US recession forecasting results to the reduced form methodology of Chapter 1, I find mixed results. While some metrics indicate similar performance between

the two methods, the ANN makes less extreme forecasting errors on average.

The third chapter uses a seemingly unrelated regressions (SUR) approach with a mixed frequency framework to forecast GDP of 10 OECD countries. This chapter adds to the literature as a way to efficiently include cross country information in GDP forecasting equations, as well as being an effective methodology when the researcher is constrained by small sample sizes. Overall we find that SUR outperforms OLS for the majority of countries and forecasting horizons, however as sample sizes are extended this benefit is reduced.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my advisor, Professor Yu-chin Chen, for her continued support throughout the dissertation process. Her feedback and advice has been a key driver in helping me shape my way of thinking and progress my research into something that adds to the existing literature. I am also grateful to my committee members Professor Eric Zivot and Professor Philip Brocks, as well as my GSR, Professor Thomas Gilbert, for their help and extremely useful feedback over the last few years of my PhD.

Can Daily Financial Data Help Forecast Economic Downturns?

1 Chapter 1

1.1 Introduction

Predicting when a recession will occur is of great importance and interest to policy makers, such as the government and Federal Reserve, as well as private economic agents who have vested interests in the direction of economic activity, whether it be for personal wealth reasons or general job market opportunities. Thousands of newspapers, newsletters, television shows, blogs and many more, constantly report on these changing economic conditions. However, predicting the turning point of business cycle phases in real time (constantly) is difficult as business conditions are not observable in real time. The Business Cycle Committee of the NBER, who decides when the US economy is in a recession, will make an announcement of when a recession began long after the fact. For example, the US recession that began in December 2007 was not officially announced until December 2008. This is a common occurrence, and over the past 30 years the NBER has made its announcements 6 to 20 months after the corresponding peak or trough in economic activity, Fossati (2015). A natural country of comparison is Canada who, similarly, use a committee called the Business Cycle Council of the C.D. Howe Institute, to determine recession periods. They also announce actual recession periods at a considerable delay. This makes it of high interest to predict if we are in or about to enter a recession, as it may not be obvious from key macroeconomic variables, and the business cycle dating committees of the US and Canada will not announce a recession to the public until long after it has begun.

The common strategy to model business conditions in real time is by generating recession probabilities using binary models, where both the recession indicator and explanatory data are recorded at the same frequency - monthly. This paper adds to the literature in multiple ways: (1) I use NBER and C.D. Howe Institute defined recession indicators as a binary dependent variable in a mixed frequency probit model (MFP), taking advantage of the readily available daily, weekly and monthly financial and macroeconomic data. I then forecast recession probabilities up to a 3 month horizon. (2) I introduce a new method to include mixed frequency data in a binary model that could be useful for topics outside of recession prediction. (3) I subset variables into different asset markets and carry out factor analysis. This allows me to see what asset markets are key leading indicators of US and Canadian recessions. (4) Part of my analysis focuses on recession 'onset' prediction, defined later as the first 5 months of a recession. Arguably this is the most important time to accurately predict a recession due to the considerable reporting lag by the respective business cycle committees. (5) The mixed frequency framework allows me to update, or 'Nowcast', current recession probabilities as frequently as daily. Hence, providing the public with useful and constant information about current business conditions.

There has been much previous research on forecasting US recessions with matched frequency data. Popular techniques include using binary and Markov switching models. Fossati (2015) used dynamic factors estimated from panels of macroeconomic and financial indicators to predict future recessions using probit models, concluding models that include the 3-month less 10 year term spread, a stock market dynamic factor and a real economic factor achieve the best out of sample fit. Huang and Startz (2020) find that augmenting existing Markov-switching dynamic factor models with additional information on the stock return volatility improves prediction of the state of the economy. Literature on forecasting recessions with mixed frequency data is more limited, and focuses on identifying recessions using GDP turning points instead of the recession indicator directly. One example is Balcilar (2016) who found including the monthly economic policy uncertainty index improves forecasting of quarterly GDP turning points in a mixed frequency Markov switching vector autoregressive model, compared to aggregation of data into quarterly frequency.

Comparatively there has been little work focusing on Canada. Using a matched frequency probit model with dynamic factors, Fossati (2018) found that Canadian real activity factors are particularly successful at predicting recessions at short horizons, and that Canadian bond and exchange rate factors improve recession forecasts at longer horizons. I find no papers directly predicting Canadian recessions using mixed frequency. A more extensive literature review of both the US and Canada can be found in section 1.2.

In this paper I follow Ludvigson and Ng (2009) and Fossati (2015, 2018) in estimating factors representing the bond, stock, exchange rate, commodity and real activity markets from panels of macroeconomic and financial data. These daily, weekly and monthly factors are then used to predict future US and Canadian recession dates, as published by the NBER and the Business Cycle Council of the C.D. Howe Institute respectively, in a mixed frequency probit model (MFP). My main findings show that when daily financial data is included in the forecasting model for the US, out of sample predictive performance improves at the 1,2 and 3 month forecasting horizons. Reductions of up to 17% in the quadratic probability score (QPS) are found, depending on the forecast horizon, compared to aggregating data at the monthly frequency. When focusing on recession onset prediction these improvements increase up to 30%. However, for Canada, where only weekly financial data for the whole sample is available, there are mixed results. I find improved recession onset prediction performance at all forecast horizons, with up to a 22% decrease in the QPS. But evaluation statistics covering the full out of sample period show no improvement in predictive power when compared to probit models that aggregate financial data at the monthly frequency. Although I use revised macroeconomic data for both US and Canada, as opposed to real time vintage data, this paper seeks to emphasize the benefit and use of mixed frequency forecasting approaches over matched frequency approaches. More information can be found for my choice to use revised data in Section 1.5.

Additionally, I use the benefit of daily and weekly frequency data to nowcast the current months recession probabilities, updating forecasts on a daily and weekly basis in the US and Canada respectively. By including the current months financial data into the MFP I find that I can improve forecasting performance measured by the QPS in the US by up to 14%, compared to a 1 month ahead forecast. The Diebold Mariano test also shows statistically significant improvements in forecasting performance at the 10% level, depending how many days of the current months financial data are included in the model. However, there are no significant improvements from nowcasting using financial data in Canada.

Finally, I find that daily and weekly bond market factors are key leading indicators, especially at detecting the onset of a recession, in the US and Canada respectively. The stock and real activity market also play important roles in improving forecasting performance. However, this is only when the stock market data is aggregated at the monthly frequency, due to daily data causing volatile results. Exchange rate and commodity market data at any frequency is not useful in predicting future recessions in the US and Canada.

1.2 Literature Review

1.2.1 Matched Frequency

Estrella and Mishkin (1998) find that the 3-month less 10-year term spread and stock price indices are the most useful predictors of future US recessions. Similarly, Wright (2006) finds that using the level of the federal funds rate together with the term spread improves the performance of the predictive probit models. Katayama (2009) analyzed the performance of several binary class models for NBER recessions using combinations of 33 macroeconomic and financial indicators. He finds that the combination of the term spread, month to month changes in the SP 500 index and the growth rate of non-farm employment generate the sequence of out of sample recession probabilities that better fits NBER recession dates. Fossati (2015) uses dynamic factors estimated from panels of macroeconomic and financial indicators to predict future recessions using probit models. He concludes that at 3 month forecast horizons probit models that include the 3-month less 10 year term spread, a stock market dynamic factor and a real economic factor achieve the best out of sample fit. More recently, Huang and Startz (2020) find that augmenting existing Markov-switching dynamic factor models with additional information on the stock return volatility improves prediction of the state of the economy. They beat both the peak and trough announcements for recent recessions by the NBER by several months.

Comparatively there has been little work focusing on Canada. Atta-Mensah and Tkacz (1998) find that that the Canadian yield spread, defined as the difference between long term bond yields and the 3-month commercial paper rate, is the most useful indicator to predict recessions in Canada. Bernard and Gerlach (1998) show that when the US yield spread is included as a predictor, recession forecasts improve at medium and long term horizons. Hao and Ng (2011) include inflation, employment, monthly GDP, housing starts and a number of other financial and real activity indicators in a dynamic probit model to forecast Canadian recessions. Recently, the interest has moved to factor-augmented models, for example Fossati (2018). He uses bond and exchange rate, stock and real activity dynamic factors from Canada and the US to forecast Canadian recessions in a probit model at various horizons. Fossati finds that Canadian real activity factors are particularly successful at predicting turning points at short horizons, and that Canadian bond and exchange rate factors improve recession forecasts at 6 to 12 month horizons. Exclusion of US data results in no significant deterioration in predictive performance, but US interest rate spreads are part of the best performing model at longer forecast horizons.

1.2.2 Mixed Frequency

The literature on forecasting recessions with mixed frequency data is much more limited. I begin by looking at research that has focused on the US. Balcilar (2016) use a mixed frequency Markov switching vector autoregressive (MF-MS-VAR) model to predict regimes in quarterly US GDP using the monthly economic policy uncertainty index as the leading indicator. They find that their model performs better out of sample with mixed frequency data compared to that of matched frequency. Bessec (2015) uses a Markov switching factor mixed data sampling (MS-factor MIDAS) model to extract probabilities in turning points of quarterly US GDP using monthly financial variables. They also conclude that economic turning points are detected more successfully with mixed frequency models versus models that aggregate the higher frequency data.

Next I focus on the Europe. Camacho (2014) use a Markov switching dynamic factor model to forecast quarterly GDP of the euro area, using a mix of quarterly and monthly variables. Filtered probabilities are then extracted and updated whenever new data is released to forecast economic turning points in the euro area in real time. Bessec (2015) uses monthly bond market, stock price and oil price data to forecast turning points in UK quarterly GDP. They find the Markov switching MIDAS model performs better than a matched frequency model in determining US economic turning points, but has little benefit for forecasting recessions in the UK versus matched frequency. Foroni (2015) uses a MF-MS-VAR to forecast quarterly GDP growth in the euro area using four monthly indicators; the Economic Sentiment Indicator (ESI), the M1 monetary aggregate, headline industrial production and the slope of the yield curve. They find their model works well for nowcasting and short term forecasting of the euro area GDP growth. Finally, Pirschel (2016) uses a linear mixed frequency Bayesian VAR to provide monthly real time recession probabilities for the euro area using a number of monthly macroeconomic and financial indicators. The Y variable in this paper is also quarterly GDP.

Other useful, but not directly related to recession prediction, literature include Auroba (2009), Kumar (2013) and Andreou, Ghysels and Kourtellos (2013). Auroba (2009) use a dynamic factor model with daily term premium, weekly initial jobless claims, monthly employment and quarterly GDP to develop their own Business Conditions Index, where a large negative value indicates poor business conditions, and a value of 0 indicates normal business conditions. Their dynamic factors model updates the Business Conditions Index on a weekly basis. Kumar (2013) employs the same model as Auroba (2013) but for Canadian business conditions. Finally, Andreou, Ghysels and Kourtellos (2013) use daily financial data, in combination with monthly and quarterly macroeconomic variables, in a factor MIDAS model to estimate US GDP growth. They find that the inclusion of financial data at the daily frequency improves estimation results.

1.2.3 Binary Mixed Frequency Models

There is very little literature that applies mixed frequency to binary variables. As to date I find three papers; 1) Freitag (2014) uses a Probit MIDAS to analyze the relationship between sovereign downgrades and sovereign CDS premiums for Euro countries. 2) Audrino (2019) uses a Logit MIDAS to obtain bank failure probabilities in the US. 3) Jiang (2020) uses a U-MIDAS Logit model to study the default of listed companies in mainland China. All papers find improvement of forecast accuracy when the model allows for inclusion of higher frequency data. Papers 1) and 2) both carry out Monte Carlo simulations.

1.3 Motivation

1.3.1 Why is Daily and Weekly Financial Data Useful?

Theory suggests that the forward looking nature of financial assets prices should contain information about the future state of the economy and therefore should be considered as extremely relevant for macroeconomic forecasting, Andreou, Ghysels and Kourtellos (2013). The fact that the Y variable of interest in this paper (recession or not) is monthly, usually restricts researchers to aggregate the financial time series to match the dependent variable frequency. For example, the log of stock returns over a whole month versus daily stock returns, and average bond yields over the month versus daily bond yields.

Not using the readily available high frequency data such as the daily financial predictors to perform the monthly forecasts of recession probabilities has two important implications: (1) you lose the possibility of having real time daily, weekly or bi-weekly updates of the recession probabilities and (2) you lose information through temporal aggregation, Andreou, Ghysels and Kourtellos (2013). In other words, important within month information that can help in predicting current and future recessions may be 'washed' out by the assumption of equal weighting of data.

Andreou, Ghysels and Kourtellos (2010) show that the estimated slope coefficient of a regression model that imposes a standard equal weighting aggregation scheme, ignoring the fact processes are generated from a mixed data environment, yield asymptotically inefficient and inconsistent estimates. Both these asymptotic inefficiencies and inconsistencies can have adverse effects on forecasting.

1.3.2 Why use a Mixed Frequency Probit Model?

In this section I justify my use of a reduced form binary modeling approach, the MFP, versus other potential mixed frequency empirical methods. In previous literature a common method to deal with mixed frequency data has been using state space models. However, as well as being computationally complex which increases with the number of variables involved, they also require the correct specification of the model in high frequency, which is even more complex than usual given the missing observations in the dependent variable, Foroni (2013).

Andreou, Ghysels and Kourtellos (2013) explains that a mixed data sampling (MIDAS) regression model can be viewed as a reduced form representation of the linear projection that emerges from a state space model approach - by reduced form he means that the MIDAS regression does not require the specification of a full state space system of equations. They lie out some disadvantages of state space models versus reduced form mixed frequency models as (1) they are more prone to specification errors, as a full system of equations and latent factors are required and (2) it requires a lot more parameters to achieve the same goal. The system of equations requires many parameters for the measurement equation, state dynamics and their error processes. Therefore, state space models are far more complex in terms of specification, estimation and computation of forecasts, compared to the reduced form approach I propose in this paper.

As Andreou, Ghysels and Kourtellos (2013) explain, the Kalman filter approach to estimate state space models is often feasible when dealing with a small system of mixed frequencies such as Aruoba (2009) who use 6 time series. Instead, my analysis deals with a larger number of daily and weekly variables (upwards of 20) and therefore the approach I use is regression based and reduced form - notably not requiring to model the dynamics of each and every daily predictor series and estimate a large number of parameters.

There are, however, advantages to using state space models. In particular, once a state space system is estimated it is possible to produce forecasts at any horizon. In contrast, my MFP regressions need to estimate projection equations tailored to each horizon. However, since MFP regressions are relatively easy to estimate, the fact one has potentially different MFP regressions across different horizons does not come at a substantial cost. My paper also only focuses on a handful of forecast horizons, and therefore the cost of re-estimating a model for each horizon is low. Since the Kalman filter produces multi-horizon forecasts via iteration, specification errors, which pose a serious problem with high frequency data, are compounded forward across multiple horizon forecasts. Previous literature, including Findley (1983), Findley (1985), Lin and Granger (1994), Clements and Hendry (1996), Bhanzali (1999) and Chevillon and Hendry (2005), have shown iterated forecasting can feature poor performance in the presence of specification errors.

Kuzin (2011) compares a MF-VAR, cast into state space form, with the reduced form MIDAS approach to forecast Euro area GDP. The MF-VAR does not restrict the dynamics but suffers from the curse of dimensionality. The authors argue that it is difficult to rank the different approaches, so they compare their performance empirically. They find that the MF-VAR performs better for longer horizons (up to 9 months), whereas MIDAS performs better for shorter horizons (up to 4 to 5 months). Seeming my paper is based on shorter term forecasts up to 3 months and nowcasts, Chen and Tsay's (2011) adaption of MIDAS seems applicable.

Various mixed frequency Markov switching models are also popular in the literature, see Balcilar (2016) and Bessec (2015) among others in section 1.2.2. However, in all previous literature this and the above methods have all required the use of GDP to forecast economic turning points. Recession probabilities are then derived from this. Using GDP as opposed to the binary recession indicator released by the business cycle dating committees has issues. Firstly, it is released on a quarterly basis versus recession indicators which are released monthly. This severely reduces the number of Y observations for when there is a recession, which are already sparse. Secondly, the amount of times and magnitude to which GDP is revised after being first published can cause issues. Between 1990-2010, quarterly growth rates of US GDP were revised by an average of 0.26 points three years after its first publication. This revision reaches up to 0.37 points for the recession quarters, as opposed to 0.21 points in expansion, Bessec (2015). The delay in accurate GDP figures poses issues when attempting to accurately forecast recessions in real time.

Finally, my MFP is very similar to using MIDAS, but in a probit set up. Reasons for using my specific MFP vs a MIDAS probit are explained in section 1.4.3.

1.4 Methodology

1.4.1 Probit Model

I begin by looking at the standard probit model that will form a benchmark to compare my mixed frequency estimation to. Let y_{t+h} be a binary variable that equals 1 if month $t+h$ is subsequently declared as a recession and 0 otherwise. A forecast of the probability of a recession in month $t+h$ (p_{t+h}) from a probit regression is given by

$$p_{t+h} = Prob(y_{t+h} = 1|z_t) = \Phi(\beta' z_t) \tag{1}$$

where $\Phi(.)$ is the standard normal cumulative distribution function, β is a vector of coefficients, and z_t is a $k \times 1$ vector of predictors including an intercept.

1.4.2 Mixed Frequency Probit Model

In this paper I extend the standard probit model to allow for the inclusion of daily and weekly financial and macroeconomic data. Since the NBER Business Cycle Dating Committee of the US and the Business Cycle Council of the C.D. Howe Institute of Canada announce recessions at a monthly frequency, I will need a mixed frequency probit (MFP) to forecast future recession probabilities in both countries. I extend Chen and Tsay (2011) mixed frequency approach to the probit model. In their paper they allow for a polynomial to weight daily exchange rate data to forecast quarterly changes in commodity prices. Consider first a mixed frequency model with only one predictor x_1. Following the notation in Ghysels (2007) consider the following h-month ahead predictive regression.

$$y_t^* = \beta_0 + \beta_1 W(L^{1/m}, \theta) x_{1,t-h}^m + \epsilon_t, where \tag{2}$$

$$W(L^{1/m}, \theta) = \sum_{k=1}^{K} b(k;\theta) L^{(k-1)/m}, and \tag{3}$$

$$L^{s/m} x_{1,t}^m = x_{1,t-s/m}^m \tag{4}$$

Here y_t^* is a latent variable which represents the state of the economy as measured by the Business Cycle Dating Committees in the US and Canada, t denotes the basic time unit for the lower frequency data (monthly) from 1 to T, m and x^m indicate higher sampling frequency and observations, which is indexed from 1 to K (where K is finite). $L^{1/m}$ is the lag operator in frequency-m space, $b(k;\theta)$ is the weight on each of the K lagged higher frequency predictors and ϵ_t is a white noise process. In this paper the higher frequency data is at the daily and weekly level. For example, with daily frequency if $K = 15$ we would use the last 15 trading days of the current month to forecast recession probabilities h months in the future. This is as opposed to aggregating all the current months trading days into one observation, as past literature has done when predicting future binary economic states.

I can generalize equation (2) to contain q sets of mixed-frequency predictors, as well as $r - q$ sets of matched frequency predictors as follows:

$$y_t^* = \beta_0 + \beta_1 W_1(L^{1/m}, \theta_1) x_{1,t-h}^m + ... + \beta_q W_q(L^{1/m}, \theta_q) x_{q,t-h}^m + \beta_{q+1} x_{q+1,t-h} + ... + \beta_r x_{r,t-h} + \epsilon_t, where \tag{5}$$

$$W_i(L^{1/m}, \theta_i) = \sum_{k=1}^{K} b_i(k; \theta_i) L^{(k-1)/m}, and \tag{6}$$

$$L^{s/m} x_{i,t}^m = x_{i,t-s/m}^m, \forall i = 1, ..., q \tag{7}$$

Parameters β_1,...,β_q measure the aggregate impact of predictors of the mixed frequency (daily and weekly) data $x_{1,t-h}$,...,$x_{q,t-h}$ on the lower frequency (monthly) y_t^*, provided that the sum of the weighting polynomial $W_1(L^{1/m}, \theta_1), ..., W_q(L^{1/m}, \theta_q)$, are normalized to 1. Parameters β_{q+1},...,β_{q+r} measure the aggregate impact of predictors of the matched frequency (monthly) data $x_{q+1,t-h}$,...,$x_{q+r,t-h}$ on y_t^*.

To weight the higher frequency daily financial data in my MFP a $K \times n$ Vandermonde matrix is used:

$$V = \begin{bmatrix} 1 & 1^1 & 1^2 & \dots & 1^{n-1} \\ 1 & 2^1 & 2^2 & \dots & 2^{n-1} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & K^1 & K^2 & \dots & K^{n-1} \end{bmatrix} \tag{8}$$

This follows Almon (1965) that assumes each lag coefficient can be approximated by a polynomial of degree $n - 1 < K$. Therefore, instead of having to estimate $1 + Kq$ parameters (the intercept and weights on each K higher frequency observation of the q variables) the model is reduced to $1 + nq$ parameters to estimate. This ignores matched frequency coefficients that need to be estimated in addition.

Using V from (8) I can re-write (5) as follows in matrix notation:

$$Y = \beta_0 + \beta_1 X_1 V \alpha_1 + ... + \beta_q X_q V \alpha_q + \beta_{q+1} X_{q+1} + ... + \beta_r X_r + \epsilon \tag{9}$$

Where X_i for $i = 1, 2, ..., q$ is a $(T \times K)$ matrix of higher frequency financial and macroeconomic data, sampled at a daily and weekly rate. X_i for $i = q + 1, ..., r$ is a vector of length T of matched frequency financial and macroeconomic data, sampled at a monthly rate. Y is a vector of length T of monthly binary variables indicating a recession or not. β_i measures the aggregate impact of X_i on Y. V is as described in (8) and α_i is a $n \times 1$ vector of coefficients that form the polynomial of degree $n - 1$, which weights the higher frequency data. Assuming I am trying to estimate recession probabilities one month ahead, using data from every day of the month, a visual representation of the matrices for the higher frequency data are shown below.

$$Y = \begin{bmatrix} y_{feb} \\ y_2 \\ \vdots \\ y_T \end{bmatrix} X_i = \begin{bmatrix} x_{i,jan31st} & x_{i,jan30th} & \cdots & x_{i,jan1st} \\ x_{i,2-h} & x_{i,2-h-1/K} & \cdots & x_{i,2-h-(K-1)/K} \\ \vdots & \vdots & \cdots & \vdots \\ x_{i,T-h} & x_{i,T-h-1/K} & \cdots & x_{i,T-h-(K-1)/K} \end{bmatrix} \tag{10}$$

Note that the inclusion of all trading days in the month is not needed. If $K = 15$ then only the 15 most recent trading days from that month are included. Simplifying (9) further:

$$Y = \beta_0 + Z_1 \gamma_1 + ... + Z_q \gamma_q + \beta_{q+1} X_{q+1} + ... + \beta_r X_r + \epsilon \tag{11}$$

Where $Z_i = X_i V$ and $\gamma_i = \beta_i \alpha_i$. The parameter γ can then be estimated via maximum likelihood, and aggregate effects of X_i, $\forall i = 1, ..., q$, on Y, as well as the higher frequency weighting function, can be solved through manipulation of the $\gamma' s$.

1.4.3 Identification of β

By restricting the weights on the higher frequency data to sum up to 1, we can identify $\beta_1, ..., \beta_q$. This means letting $\mathbf{1}^\top V\alpha_i = 1$ for $i = 1, ..., q$. We can the use the fact that $\gamma_i = \beta_i \alpha_i$ and therefore $V\gamma_i = \beta_i V \alpha_i$ to obtain:

$$\hat{\beta}_i = \mathbf{1}^\top (V\hat{\gamma}_i), \forall i = 1, ..., q, and \tag{12}$$

$$var(\hat{\beta}_i) = \mathbf{1}^\top V var(\hat{\gamma}_i) V^\top \mathbf{1} \tag{13}$$

Equation (12) demonstrates that the aggregate impact of X_i on Y can be obtained directly without separating out β_i's versus b_i's like in the MIDAS specification given in (5) and (6). In the MFP we combine the β_i's versus b_i's into free parameters that can be parameterized with the Vandermonde matrix as in (12). Hence, the MFP involves a one step procedure that automatically embeds the identification condition for β_i, unlike a typical MIDAS regression.

Addtionally the typical weighting restrictions used in other binary MIDAS models, for example 'Exponential Almon Lags' and 'Beta Lag', Ghysels (2007), are no longer needed. The Almon lag polynomial is well known to offer useful approximations for a variety of weighting functions, provided that n is large enough. This means (12) can deliver consistent estimates of the aggregate impact parameters under a wide range of true underlying weighting functions. Overall, the polynomial weighting function is more flexible than traditional MIDAS methods, one reason being that weights do not have to be positive.

1.5 Data

1.5.1 US

My entire data sample period for Y runs from October 1971 to January 2019, which includes the last 6 US recessions . The Y variable is a binary indicator that takes the form:

$$y_t = \begin{cases} 1, & \text{if NBER defined recession.} \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

For the X variables a full table of the data can be found in Table 1.1 in the data appendix. I split my data into four separate markets; (1) bond market containing 15 financial indicators, (2) exchange rate market containing 4 different rates, (3) stock market containing 4 indicators, and (4) macroeconomic indicators containing 6 variables. For each of the asset classes (1)-(3) I can observe daily, weekly or monthly data, and for the macroeconomic indicator (4), data is observed at a weekly and monthly frequency depending on the variable. This variable selection has shown to have good real time predictability of US business conditions in the previous literature of Fossati (2015), Camacho (2014), Chauvet and Piger (2008) and many more.

Two issues I have with the data are: 1) Co-linearity between variables. For example, yields on 5 year Treasury Bonds are highly correlated with yields on 10 year Treasury Bonds. The same can be said for the macroeconomic indicators which vary with economic conditions and hence are correlated with one another. 2) The number of X variables compared to the Y variable observations. This is especially apparent with mixed frequency data, where, for example, 4 parameters need to be estimated per variable if using a cubic polynomial to weight the data. To overcome these problems I use Factor Analysis to capture variability among observed, correlated variables in terms of a lower number of unobserved variables.

Prior to estimation, data is first transformed to be stationary - see the data appendix for relevant transformations. Since real activity variables are usually available with some lag, I account for data availability at time t by using the last known value $x_{i,t-l}$, where l indicates the publication lag of variable i. Publication lags for US indicators are adopted from Katayama (2009) and are presented in the data appendix. Although many macroeconomic variables are subject to revisions after their initial data release I use the final revised data in my prediction model. Data revisions appear to impact real activity factors less than individual real activity variables, and hence final results of models with factors containing revised and non revised data is not significantly different. This has been found and documented by Fossati (2015), Berge and Jorda (2011) and Chen et al. (2011). Therefore because I use factors estimated from US and Canadian macroeconomic variables, instead of using individual macroeconomic variables in my model, I chose to use revised data only.

Factor analysis is then run separately on each asset class (1), (2), (3) and (4) to get a set of bond, exchange rate, stock and macroeconomic market factors. The factor analysis model takes the following form

$$X = \Lambda F + \epsilon \tag{15}$$

Where X is a $(N \times T)$ matrix of observed data and Λ is an $(N \times M)$ matrix of factor loadings, where M corresponds to the number of common factors being estimated. F is a $(M \times T)$ matrix of latent factor scores and ϵ is the error term. There is no strict rule on deciding how many factors to use. Currently I analyze the cumulative variance explained by the factors and stop adding additional factors when the marginal variance explained from the extra factor is low. This leads me to use 3 factors for the bond market, 1 factor for the exchange rate market, 1 factor for the stock market and 2 factors for the real market. The MFP regression I then run is as follows:

$$Y = \beta_0 + \underbrace{F_{1,D}^{mkt_i} V_D \gamma_{D1} + ...}_{\text{daily freq}} + \underbrace{F_{1,W}^{mkt_i} V_W \gamma_{W1} + ...}_{\text{weekly freq}} + \underbrace{F_{1,M}^{mkt_i} \delta_1 + ...}_{\text{monthly freq}} + \epsilon \tag{16}$$

Where Y is a vector of binary indicators showing whether the economy is in a recession or not. $F_D^{mkt_i}$ and $F_W^{mkt_i}$ are daily and weekly frequency factors for market i, where mkt_i refers to the four asset markets; bond, stock, exchange rate and real. V_D and V_W are the associated Vandermonde matrices for the daily and weekly frequency data. The γ's are the parameters to estimate from the mixed frequency part of the model, of which can be used to find the aggregate impact of F on Y as stated in section 1.4.3. Finally, $F_M^{mkt_i}$ are the monthly factors for asset market i which are the same frequency as Y and hence enter the probit regression in the standard way.

Once I have dealt with missing data, the daily mixed frequency component of the MFP ($F_D^{mkt_i}$) consist of 15 trading days worth of information for each month (K=15). The weekly mixed frequency component of the MFP ($F_W^{mkt_i}$) consist of 4 weeks worth of information for each month (K=4). In this paper I do not look at the optimal choice of K, and use the highest K possible given the data restrictions. To weight the higher frequency data I use a polynomial of degree 3 and 2 for daily and weekly data respectively. I deem a polynomial of degree 3 flexible enough to capture any potential weighting function on daily data. For the weekly data, if I use a polynomial of degree higher than 2 I may as well include the weekly data points separately as it would require the same (or more) parameters to be estimated. A polynomial of degree 2 is still flexible enough to incorporate most weighting schemes.

Initially, before moving onto nowcasting, I focus on short horizon forecasts of recession probabilities - specifically 1, 2 and 3 months ahead. Recession probabilities can be estimated from equation (16) via maximum likelihood.

For direct comparison of my MFP I use a benchmark model. For each asset market, data that is used at the daily and weekly frequency in the MFP is aggregated to the monthly frequency. I then estimate monthly factors for each asset market as in equation (43). This gives me the below regression:

$$Y = \alpha_0 + F_{1,M}^{mkt_i}\theta_1 + ... + F_{p,M}^{mkt_i}\theta_p + u \tag{17}$$

Where Y is the same vector of binary indicators as in the MFP, indicating whether the economy is in a recession or not. $F_{1,M}^{mkt_i}$ through $F_{p,M}^{mkt_i}$ are monthly factors estimated from asset market i. The number of factors p included in the benchmark model depends on the MFP. For example, if the MFP contains 3 daily bond factors and 2 monthly real activity factors, the benchmark model will contain 3 monthly bond factors and 2 monthly real activity factors. This setup allows for exact comparison of using the daily and weekly frequency versus the monthly aggregated data. The benchmark matched frequency probit model is estimated via maximum likelihood.

1.5.2 Canada

My entire data sample period for Y runs from March 1972 to June 2018, which includes the last 5 Canadian recessions. The Y variable is a binary indicator that takes the form:

$$y_t = \begin{cases} 1, & \text{if C.D. Howe Institute defined recession.} \\ 0, & \text{otherwise.} \end{cases} \tag{18}$$

For the X variables a full table of the data can be found in Table 1.2 in the data appendix. In a similar way to the US I split my data into five asset markets; (1) bond market containing 11 financial indicators, (2) exchange rate market containing 4 different rates, (3) stock market containing 3 indicators, (4) commodity market containing 4 indicators and (5) macroeconomic indicators containing 4 variables. Asset markets (1)-(3) and (5) have been used in the previous literature, e.g. Fossati (2018) and Hao and Ng (2011), and have been found to be good individual predictors. Asset market (4) was added due to the importance of commodity exports to the Canadian economy. According to commodity.com in 2018 Canada was the 4th largest exporter of crude oil, 7th largest exporter of gold, 2nd largest exporter of wheat and largest exporter of aluminum in the world. Therefore the prices of these 4 commodities have been added as they could potentially significantly impact the health of the Canadian economy.

Prior to estimation, data is first transformed to be stationary - see the data appendix for relevant transformations. Again, since real activity variables are usually available with some lag, I account for data availability at time t by using the last known value $x_{i,t-l}$, where l indicates the publication lag of variable i. Publication lags for Canadian real activity indicators are obtained from Statistics Canada. I then follow the same steps as explained in section 1.5.1, which are summarized again below.

Step 1) Factors are estimated for each asset market, using the setup explained in equation (15), to be used in the MFP. Specifically, I extract 3 factors for the bond market, 1 factor for the exchange rate market, 1 factor for the stock market, 1 factor for the commodity market and 1 factor for the real activity market.

Step 2) Estimate equation (16) using the factors from the Canadian data, and extract the recession probabilities. The daily mixed frequency component of the MFP ($F_D^{mkt_i}$) consists of 15 trading days worth of information for each month (K=15). The weekly mixed frequency component of the MFP ($F_W^{mkt_i}$) consists of 4 weeks worth of information for each month (K=4). A polynomial of degree 3 (cubic) is used to weight the daily data and of degree 1 (linear) to weight the weekly data. Before moving onto nowcasting, I focus on short horizon forecasts of recession probabilities - specifically 1, 2 and 3 months ahead.

Step 3) Compare the MFP for Canada to the benchmark model explained in equation 17.

1.6 Results

1.6.1 Forecast Evaluation

The first method used to evaluate forecast accuracy of the MFP and benchmark models is the Quadratic Probability Score (QPS). The QPS is equivalent to the mean square area when using binary models and is defined by

$$QPS = \frac{2}{T} \sum [\hat{p}_{t+h} - y_{t+h}]^2 \tag{19}$$

where T is the number of forecasts, $\hat{p}_{t+h}$ is the predicted probability of recession for month $t+h$ for a given model and y_{t+h} is the realized recession indicator in the month t+h. The QPS can take values from 0 to 2 with smaller values indicating more accurate predictions.

As discussed in the Introduction, the business cycle dating committees of the US and Canada delay their announcements of recession periods. Arguably the most important time for economic agents to know whether the economy is in a recession is the first few months, which is not always immediately obvious by observing single economic indicators. Therefore a model that is able to identify whether the economy is in a recession in the early stages is beneficial to these economic agents. I devise an additional evaluation measure which slightly adapts equation (19). I call this measure QPS_{onset}, where *onset* refers to the onset of a recession which I define as the first 5 months of a recession period. In other words, I want to see if the MFP has particularly good prediction power at the beginning of recession periods, versus that of the benchmark model.

$$QPS_{onset} = \frac{2}{T_{onset}} \sum [\hat{p}_{onset,t+h} - \mathbf{1}]^2 \tag{20}$$

T_{onset} refers to the number of forecasts during recession onsets, for example, if there are 3 recessions in the period being examined, $T_{onset} = 15$. $\hat{p}_{onset,t+h}$ is the predicted probability of a recession for month $t+h$, when period $t+h$ falls within the first 5 months of a recession. Finally, $\mathbf{1}$ is a vector indicating that the true state of the economy is a recession.

The third evaluation method I use is the log probability score (LPS), which is given by

$$LPS = -\frac{1}{T} \sum [y_{t+h} log(\hat{p}_{t+h}) + (1 - y_{t+h}) log(1 - \hat{p}_{t+h})] \tag{21}$$

The LPS can take values from 0 to $+\infty$ and smaller values indicate more accurate predictions. Compared to the QPS, the LPS score penalizes large errors more heavily.

For each of these first 3 evaluation methods I divide the statistic from the MFP model by that of the benchmark model. A value less than 1 indicates that the MFP is a better model. The amount below 1 can be interpreted as the % improvement from using the mixed frequency model versus matched frequency. For example, a value of 0.8 indicates a 20% improvement in the evaluation statistic.

The fourth evaluation method I use is the Diebold-Mariano test with a squared loss function. This test compares the predicted probabilities of the MFP and the benchmark model with the actual values of Y that occurred. For the purpose of this paper the H_A is that the MFP is more accurate than the benchmark model, and hence a low p-value is desired. I choose a squared loss function, as opposed to a linear loss function, as I deem incorrect recession predictions to be costly to the economy and therefore want to penalize these at a higher rate.

1.6.2 In Sample Results - US

The in sample period uses data for X from September 1971 to October 2018 which I use to forecast recession probabilities at alternative horizons: $h = 1$, 2 and 3 months. Given the way I have set up my data there are multiple variations of models, in terms of X variables, I can use. Firstly, I have estimated factors that represent 4 asset markets; bond, exchange rate, stock and the real market, but the best model does not necessarily contain all asset markets. Secondly, the first 3 asset markets have daily, weekly and monthly frequency available. I do not assume that I must always use the highest frequency available for the model. For example, it may be the case that monthly stock factors provide a better fit than using daily stock factors due to the fact that daily stock information is volatile.

The obvious next question is how do I classify what the best model is? For my research I deem the most important aspect to be the out of sample forecasting exercise, as this provides real time recession predictions, and hence decide the best model from the forecast evaluation of the out of sample results. These are reported in the next section. I analyze all the possible combinations of asset markets and frequencies, and report the best model based on the out of sample forecast evaluation methods explained in the previous section. For the US the best model uses 3 daily bond market factors, a monthly stock market factor and 2 monthly real market factors.

The asset markets that appear in the best model are in line with the previous literature. Estrella and Mishkin (1998), Katayama (2009), and Owyang et al. (2015), among others, find that the real economic activity indicators that form the real market factor in this paper improve recession forecasts, particularly at short horizons. Wright (2006) find that the term spread and federal funds rate have high predictive power for recessions at short and long forecast horizons. The bond market factors are functions of these variables. Fossati (2015) finds that the best performing model at a 3 month forecast horizon includes a monthly stock market factor, monthly real market factor and the monthly 3 month less 10 year spread on government bond yields. This is very similar to what I find, apart from the fact I used a daily frequency bond market factor, instead of just the monthly 3 month less 10 year spread.

Figure 1.1 shows the in sample recession probabilities for forecast horizons 1, 2 and 3 months. The red line shows recession probabilities calculated from the MFP, and the black line shows probabilities calculated from the benchmark model. A common occurrence is the ability of the MFP model to send a stronger signal at the onset of a recession compared to that of the benchmark matched frequency model. This is especially evident in detecting the most recent recession for all forecast horizons. The MFP also produces higher overall recession probabilities in the fourth and fifth US recessions in the sample. Recession probabilities for the first three recessions are very similar in both models. In non-recession periods the models behave very similarly, with occasional short lived false positives. These can usually be attributed to specific negative shocks, for example in Summer 1998 when the $S\&P$ 500 fell 20% in a short span of time. Fossati (2015) and other previous research detect the same false positives.

Table 1.3 reports the evaluation methods explained in the previous section. For the QPS and LPS I divide the statistic from the MFP model by that of the benchmark model. A value less than 1 indicates that the MFP is a better model. For all evaluation methods the MFP performs better than the matched frequency benchmark, especially when focusing on the onset of a recession, defined as the first 5 months of a recession period, with improvements up to 24%. The p value from the Diebold Mariano test is reported in the last row. For $h = 3$ months, we reject H_0 at the 1% level of significance, in favor of H_A, which states that the MFP is more accurate than the benchmark.

1.6.3 In Sample Results - Canada

The in sample period uses data for X from February 1972 to March 2018, which I use to forecast recession probabilities at alternative horizons: $h = 1$, 2 and 3 months. Again I decide the best model based on forecast evaluation of the out of sample results, and only report these. Firstly, I must choose which of the 5 asset markets (bond, exchange rate, stock, commodity and the real market) to include in the model. Secondly, I must choose the frequencies of each of the asset market factors to include. Unlike the US, the Canadian bond market data is only available at the weekly frequency, not daily, for the full sample. The Canadian stock market factor is only available at the monthly frequency. For Canada the best model uses 3 weekly bond market factors, a monthly stock market factor and a monthly real market factor. Fossati (2018) also finds that a Canadian real market factor is a preferred variable for generating short term (1 to 3 month) recession probabilities of the Canadian economy. He finds that excluding this leads to a substantial deterioration in fit, with larger QPS and LPS values. He also finds that monthly Canadian bond yields improve prediction at the longer forecast horizons (6 to 12 months), however I find that weekly bond factors are crucial for forecasting at short horizons of 1 to 3 months. Excluding these leads to worse fit. The difference may be due to my MFP including higher frequency data.

Figure 1.2 shows the in sample recession probabilities for forecast horizons 1, 2 and 3 months. The red line shows recession probabilities calculated from the MFP, and the black line shows probabilities calculated from the benchmark model. For the first 4 recessions there is slight improvement in recession onset prediction and the MFP probability usually peaks at a higher value than the benchmark model. Both models perform similarly in non-recession periods with less false positives than the US. The most noticeable false positive occurs in the mid 1980's and is found in other literature, see Fossati (2018). Overall, in sample improvements for Canada are visually not so obvious compared to the US, which may emphasize the importance of daily frequency bond market factors over weekly frequency.

Table 1.4 reports the evaluation methods explained in the previous section. All QPS and LPS values are below 1, indicating that the MFP model performs better than the benchmark matched frequency model. I observe improvements up to 12%, which is less than those reported for the US. Of the first 3 evaluation methods, the only one that is better than the US is the QPS on the full sample at the 1 month forecast horizon. Again this may emphasize the importance of daily versus weekly bond data frequency in forecast accuracy in mixed frequency regressions. Finally, the Diebold Mariano test suggests that the MFP is more accurate than the benchmark model at the 5% and 10% significance level for the 1 and 2 month forecast horizons respectively.

1.6.4 Out of Sample Results

The out of sample forecasting of recessions comes with some complications. The Y variable indicating whether the US and Canadian economies are in a recession or not is observed with a lag. Nyberg (2010) assumes this 'publication lag' in the recession indicator is 9 months when forecasting US recessions. However, because the Great Recession was not officially declared a recession in the US until 12 months after the fact I will assume a 12 month lag for both the US and Canada. This assumption can be relaxed to be less or more than 12 months with ease, however it does not significantly impact results.

It is important that when carrying out the out of sample forecast I only use information known at time t to forecast recession probabilities in $t + h$. For example, if I am standing at the end of July 2018 and I want to forecast the probability that a recession will begin in August 2018 then only X and Y data known at the end of July 2018 can be used. Since financial markets data is released in real time I can use the exchange rate, bond and stock market factors up to and including the end of July 2018. This is with the exception of the $S\&P$ 500 price to earnings ratio which is observed with a 2 month lag due to earnings data being released on a quarterly basis. The real market factor contains data that is released with a publication lag. I follow Fossati (2015) and assume this is a 1 period lag, so in the example above I would only use values up to the end of June 2018. The rest of the variables with publication lags are stated in the Data Appendix.

The publication lag in the Y binary indicator is where the main complication comes from. Since I assume there is a 12 month lag, using the example above, I would only be able to estimate the probit model up to August 2017. The estimated parameters from this probit model along with the X data known at the end July 2018 can be used to extract the real time 1 month ahead recession probability for August 2018. This method can easily be generalized to h month ahead predictions.

With this in mind, I use recursive estimation for the out of sample period, expanding the time window one month at a time. Factors for the MFP and benchmark model are re-estimated every time an additional month of X data becomes available. Finally, I re-estimate the probit model as I add each additional month of data. Rolling windows is another potential technique to use, however because of the lack of recessions that have occurred in my sample in the US (6) and Canada (5) I do not. The above methodology allows me to estimate real time h month ahead recession probabilities for the US and Canada.

1.6.5 Out of Sample Results - US

The out sample period forecasts real time recession probabilities from November 1988 to January 2019, which includes the last 3 US recessions. As explained in the in sample results section there are multiple variations of the model that can be run. After running all variations the best out of sample model at the 1, 2 and 3 month ahead forecast horizons includes 3 daily bond market factors, a monthly stock market factor and 2 monthly real market factors.

This setup is used for both the MFP and the benchmark model for direct comparison. I explain why these particular asset markets are chosen in more detail later in this section. Figure 1.3 shows the out of sample recession probabilities for forecast horizons 1, 2 and 3 months. The red line shows recession probabilities calculated from the MFP, and the black line shows probabilities calculated from the benchmark model.

Visually we can see that the MFP gives a higher probability of a recession for all recessionary periods at all forecast horizons, $h = 1, 2, 3$ months. More importantly, arguably, there is much stronger detection of a recession onset in every recessionary period for all the forecast horizons when comparing the MFP against the benchmark. This is especially evident in the most recent US recession of our sample. In non recession periods both models perform similarly with the only significant false positive occurring in 1998. As explained in previous sections this can be attributed to the $S\&P$ 500 falling 20% in a short span of time. The false positive is more prevalent in the MFP model, indicating that the daily bond market factors are driving this additional sensitivity compared to the benchmark - this is because the only difference between the 2 models is the daily bond market factors versus the monthly bond market factors.

Table 1.5 reports the evaluation methods used to assess the performance of the MFP against the benchmark. For the QPS and LPS I divide the statistic from the MFP model by that of the benchmark model, as I did for the in sample results. A value less than 1 indicates that the MFP is a better model. For all evaluation methods the MFP performs better than the matched frequency benchmark, especially when focusing on the onset of a recession, defined as the first 5 months of a recession period. Improvements can reach up to 30%. This confirms the visual assessment given above. The p value from the Diebold Mariano test is reported in the last row. For all forecast horizons, we reject H_0 at the 1% level of significance, in favor of H_A, which states that the MFP is more accurate than the benchmark over the whole out of sample period.

By including and excluding certain asset market factors in the models I can gain insights into how these markets interact with the US economy, and which are key leading indicators for recessions. Monthly real market factors are key in predicting recessions, and model performance drops significantly without these. However, in the US I also had the ability to use Weekly Initial Jobless Claims as a higher frequency weekly macroeconomic variable, but it did not improve model performance and hence was omitted from the best model. Daily bond market factors are also key for the superior performance of the MFP over the benchmark model. Specifically, these are what drive the improved recession onset predictability compared to a model that uses monthly bond market factors. Weekly corporate bond data do not improve model prediction and are therefore excluded. Daily stock market factors introduce much volatility and false positives into the recession prediction and are hence omitted, however using just monthly stock factors improve model prediction. Finally, daily and monthly exchange rate factors do not improve recession prediction and are not included in the best models.

1.6.6 Out of Sample Results - Canada

The out sample period forecasts real time recession probabilities from August 1988 to June 2018, which includes the last 2 Canadian recessions. As explained in the in sample results section there are multiple variations of the model that can be run. After running all variations the best out of sample model at the 1, 2 and 3 month ahead forecast horizons includes 3 weekly bond market factors, a monthly stock market factor and a monthly real market factor - similar to the best model found for the US. This setup is used for both the MFP and the benchmark model for direct comparison. I explain why these particular asset markets are chosen in more detail later in this section. Figure 1.4 shows the out of sample recession probabilities for forecast horizons 1, 2 and 3 months. The red line shows recession probabilities calculated from the MFP, and the black line shows probabilities calculated from the benchmark model.

Visually, unlike the US we do not observe out of sample improvements from the higher frequency data in the most recent recession in our sample, which may emphasize the importance of daily frequency bond market factors over weekly frequency. For the recession in the early 1990's results between the MFP and benchmark look different, with improvements in recession onset prediction for all forecast horizons. Both models perform similarly in non-recession periods with no notable false positives, besides one in 1995/96 at the 1 month forecast horizon.

Table 1.6 reports the evaluation methods used to assess the performance of the MFP against the benchmark. For the QPS and LPS I divide the statistic from the MFP model by that of the benchmark model, as I did for the US. For the QPS and LPS which cover the full out of sample period, the MFP performs worse than the matched frequency benchmark. The QPS onset, which measures the models effectiveness at detecting the beginning of a recession, shows significant improvement in results for the MFP versus the benchmark at all forecast horizons. This improvement is up to 22% depending on the forecast horizon. The Diebold Mariano statistic does not conclude that the MFP improves prediction performance over the whole out of sample period. These results outline the potential importance that daily financial data can have in predicting future recessions over weekly and monthly frequency. It may also indicate that news and economic shocks are incorporated in Canadian financial market prices at a slower rate than in the US. Unfortunately, daily bond market data is not available for the whole of the sample period for Canada, and is only recorded, with missing observations, beginning in 1991. However, like the US I have still improved in recession onset prediction by using higher frequency financial market data in the model.

In terms of asset markets that are key leading indicators for Canadian recessions, results are similar to those of the US. Monthly real market factors are key in predicting recessions, and model performance drops significantly without these. Bond market factors, whether they be monthly or weekly frequency improve prediction especially at the recession onset. Stock market factors were not available at the daily frequency for Canada, but the monthly frequency improves forecasting performance. As with the US, daily and monthly exchange rate factors do not improve recession prediction and are not included in the best models. Finally, the commodity market both at the daily and monthly frequency did not improve the out of sample forecast error and were therefore not included.

1.6.7 Higher Frequency Data Weighting Functions

Much of the past mixed frequency literature has neglected to inspect the weighting functions of the higher frequency data. This is important to check if the data is truly weighted differently dependent on the higher frequency lag, as opposed to an equal weighting function that the matched frequency methodology imposes. Additionally it can be used as a robustness check to make sure the weighting functions are stable over time, and as a way to gain intuition with regards to which data within the time period has the largest impact on the forecasts.

To evaluate the higher frequency data for the US and Canada I extract the weighting functions for each of the out of sample periods. For example, with the US I estimated 361 probit regression models for the recursive out of sample estimation procedure. This gives me 361 weighting functions for each higher frequency variable, that can be plotted and examined. To extract the weighting function from the regression coefficients we can use the fact that the vector $\hat{\gamma}$, which are the estimated coefficients from the mixed frequency model, are formed of two parts; the aggregate impact $\hat{\beta}$ multiplied by the vector of polynomial coefficients $\hat{\alpha}$. This leaves us with one equation, one known vector of parameters $\hat{\gamma}_i$, the unknown scalar parameter $\hat{\beta}$ and the unknown vector of parameters $\hat{\alpha}$. However, using the assumption that the weights on the higher frequency data must sum to 1, explained in Section 1.4.3, we can identify $\hat{\beta}$ as follows.

$$\hat{\beta}_i = \mathbf{1}^\top (V\hat{\gamma}_i), \forall i = 1, ..., q \tag{22}$$

A simple transformation that divides the vector $\hat{\gamma}$ through by the scalar $\hat{\beta}$ uncovers the vector of coefficients, $\hat{\alpha}$, for the weighting function. Weighting functions can then be plotted for the higher frequency variables overtime.

The top panel of Figure 1.5 shows the cubic polynomial weighting function for the first daily US bond market factor in the one month ahead forecasting model. The corresponding lag of 1 on the x-axis refers to the most recent day in the month, and a lag of 15 refers to the least recent day of the month in the data. The weighting function for each probit model estimated in the recursive out of sample procedure has been included to check for stability of the weighting function, and can also be used to see if there has been a structural break in the weighting function over time. Each weighting function has a distinct shape with notably more weight put on the most recent observation, and less weight put on the remaining observations. This could indicate that the most recent daily financial data matters the most when forecasting recessions in the US. For reference, a flat line would indicate an equal weighting function for the higher frequency data, implying higher frequency data does not add information to the forecasting model compared to matched frequency data. The weighting functions appear to be stable over time.

The bottom panel of Figure 1.5 shows the linear polynomial weighting function for the first weekly Canadian bond market factor in the one month ahead forecasting model, with the corresponding lag of 1 on the x-axis referring to the most recent week in the month. Unlike the US there are not obvious higher weights on certain weeks. The first and fourth week of the month have the highest absolute weights, but whether these weighting functions are statistically different from a flat line for all out of sample periods is unknown. This may explain why improvements from the inclusion of mixed frequency data to forecast Canadian recession are more muted than in the case for the US.

1.7 Nowcasting

As I discussed at the beginning of the paper, one of the main advantages to using mixed frequency data is the ability to update estimates at a more frequent rate than with matched frequency data. In this paper I have explained the importance of the constant interest and need from economic agents to know current economic conditions. With the MFP framework, that uses daily and weekly financial data, I can potentially update the current and future months recession probabilities every trading day once data is released, and hence serve as constant source for forecasting the current un-observable state of the US and Canadian economies. I plan to follow Andreou, Elena and Ghysels (2013) and Gomez-Zamudio and Ibarra (2017) method of nowcasting with leads. Both these papers use current quarter daily financial data to update current and future quarter GDP estimates for the US and Mexico. As explained in Andreou, Elena and Ghysels (2013) there are quite a few variations in the specification of a nowcasting with leads regression model. One particular way which I will focus on is shown below:

$$y_t^* = \beta_0 + \beta_1 W_i(L^{1/m}, \theta)x_{1,t-h+1}^m + \beta_2 W_j(L^{1/m}, \theta)x_{1,t-h}^m + \epsilon_t, where \tag{23}$$

$$W(L^{1/m}, \theta) = \sum_{k=1}^{K} b(k;\theta)L^{(k-1)/m}, and \tag{24}$$

$$L^{s/m}x_{1,t}^m = x_{1,t-s/m}^m \tag{25}$$

Where y_t^* is a latent variable which represents the state of the economy, t denotes the basic time unit for the lower frequency data (monthly) from 1 to T, m and x^m indicate higher sampling frequency and observations, which is indexed from 1 to K (where K is finite). $L^{1/m}$ is the lag operator in frequency-m space, $b(k;\theta)$ is the weight on each of the K lagged higher frequency predictors and ϵ_t is a white noise process.The difference between nowcasting with leads and equation 2 is the second term on the RHS of equation 23. This $x_{1,t-h+1}^m$ term refers to the 'lead' and will be weighted differently compared to the previous month. For example, if the forecast horizon is 1 month and we are forecasting a recession probability for February, we usually only use the information known up to the end of January. However, with nowcasting, once we enter the month of February we do not need to wait until the end of the month to update our X data to forecast a March recession probability. Instead we can take account of the fact that financial data is released without delay or measurement error, and update the current February recession probability, potentially, as every trading day of February passes. This allows for constant updating of economic conditions. Visually, and intuitively, the matrices may look like the following:

$$Y = \begin{bmatrix} y_{feb} \\ y_{mar} \\ \vdots \\ y_T \end{bmatrix} X_i = \begin{bmatrix} x_{i,feb15th} & \cdots & x_{i,feb1st} \\ x_{i,mar15th} & \cdots & x_{i,mar1st} \\ \vdots & \cdots & \vdots \\ x_{i,T-h+j/K} & \cdots & x_{i,T-h+1/K} \end{bmatrix} X_j = \begin{bmatrix} x_{i,jan31st} & \cdots & x_{i,jan1st} \\ x_{i,feb28th} & \cdots & x_{i,feb1st} \\ \vdots & \cdots & \vdots \\ x_{i,T-h-1/K} & \cdots & x_{i,T-h-(K-1)/K} \end{bmatrix} \tag{26}$$

Where X_i is the matrix referring to the 'lead' if we were 15 trading days into February, which can be updated as new information is released, and X_j is information from 1 month previous. As Andreou, Elena and Ghysels (2013) explains, conventional nowcasting typically refers to within period updates of forecasts. For example, updating the current months recession probability as shown above. Nowcasting with leads can be viewed as current month updates of current months recession probabilities, but also of any future horizon recession probability forecast (i.e. for h$= 2, 3$ months).

A common method for nowcasting is to use state space models. While both state space and my MFP model can produce multiple horizon forecasts, a subtle difference is that the MFP can produce direct, as opposed to iterated h step ahead forecasts. Arguably iteration-based forecasts can suffer from misspecification, which can be compounded across multiple horizons that may produce inferior forecasts; see Marcellino, Stock and Watson (2006). However, for the purpose of my paper I will focus on the conventional definition of nowcasting, updating just the current months recession probability to see if this improves out of sample prediction.

1.7.1 Results

I follow the same out of sample forecasting procedure as described before, using the model explained in equation 16, adding additional leads for the financial data as laid out in equation 23. However, one small complication is the number of variables that are needed to be estimated. For example, in the case of the US the best model found for out of sample estimation included 3 daily bond market factors, 1 monthly stock market factor and 2 monthly real market factors. Assuming I use a polynomial of degree 3, I need to estimate 16 parameters (including a constant). When I add the lead terms for the daily bond market factors this could potentially increase to 28 parameters to estimate, which is high relative to my number of observations. To overcome this problem I have two potential solutions; 1) adjust down the degree of the polynomial on the 'lead' data as less data points do not need such a flexible function to weight them. 2) Remove some of the lagged mixed frequency factors which represent older information, but keep all the 'lead' factors which represent the most up to date information. Financial markets absorb shocks quickly into prices so the most up to date information is the most relevant. Both these strategies decrease the number of parameters to estimate, and allow the maximum likelihood estimation to converge.

Specifically, the degree of polynomial I use to weight the 'lead' data points increases as the number of leads increases. For one lead the data is included directly into the MFP, for 2 leads I use a polynomial of degree 1, for 3 to 9 leads I use a polynomial of degree 2 and for 10 to 14 leads I use a polynomial of degree 3. For each model I only include 2 of the lagged bond factors, but all 3 of the lead bond factors. This decision was made after trying the different variations, and was based on the best evaluation results.

1.7.2 US

Figure 1.6 compares the out of sample evaluation statistics of the MFP model containing various numbers of leads, with the MFP model that contains no leads for the 1 month forecast horizon. The top panel of Figure 1.6 divides the QPS from the MFP containing leads by the QPS of the MFP with no leads, for the full out of sample period. The x-axis refers to the number of trading days I use as leads. For example, $lead = 8$ compares the MFP with 8 extra days of bond market data from the current month, to the MFP that contains no leads. A ratio of less than 1 indicates that the extra information from the leads reduces the error rate of the current months recession prediction. The red dot concludes whether the the extra information improves recession probability prediction at the 10% level of significance, according to the Diebold Mariano Test. For all leads examined (1-14) the additional bond market information used to nowcast the current months recession probability reduces the QPS by up to 14%. For 6 of the leads (2, 3, 4, 8, 13 and 14 days) I find that the recession probability predictions are more accurate than the MFP containing no leads, as shown by the red dots. Further analysis also shows a reduction in the false positive that was found in the top panel of Figure 1.3 in 1998. Surprisingly there is no clear relationship between number of leads and model performance, shown by a correlation of 0.11 between the QPS ratio and the number of leads included.

The bottom panel of Figure 1.6 divides the QPS onset from the MFP containing leads, by the QPS onset of the MFP with no leads. For all leads examined, the additional daily bond market data improves the QPS onset error rate by up to 15%. Unlike the top panel, there is a stronger relationship between the QPS onset ratio and the number of leads. The correlation is -0.29, which is still weak but is in the direction I would expect.

1.7.3 Canada

The same process of nowcasting was carried out for Canada using weekly bond market data as leads. When leads of 1,2 and 3 weeks of bond data were included in the model there were no improvements in forecasting performance at the 10% significance level according to the Diebold Mariano test, compared to the forecasts of the MFP used for Canada with no leads. As with the out of sample results this may explain the importance of using daily financial data in the model, and may also indicate that financial markets in Canada incorporate news and economic shocks into financial prices at a slower rate than the US.

1.8 Conclusion

This paper uses factors representing the bond, stock, exchange rate, commodity and real markets estimated from panels of macroeconomic and financial data. These daily, weekly and monthly factors are then used to predict future US and Canadian recession dates, as published by the NBER and the Business Cycle Council of the C.D. Howe Institute respectively, in a mixed frequency probit model. My main findings show that when daily financial data is included in the forecasting model for the US, out of sample predictive performance improves at the 1,2 and 3 month forecasting horizons with reductions of up to 17% in the quadratic probability score (QPS) depending on the forecast horizon. This is compared to aggregating data at the monthly frequency. When focusing on recession onset prediction these improvements increase up to 30%. However, for Canada, where only weekly financial data for the whole sample is available, there are mixed results. I find improved recession onset prediction performance at all forecast horizons, with up to a 22% decrease in the QPS. But evaluation statistics covering the full out of sample period show no improvement in predictive power when compared to probit models that aggregate financial data at the monthly frequency.

Additionally, I use the benefit of daily and weekly frequency data to nowcast the current months recession probabilities, updating forecasts on a daily and weekly basis in the US and Canada respectively. By including the current months financial data into the MFP I find that I can improve forecasting performance measured by the QPS in the US by up to 14%. The Diebold Mariano test also shows statistically significant improvements in forecasting performance at the 10% level, depending how many days of the current months financial data is included in the model. However, there are no significant improvements from nowcasting using financial data in Canada.

By dividing data into the 5 asset markets (4 for the US) and extracting factors from each of these, I am able to gain valuable insights into which markets are key leading indicators for US and Canadian recessions. Results are similar for both the US and Canada. I find that daily and weekly bond market factors are key leading indicators, especially at detecting the onset of a recession, in the US and Canada respectively, compared to aggregated monthly bond market data. The stock and real market also play important roles in improving forecasting performance. However, this is only when the stock market data is aggregated at the monthly frequency, due to daily data causing volatile results. Exchange rate data at any frequency is not useful in predicting future recessions in the US and Canada, and the same can be said for commodity market data for Canada.

1.9 Data Appendix

For US data in Table 1.1, the Bond and Exchange rate market data are from FRED (St. Louis Fed) unless AC (authors calculation) is stated. All macroeconomic indicators are from FRED. The SP500 Industrial Index and SP500 PE ratio are from GFD (Global Financial Data). Finally, the S&P 500 and Dow Jones index closing price is from Yahoo Finance.

For Canadian data in Table 1.2, the Bond market data are from Statistics Canada, Exchange rate data is from FRED, Stock market data is from GFD, Commodity prices are from GFD and all Macroeconomic indicators are from FRED, apart from Housing Starts which is from Statistics Canada. AC indicates authors calculation.

Table 1.1: US Variables, Chapter 1

Variable	Daily	Weekly	Monthly	Description
Bond Market				
Fed Funds 2 (0)	Y	N	Y	Interest Rate: Federal Funds (Effective) (% per annum)
3m Tbill 2 (0)	Y	N	Y	Interest Rate: US Treasury Bills, 3-Mo. (% per annum)
6m Tbill 2 (0)	Y	N	Y	Interest Rate: US Treasury Bills, 6-Mo. (% per annum)
1y Tbond 2 (0)	Y	N	Y	Interest Rate: US Treasury, 1-Yr. (% per annum)
5y Tbond 2 (0)	Y	N	Y	Interest Rate: US Treasury, 5-Yr. (% per annum)
10y Tbond 2 (0)	Y	N	Y	Interest Rate: US Treasury, 10-Yr. (% per annum)
AAA bond 2 (0)	N	Y	Y	Bond Yield: Moody's AAA Corporate (% per annum)
BAA bond 2 (0)	N	Y	Y	Bond Yield: Moody's BAA Corporate (% per annum)
3m spread 1 (0)	Y	N	Y	3m Tbill - Fed Funds (AC)
6m spread 1 (0)	Y	N	Y	6m Tbill - Fed Funds (AC)
1y spread 1 (0)	Y	N	Y	1y Tbill - Fed Funds (AC)
5y spread 1 (0)	Y	N	Y	5y Tbill - Fed Funds (AC)
10y spread 1 (0)	Y	N	Y	10y Tbill - Fed Funds (AC)
AAA spread 1 (0)	N	Y	Y	AAA bond - Fed Funds (AC)
BAA spread 1 (0)	N	Y	Y	BAA bond - Fed Funds (AC)
Exchange Rate Market				
Ex. Rate Switzerland 3 (0)	Y	N	Y	Foreign Exchange Rate: Swiss Franc per US$
Ex. Rate Japan 3 (0)	Y	N	Y	Foreign Exchange Rate: Yen per US$
Ex. Rate UK 3 (0)	Y	N	Y	Foreign Exchange Rate: Cents per Pound
Ex. Rate Canada 3 (0)	Y	N	Y	Foreign Exchange Rate: Canadian$ per US$
Stock Market				
SP 500 3 (0)	Y	N	Y	S&P 500 Index Closing Price
DJ Index 3 (0)	Y	N	Y	Dow Jones Index, Closing Price
SP Industrials 3 (0)	N	N	Y	S&P 500 Industrials Index Closing Price
SP PE ratio 3 (2)	N	N	Y	S&P 500 Index: Price Earnings Ratio (%)
Macroeconomic Indicators				
IPI 3 (1)	N	N	Y	Industrial Production Index, Total Index
PILT 3 (1)	N	N	Y	Personal Income Less Transfer Payments
MTS 3 (1)	N	N	Y	Manufacturing and Trade Sales
Emp: Total 3 (1)	N	N	Y	Employees On Nonfarm Payrolls: Total Private
Housing Starts 3 (1)	N	N	Y	Total New Privately Owned Housing Units Started
Unemp claims: Weekly 3 (1)	N	Y	N	Unemployment Insurance Weekly Claims

1 = no transformation, 2 = first difference, 3 = log of first difference, number in parentheses = publication lag in weeks or months

Table 1.2: Canada Variables, Chapter 1

Variable	Daily	Weekly	Monthly	Description
Bond Market				
10y+ Gov Bond 2 (0)	N	Y	Y	Government marketable bonds average yield, 10 years
5-10y Gov Bond 2 (0)	N	Y	Y	Government marketable bonds average yield, 5-10 years
3-5y Gov Bond 2 (0)	N	Y	Y	Government marketable bonds average yield, 3-5 years
1-3y Gov Bond 2 (0)	N	Y	Y	Government marketable bonds average yield, 1-3 years
3m Prime Corp Paper 2 (0)	N	Y	Y	3 months prime corporate paper
2m Prime Corp Paper 2 (0)	N	Y	Y	2 months prime corporate paper
1m Prime Corp Paper 2 (0)	N	Y	Y	1 month prime corporate paper
10y+ spread 1 (0)	N	Y	Y	Yield Spread b/t 10-yr bond and 3-m prime (AC)
5-10y spread 1 (0)	N	Y	Y	Yield Spread b/t 5-10-yr bond and 3-m prime (AC)
3-5y spread 1 (0)	N	Y	Y	Yield Spread b/t 3-5-yr bond and 3-m prime (AC)
1-3y spread 1 (0)	N	Y	Y	Yield Spread b/t 1-3-yr bond and 3-m prime (AC)
Exchange Rate Market				
Ex. Rate Switzerland 3 (0)	Y	N	Y	Foreign Exchange Rate: Swiss Franc per Can$
Ex. Rate Japan 3 (0)	Y	N	Y	Foreign Exchange Rate: Yen per Can$
Ex. Rate US 3 (0)	Y	N	Y	Foreign Exchange Rate: US$ per Can$
Ex. Rate UK 3 (0)	Y	N	Y	Foreign Exchange Rate: Pound Sterling per Can$
Commodity Market				
Oil Price 3 (0)	Y	N	Y	West Texas Intermediate Oil Price
Gold Price 3 (0)	Y	N	Y	Gold Spot Price, London PM Fixing
Wheat Price 3 (0)	Y	N	Y	Wheat 2 Cash Price
Aluminum Price 3 (0)	Y	N	Y	Aluminum Spot Price
Stock Market				
TSX Index 3 (0)	Y	N	Y	Toronto Stock Exchange, composite index
TSX Value 3 (0)	N	N	Y	Toronto Stock Exchange Index, closing price
TSX Vol 3 (0)	N	N	Y	Toronto Stock Exchange, volume of shares traded
Macroeconomic Indicators				
Consumer Credit 3 (1)	N	N	Y	Consumer Credit, month-end, sa, Total outstanding balances
Manufacturing Prod 3 (1)	N	N	Y	Production in total manufacturing, sa
Emp: Total 2 (1)	N	N	Y	Employed population, total
Housing Starts 1 (1)	N	N	Y	Housing starts 12 month growth

1 = no transformation, 2 = first difference, 3 = log of first difference, number in parentheses = publication lag in weeks or months

Tables

Table 1.3: In Sample, US

Forecast Horizon		1 month	2 month	3 month
QPS Full	$\frac{QPS_{MFP}}{QPS_{benchmark}}$	0.93	0.94	0.88
QPS Onset	$\frac{QPS_{MFP}}{QPS_{benchmark}}$	0.88	0.83	0.76
LPS Full	$\frac{LPS_{MFP}}{LPS_{benchmark}}$	0.88	0.89	0.87
Diebold Mariano	p value	0.15	0.13	0.01

Table 1.4: In Sample, Canada

Forecast Horizon		1 month	2 month	3 month
QPS Full	$\frac{QPS_{MFP}}{QPS_{benchmark}}$	0.88	0.94	0.95
QPS Onset	$\frac{QPS_{MFP}}{QPS_{benchmark}}$	0.95	0.96	0.95
LPS Full	$\frac{LPS_{MFP}}{LPS_{benchmark}}$	0.90	0.94	0.95
Diebold Mariano	p value	0.03	0.08	0.11

In sample results of the best model for the US and Canada at various forecast horizons. Best model of the US consists of 3 bond market factors, 1 stock market factor and 2 real market factors. Best model for Canada consists of 3 bond market factors, 1 stock market factor and 1 real market factor.

Table 1.5: Out of Sample, US

Forecast Horizon		1 month	2 month	3 month
QPS Full	$\frac{QPS_{MFP}}{QPS_{benchmark}}$	0.86	0.83	0.87
QPS Onset	$\frac{QPS_{MFP}}{QPS_{benchmark}}$	0.89	0.70	0.71
LPS Full	$\frac{LPS_{MFP}}{LPS_{benchmark}}$	0.92	0.88	0.90
Diebold Mariano	p value	0.006	0.002	0.019

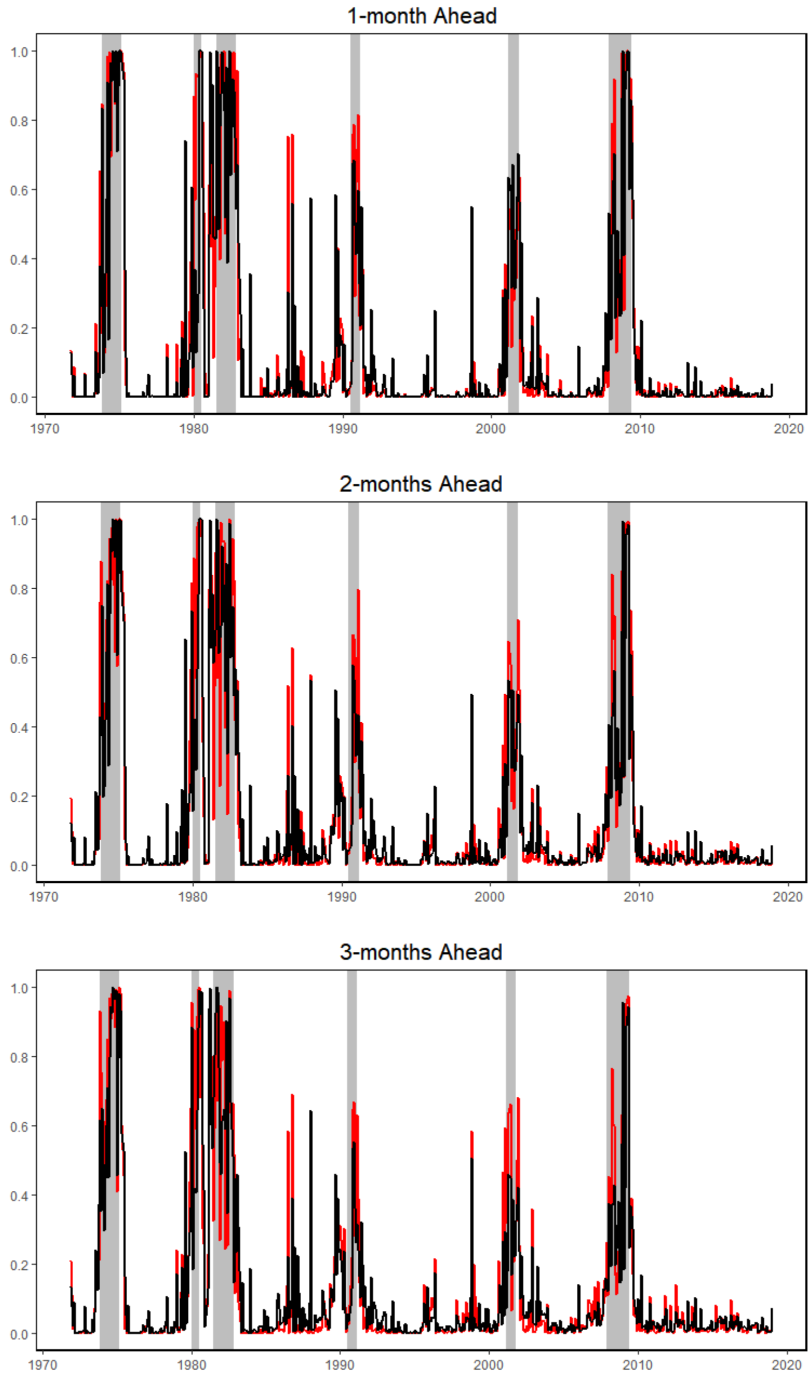
Table 1.6: Out of Sample, Canada

Forecast Horizon		1 month	2 month	3 month
QPS Full	$\frac{QPS_{MFP}}{QPS_{benchmark}}$	1.13	1.13	1.11
QPS Onset	$\frac{QPS_{MFP}}{QPS_{benchmark}}$	0.78	0.83	0.85
LPS Full	$\frac{LPS_{MFP}}{LPS_{benchmark}}$	1.26	1.20	1.25
Diebold Mariano	p value	0.95	0.93	0.88

Out of sample results of the best model for the US and Canada at various forecast horizons. Best model of the US consists of 3 bond market factors, 1 stock market factor and 2 real market factors. Best model for Canada consists of 3 bond market factors, 1 stock market factor and 1 real market factor.

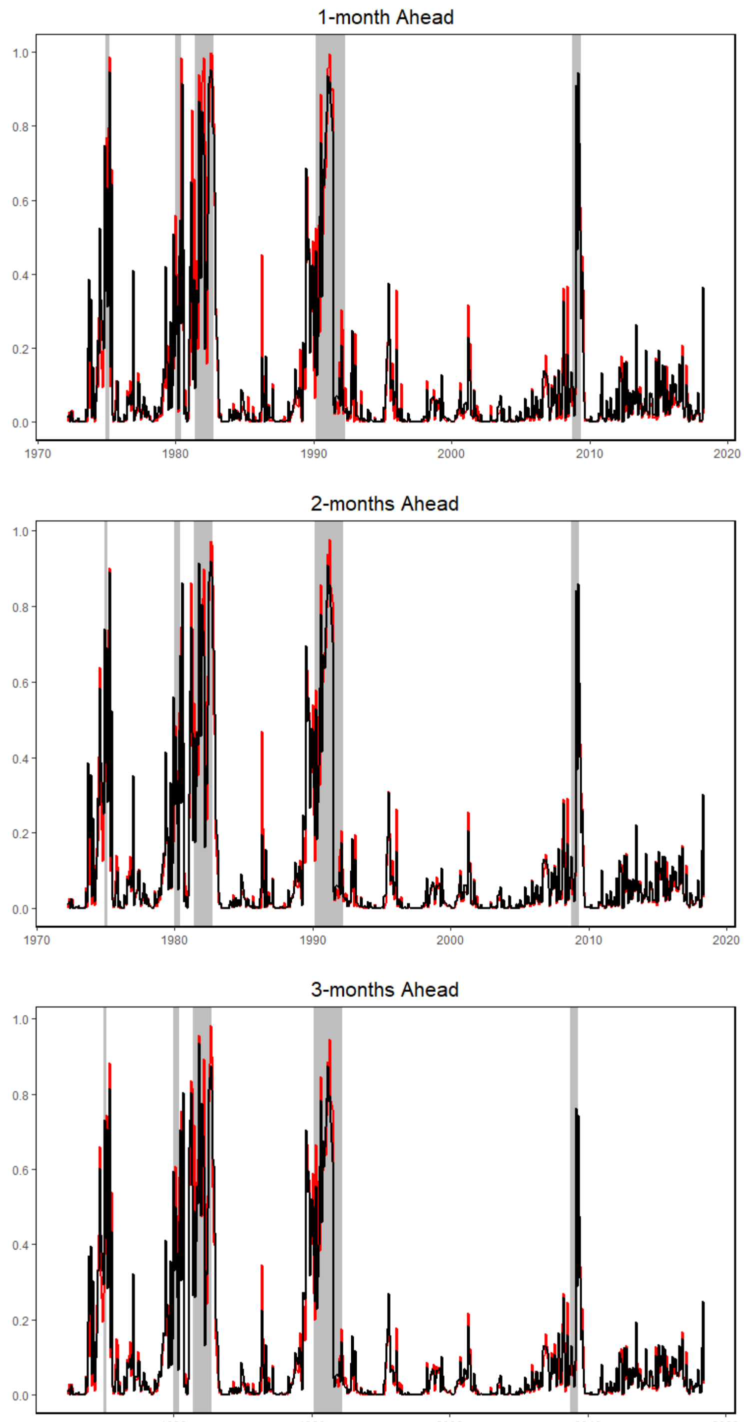
Figures

Figure 1.1: In Sample, US



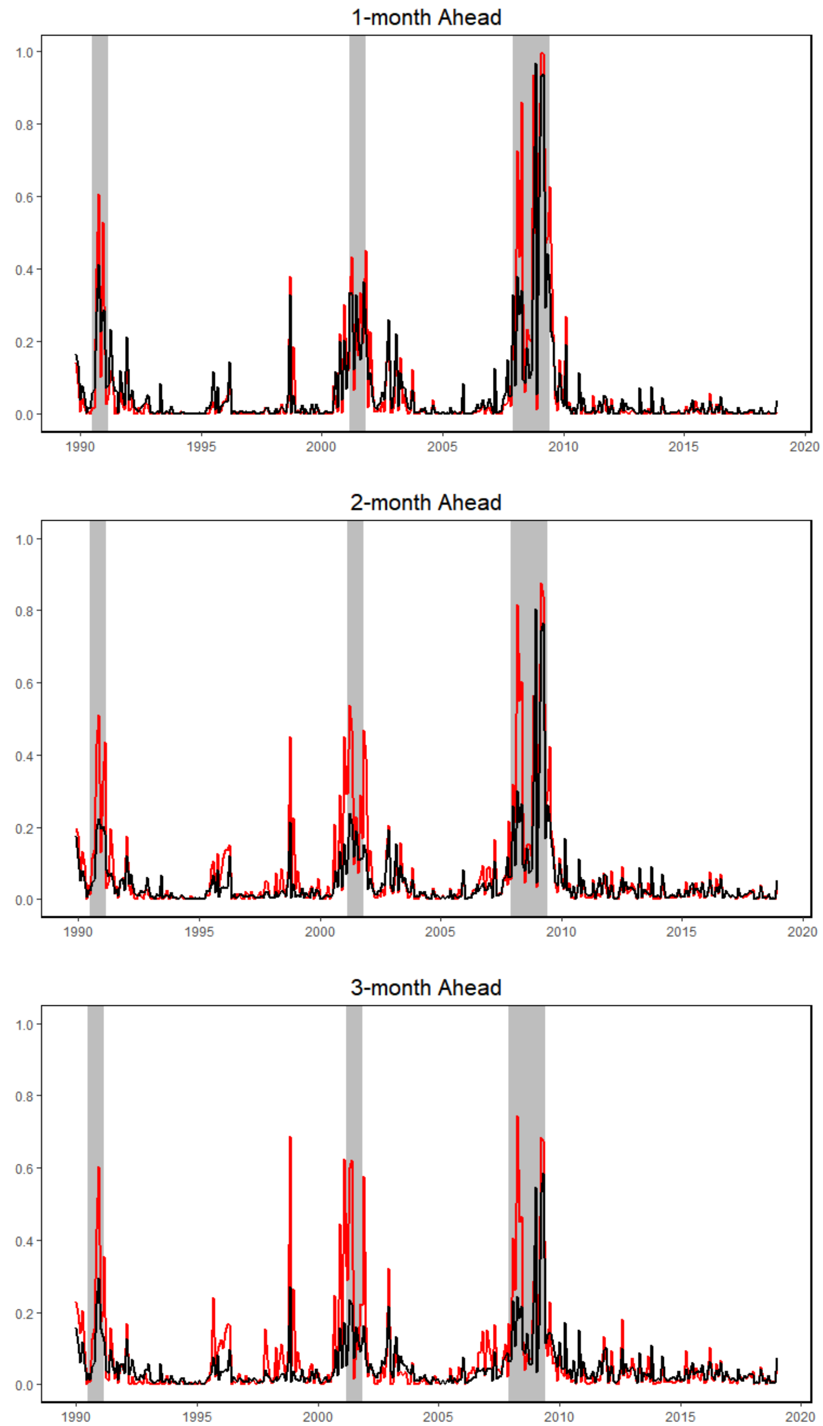
In sample US predicted probabilities of a recession for the best performing probit models at different forecast horizons: mixed frequency probit (red); benchmark model (black). Shaded areas show recessions as defined by the NBER.

Figure 1.2: In Sample, Canada



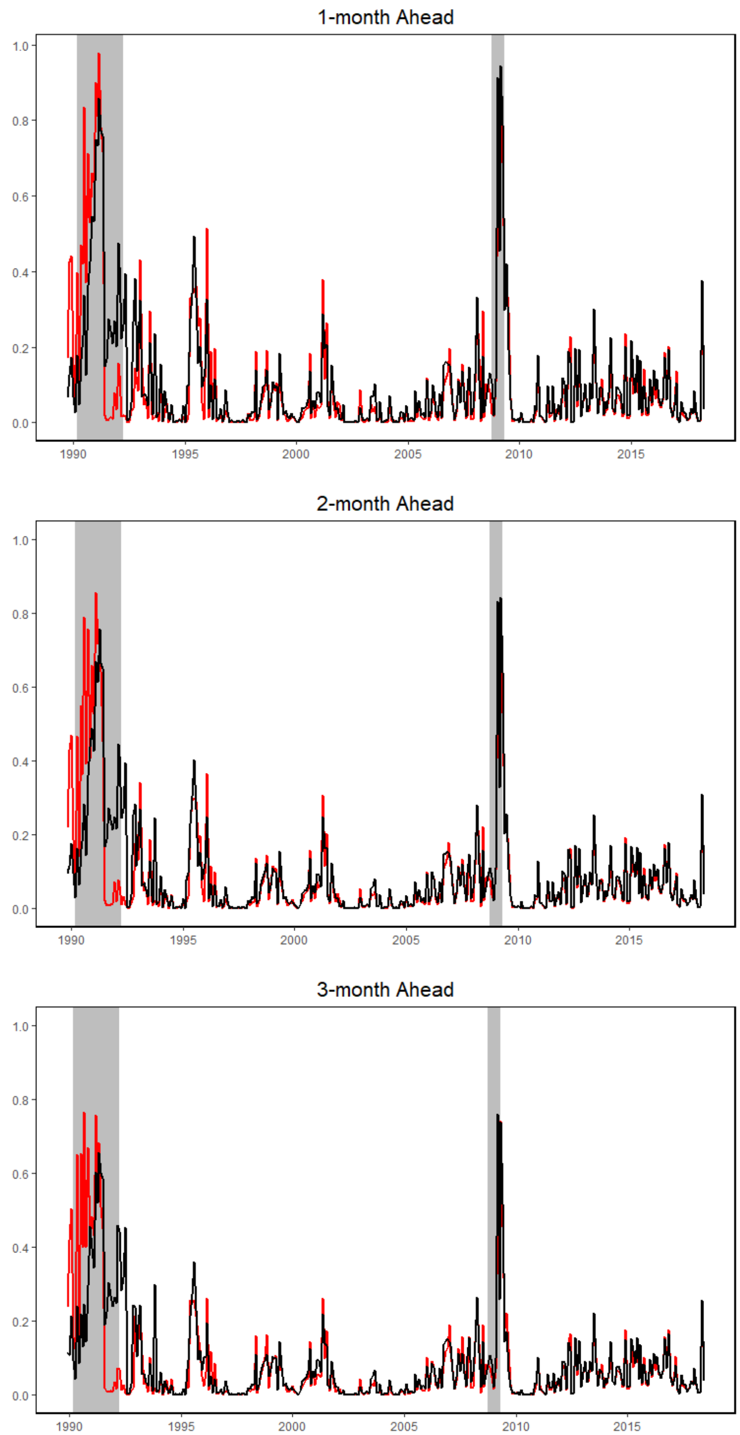
In sample Canadian predicted probabilities of a recession for the best performing probit models at different forecast horizons: mixed frequency probit (red); benchmark model (black). Shaded areas show recessions as defined by the Business Cycle Council of the C.D. Howe Institute.

Figure 1.3: Out of Sample, US



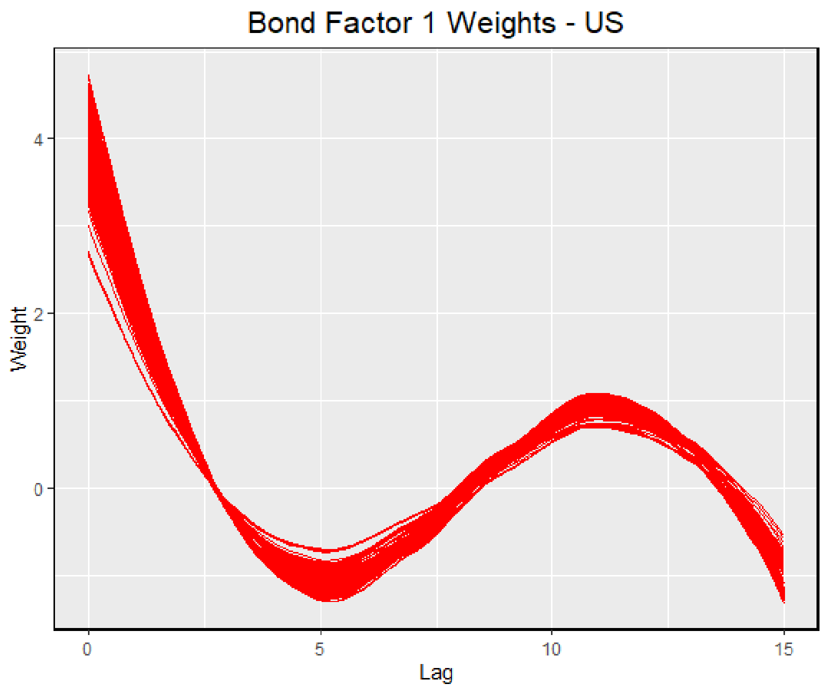
Out of sample US predicted probabilities of a recession for the best performing probit models at different forecast horizons: mixed frequency probit (red); benchmark model (black). Shaded areas show recessions as defined by the NBER.

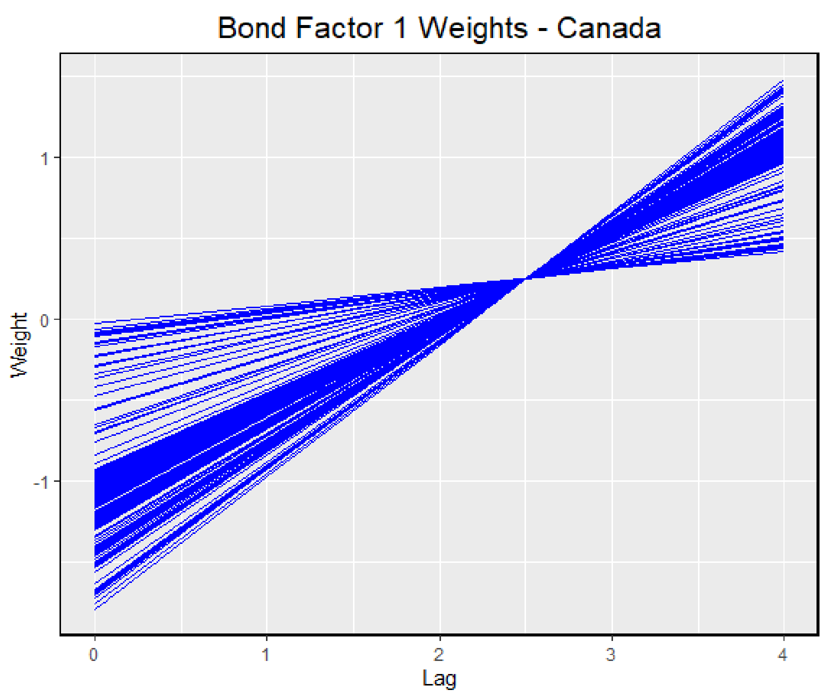
Figure 1.4: Out of Sample, Canada



Out of sample Canadian predicted probabilities of a recession for the best performing probit models at different forecast horizons: mixed frequency probit (red); benchmark model (black). Shaded areas show recessions as defined by the Business Cycle Council of the C.D. Howe Institute.

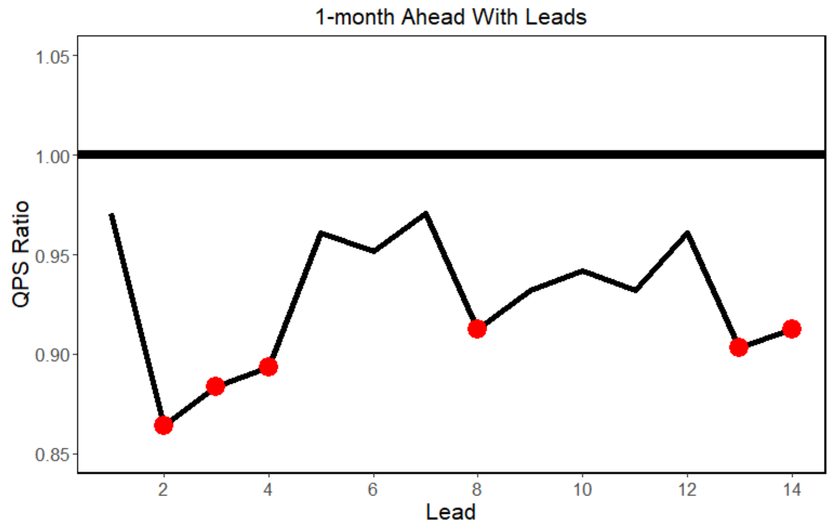
Figure 1.5: Weighting Functions

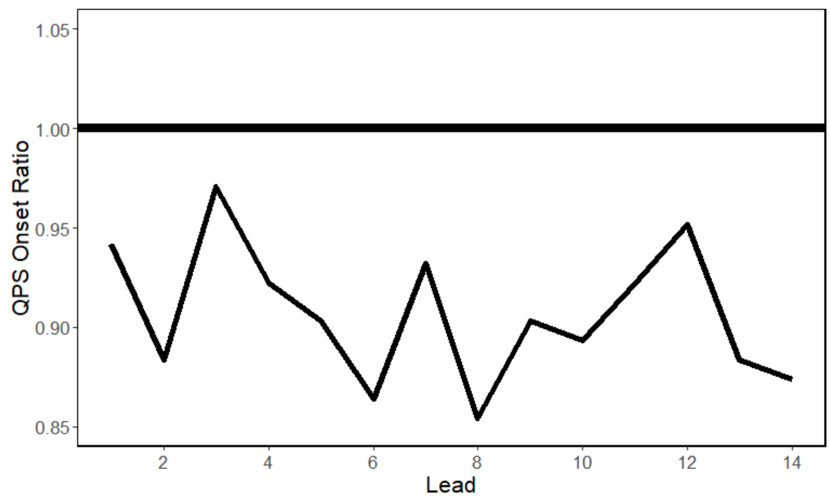




Top: Cubic polynomial weighting functions of the daily bond market factor for the US 1 month ahead out of sample models. Lag refers to the day of the month, with a lag of 1 indicating the most recent data point of the month. Bottom: Linear polynomial weighting functions of the weekly bond market factor for the Canadian 1 month ahead out of sample models. Lag refers to the week of the month, with a lag of 1 indicating the most recent data point of the month. Each line represents the weighting function for each out of sample model.

Figure 1.6: Nowcasting Results, US





Top: US out of sample QPS of the MFP nowcast with leads model divided by QPS of the MFP model with no leads at the 1-month forecast horizon. Red dots show when the nowcast with leads model is more accurate than the MFP 1-month ahead model at the 10% significance level, as per the Diebold Mariano Test. Bottom: Out of sample QPS onset of the MFP nowcast with leads model divided by QPS onset of the MFP model with no leads at the 1-month forecast horizon. Lead refers to the number of days of financial data into the current month used to nowcast the current months recession probability. Horizontal line at 1 used to show when the QPS of the nowcast with leads model is lower than QPS of MFP with no leads model.

Forecasting US Recessions Using Mixed Frequency Artificial Neural Networks

2 Chapter 2

2.1 Introduction

Forecasting recessions using reduced form approaches such as probit models has been a popular approach in the past literature. These methods can be successful if we assume a linear decision boundary to classify the binary recession indicator. However, in reality the relationship between the state of the economy with the financial and macroeconomic markets is likely to be non linear, and therefore probit models may not be the optimal forecasting methodology. Machine learning techniques, specifically artificial neural networks (ANN), contain hidden layers which allow for this non linear relationship and a more flexible non linear decision boundary to be created in order to separate the two classes of data.

There is limited recession forecasting literature using machine learning techniques. Arguably the most notable is Qi (2001) who finds that the S&P 500 index and yield spread are the best leading indicators when using a neural network to forecast the probability of a US recession. When extending this niche area of the literature further to include mixed frequency data in a ANN there has not been anything to my knowledge, with the only application of a mixed frequency ANN (MF-ANN) by Xu (2019), who applies the model to forecasting inflation in China.

Therefore, this paper adds to the literature in multiple ways: (1) I use NBER defined recession indicators as a binary dependent variable in a mixed frequency artificial neural network (MF-ANN), taking advantage of the readily available daily, weekly and monthly financial and macroeconomic data. I then forecast recession probabilities up to a 3 month horizon. (2) I introduce a new method to include mixed frequency data in a classification ANN that could be useful for topics outside of recession prediction. (3) Part of my analysis focuses on recession 'onset' prediction, defined later as the first 5 months of a recession. Arguably this is the most important time to accurately predict a recession due to the considerable reporting lag by the respective business cycle committees. (4) I implement a novel method to tune hyperparameters and select variables in an ANN with time series data.

In summary this paper extends on the limited literature of the intersection between machine learning and recession forecasting by incorporating mixed frequency data into a feed forward neural network to forecast US recessions 1 to 3 months ahead. Following the method derived by Xu (2019) to include mixed frequency data in a neural network I find that in sample a MF-ANN can drastically improve performance at all forecast horizons, in terms of forecasting accuracy, compared to a benchmark matched frequency ANN and a benchmark mixed frequency probit (MFP) model from Mitchell (2021). These improvements are as high as 81% when looking at the forecast error.

Out of sample results also show improvement of mixed and matched frequency neural networks over reduced form models, with the reduction of forecast errors up to 54%. However, results of the MF-ANN over a benchmark matched frequency ANN are mixed. Although at the shorter forecast horizon the mixed frequency model performs better, at the longer 2 and 3 month horizons the results favor, or are indifferent, for the matched frequency model. Finally, I find that daily US bond market factors, monthly US stock market factors and monthly US real activity factors are the best leading indicators for US recessions.

2.2 Literature Review

2.2.1 Reduced Form

Reduced form methods for forecasting recessions focus on using binary models, for example the probit model, to produce future recession probabilities. This has primarily centered on matched frequency data using a few key variables. For example, Estrella and Mishkin (1998) find that the 3-month less 10-year term spread and stock price indices are the most useful predictors of future US recessions. Similarly, Wright (2006) finds that using the level of the federal funds rate together with the term spread improves the performance of the predictive probit models. Katayama (2009) analyzed the performance of several binary class models for NBER recessions using combinations of 33 macroeconomic and financial indicators. He finds that the combination of the term spread, month to month changes in the SP 500 index and the growth rate of non-farm employment generate the sequence of out of sample recession probabilities that better fits NBER recession dates. This then extended to use factor analysis to allow for a larger set of predictor variables. Most notably Fossati (2015) uses dynamic factors estimated from panels of macroeconomic and financial indicators to predict future recessions using probit models. He concludes that at 3 month forecast horizons probit models that include the 3-month less 10 year term spread, a stock market dynamic factor and a real economic factor achieve the best out of sample fit. More recently, Huang and Startz (2020) find that augmenting existing Markov-switching dynamic factor models with additional information on the stock return volatility improves prediction of the state of the economy. They beat both the peak and trough announcements for recent recessions by the NBER by several months.

The literature on forecasting recessions with mixed frequency data in reduced form models is much more limited. Balcilar (2016) uses a mixed frequency Markov switching vector autoregressive (MF-MS-VAR) model to predict regimes in quarterly US GDP using the monthly economic policy uncertainty index as the leading indicator. They find that their model performs better out of sample with mixed frequency data compared to that of matched frequency. Bessec (2015) uses a Markov switching factor mixed data sampling (MS-factor MIDAS) model to extract probabilities in turning points of quarterly US GDP using monthly financial variables. They also conclude that economic turn points are detected more successfully with mixed frequency models versus models that aggregate the higher frequency data.

Camacho (2014) use a Markov switching dynamic factor model to forecast quarterly GDP of the euro area, using a mix of quarterly and monthly variables. Filtered probabilities are then extracted and updated whenever new data is released to forecast economic turning points in the euro area in real time. Bessec (2015) uses monthly bond market, stock price and oil price data to forecast turning points in UK quarterly GDP. They find the Markov switching MIDAS model performs better than a matched frequency model in determining US economic turning points, but has little benefit for forecasting recessions in the UK versus matched frequency. Foroni (2015) uses a MF-MS-VAR to forecast quarterly GDP growth in the euro area using four monthly indicators; the Economic Sentiment Indicator (ESI), the M1 monetary aggregate, headline industrial production and the slope of the yield curve. They find their model works well for nowcasting and short term forecasting of the euro area GDP growth.

2.2.2 Machine Learning

Overall the literature that uses artificial neural networks to forecasts recessions is sparse. In terms of using mixed frequency data I have found no papers. The only paper related to neural networks and mixed frequency data, to my knowledge, is by Xu (2019). I follow this methodology, explained in section 2.4.2. Below is a summary of machine learning in recession forecasting literature.

Using ANNs to forecasts recessions starts with Vishwakarma (1994), who analyzed a four layer feed forward neural network. They used manufacturing trade and sales, personal income, industrial production and employment as the input variables to date US business cycle regimes from 1965-1989. While this model produced a business cycle dating chronology closely matching that of NBER, the study published in sample results only.

The next known paper to forecast recessions using neural networks was Qi (2001), who used a feed forward neural network with one hidden layer to predict US recessions 1 to 8 quarters ahead. They only used 2 variables at any one time in their network, and found that different combinations of the yield spread, S&P 500 Index, Stock and Watson (1989) index, real money supply and Department of Commerce leading index gave the best out of sample results.

Kiani (2008) employs a neural network with selected economic and financial variables to forecast recessions in Canada, France, Germany, Italy, Japan, UK and the US. Using a wider range of variables they are able to beat the results of Qi (2001), finding the stock price index and spread between bank rates and risk free rates are the most likely candidate variables for forecasting recessions 1 to 10 periods ahead.

Zhang (2010) uses a three layer feed forward neural network to forecast turning points in the business cycle of China. Out of sample results show that some financial and economic indicators, such as steel output, Pig iron yield and the freight volume of the entire society are useful leading indicators.

Giusto and Piger (2017) propose a simple machine-learning algorithm known as Learning Vector Quantization (LVQ) for the purpose of identifying new US business cycle turning points quickly in real time. Despite its relative simplicity, the algorithm's performance appears to be very competitive with those of commonly used alternatives.

Jackson (2019) describes two machine learning methods, K-Nearest Neighbor and Neural Networks, and compares them to a Dynamic Factor Markov Switching model for determining business cycle turning points. They conclude that machine learning techniques can offer more accurate classifiers that are worthy of additional study.

Finally, Puglia and Tucker (2020) use machine learning methods to examine the power of Treasury term spreads and other financial market and macroeconomic variables to forecast US recessions.

2.3 Motivation

2.3.1 Why use Mixed Frequency Data in Recession Forecasting?

Financial assets prices, in theory, should contain information about the future state of the economy and therefore should be considered as extremely relevant for macroeconomic forecasting, Andreou, Ghysels and Kourtellos (2013). Usually researchers will aggregate financial time series to match the dependent variable frequency, for example in the case of recession forecasting all financial data would be converted to monthly frequency. This could mean taking the log of stock returns over a whole month or averaging bond yields over the month versus using a higher frequency of the data such as daily.

Andreou, Ghysels and Kourtellos (2013) explain that not using the readily available high frequency data such as the daily financial predictors to perform the monthly forecasts of recession probabilities has two important implications: (1) you lose the possibility of having real time daily, weekly or bi-weekly updates of the recession probabilities and (2) you lose information through temporal aggregation. In other words, important within month information that can help in predicting current and future recessions may be 'washed' out by the assumption of equal weighting of data.

Andreou, Ghysels and Kourtellos (2010) show that the estimated slope coefficient of a regression model that imposes a standard equal weighting aggregation scheme, ignoring the fact processes are generated from a mixed data environment, yield asymptotically inefficient and inconsistent estimates. Both these asymptotic inefficiencies and inconsistencies can have adverse effects on forecasting.

2.3.2 Why use a Neural Network?

The most common way to forecast recessions in previous literature has been to use probit and logit regressions. In the past literature most of these binary reduced form approaches only allow for a linear decision boundary to classify binary data. For example, in a 2D world this is fine if a straight line can be drawn to divide up the non recession periods from the recession periods, based on some explanatory variables. However, in reality the relationship between the binary Y variable and X variables may be non-linear. Of course non-linear terms can be added directly to the reduced form regression, but the question then arises what and how many non-linear terms should be added. This becomes additionally complex when the researcher is limited with respect to the number of terms that can be added due to a relatively small sample size. Whereas the hidden layer of a neural network creates this non linear relationship without any need of prior knowledge and allows for a more flexible non-linear decision boundary to be created in order to separate the two classes of data.

Secondly, a neural network allows for easy inclusion of high frequency data. Reduced form regression approaches have a degrees of freedom limit and therefore a constraint must be placed on the higher frequency data, for example a polynomial weighting function. Whereas with neural networks the higher frequency data can be directly added to the model with no need to reduce the number of variables by applying restrictions on the data. This is explained more in Section 2.4.2. Additionally, due to the same reasoning, more features can be included in a neural network to explain Y then would be allowed in a reduced form model.

There are many other machine learning methods that could also be used for classification. However, I choose to explore the use of ANNs due to the rapidly growing interest in the methodology and its proven performance in pattern recognition and detecting complex non-linear relationships in a large amount of variables as shown by Goh (1995) and Chong et al. (2017). Xu (2019) also explains that ANNs have been successfully applied to process non-linear issues in many areas of the economy and finance such as volatility, credit scores, risk and many more (see Pradeepkumar (2017), Beque (2017) and Xu (2016)). Therefore ANNs seem a natural methodology to use when forecasting recessions with a large amount of variables, as is this case when using mixed frequency data. In this paper I solely look at feed forward ANNs, but other ANNs such as recurrent neural networks have been shown to be successful with forecasting problems and therefore would be a natural extension to this paper. Additionally I briefly examined the machine learning methods of Random Forests and Support Vector Machines against ANNs, and for the purpose of forecasting recessions ANNs produced the smaller out of sample forecast errors, but these should be looked at in more detail in the future.

2.4 Methodology

2.4.1 Feed Forward ANN

Artificial neural networks (ANN) are useful for forecasters, especially where data is available but the data generating process is unknown. ANN's are treated as non-linear, non-parametric statistical methods due to which these are independent of the distributions of the underlying data generating processes (White, 1989). A general matched frequency form of a neural network with one hidden layer is

$$f(x) = sig\left[\alpha_0 + \sum_{j=1}^{n} \alpha_j sig\left(\sum_{i=1}^{s} (\beta_{ji} x_i + \beta_{0j})\right)\right] + \epsilon \tag{27}$$

where,

$n =$ the number of hidden nodes in the neural network

$s =$ the number of explanatory variables in the neural network

$sig(a) = 1/(1 + e^{-a})$ is the activation function converting the output to a probability

α_j for $j = 1, .., n$ is a vector of parameters from the hidden to the output layer

β_{ji} for $j = 1, .., n$ and $i = 1, .., s$ is a matrix of parameters from the input to the hidden layer

α_0 and β_{0j} for $j = 1, .., n$ are constant weights, also called the bias

x_i for $i = 1, ..., s$ are the explanatory variables

$\epsilon =$ error term

To make this specific to recession prediction, the forecasted probability of a recession h months ahead, $\hat{y_{t+h}}$, if we are standing at time t using the vector x as the leading indicators, is given by

$$\hat{y_{t+h}} = sig\left[\hat{\alpha}_{0,t} + \sum_{j=1}^{n} \hat{\alpha}_{j,t} sig\left(\sum_{i=1}^{s} \left(\hat{\beta}_{ij,t} x_{i,t} + \hat{\beta}_{0j,t}\right)\right)\right] \tag{28}$$

where the sigmoid function forces the output to be a probability between 0 and 1.

To estimate the α and β weights back propagation is used. With this technique initial values are set for the parameters, the data is passed through the network and an error term is calculated. The gradient of the loss function with respect to the weights of the network is calculated layer by layer, adjusting the weights to reduce the error term using gradient descent. This algorithm repeats itself for a programmed number of times. However, this algorithm may over-fit the model if n (number of nodes in the hidden layer) as well as the number of hidden layers is too large, causing small in sample errors but large out of sample errors. The choice of n and hidden layers is usually chosen depending on the number of explanatory variables used. In my paper I will formalize an approach to tune both these and other hyper-parameters.

2.4.2 Including Mixed Frequency Data in the ANN (MF-ANN)

In order to include mixed frequency data in the network, I will follow the methodology proposed by Xu (2019). They develop a simple ANN model for mixed frequency data through introducing a mixed data sampling (MIDAS) and unrestricted MIDAS (U-MIDAS) approach into the ANNs framework. In their paper the model is applied to an empirical task of forecasting inflation in China, and improvements are found in forecasting performance compared to a model that aggregates data into the same frequencies.

Figure 2.1 shows the mixed frequency feed forward ANN proposed by Xu (2019). X in the input layer represents the chosen variables used for the analysis. The frequency alignment layer changes the dimension of the higher frequency data so that it is the correct dimensions to run through the network. For example, if I use daily financial data to predict the monthly recession indicator then I will have K days of explanatory data per month. Originally X is a vector of length $(T \times K)$ where K represents the number of higher frequency data points in the lower frequency time frame, and T represents the number of lower frequency data points in the sample. The frequency alignment step transforms this vector to a $(T \times K)$ matrix, where the first column represents the most recent trading days data of each month, the second column represents the second most recent trading days data of each month and so on.

Previously in the mixed frequency probit (MFP) I was restricted by the number of variables I could include in the regression, hence I used a polynomial to weight the higher frequency data. This greatly reduced the number of parameters to estimate. With the ANN I do not necessarily have the degrees of freedom concern. I can therefore add the higher frequency daily data directly into the network as its own variable. For example, using the unrestricted MIDAS approach proposed by Xu (2019) with 10 daily financial variables as my X, and $K = 15$, would mean 150 inputs into the mixed frequency ANN. With the MFP this would cause extreme in sample over-fitting, which is why a polynomial would be used to weight the higher frequency data. Xu (2019) calls this the U-MIDAS ANN approach as we are not restricting the weighting functions for the higher frequency data. Theoretically the network will decide which days of the month are the most important in predicting recessions, and weight them accordingly.

Finally, the activation function used is a sigmoid function transforming the output to a value between 0 and 1, representing the probability of a recession. After data is passed through the network, the cross entropy loss function, explained in section 2.4.3, is calculated and back propagation is used to train the network by adjusting the weights and bias.

2.4.3 Training the MF-ANN

The cross entropy loss function used in the back propagation training of the model is also called the log probability score (LPS), which is given by

$$LPS = -\frac{1}{T}\sum[y_{t+h}log(\hat{p}_{t+h}) + (1 - y_{t+h})log(1 - \hat{p}_{t+h})] \tag{29}$$

where T is the number of forecasts, $\hat{p}_{t+h}$ is the predicted probability of a recession for month $t + h$ for a given model, and y_{t+h} is the realized recession indicator in the month $t + h$. The LPS can take values from 0 to $+\infty$ where smaller values indicate more accurate predictions. The LPS score penalizes large errors more heavily.

2.4.4 Tuning the Hyper-parameters of the MF-ANN

One large and important question that the machine literature has no definite answer to is how should the hyper-parameters of the neural network be tuned when we have time series data? Key hyper-parameters of my network include; the number of hidden layers, the number of nodes in each hidden layer, the regularization parameter to avoid over-fitting (decay) and the number of iterations used to train the network. Puglia (2020), who uses yield curve data to forecast US recessions in a neural network, found that changes in these hyper-parameters can lead to drastically different forecasts. Therefore it is important to come up with a methodology to select the optimal combination of hyper-parameters to be used in the final predictive model.

To do so I follow, and slightly adapt, the method used by Puglia (2020) when forecasting recessions using machine learning methods. The overall idea is to use a nested cross validation approach combined with a grid search to find the set of hyper-parameters that yield the lowest forecast errors and should therefore be used in the final predictive model. Below I outline the main steps in my hyper-parameter tuning procedure.

1) Create a grid of potential hyper-parameter combinations to be used in the neural network.

2) Split my time series into 6 data folds of roughly equal size making sure to keep the ordering of the time series. In this case each fold contains one US recession, apart from the first fold which contains 2 recessions. See Tables 2.3 and 2.4 for the exact dates for each block.

3) If I want to find what hyper-parameters to use in the final model when predicting recession probabilities for the last block, which includes the recession caused by COVID-19, I implement the following steps.

[**Part i**] Train the model on block 1 and test on block 2 for each hyperparameter combination. Save the forecast errors given by the log probability score (LPS) for each hyper-parameter combination. Train the model on block 1 and 2 and test on block 3. Save the forecast errors for each hyper-parameter combination. Train the model on block 1, 2 and 3, and test on block 4. Save the forecast errors for each hyper-parameter combination. Train the model on block 1, 2, 3 and 4, and test on block 5. Save the forecast errors for each hyper-parameter combination.

[**Part ii**] This gives 4 forecast errors for each hyper-parameter combination. I take the average over the 4 errors and select the hyper-parameter combination with the lowest error.

[**Part iii**] Run parts i and ii n times to make sure hyper-parameter selection is stable. Results can differ between each run due to different weight initialization in the neural network package in R.

4) The hyper-parameter combination selected from the previous step is then used in the out of sample forecasting for recession probabilities for block 6. Every time there is a new recession in the US a block can be added to this procedure and the hyper-parameters can be re-tuned.

5) Steps 3 and 4 can then be generalized to find the hyper-parameters for any block/fold in my data sample.

As Puglia (2020) mention in their paper, this division of data and cross validation strategy represents the best balance that can be achieved between; 1) size of the training/test sets on hand, 2) sample size of the outer loop forecast accuracy estimate and 3) recession indicator imbalance in the data. However, one downside to this methodology is that more weight is put on earlier recession periods when tuning the hyperparameters for current recession periods.

2.5 Data

My entire data sample period for Y runs from October 1971 to December 2020, which includes the last 7 US recessions . The Y variable is a monthly binary indicator that takes the form:

$$y_t = \begin{cases} 1, & \text{if NBER defined recession.} \\ 0, & \text{otherwise.} \end{cases} \tag{30}$$

A list of X variables that will be used as potential features in the ANN can be found in Table 2.1 and 2.2 in the data appendix. Data in Table 2.1 has been shown to be good predictors of US recessions in the past literature, for example Fossati (2015), Camacho (2014), Chauvet and Piger (2008) and many more. Due to an ANNs ability to use a large volume of data I have also included Canadian financial and macroeconomic indicators in Table 2.2 as potential predictors for US recessions. Historically the financial markets of US and Canada have been highly integrated and Fossati (2018) found that US financial markets can be useful in forecasting Canadian recessions.

However, even though ANN's are known for being able to handle a large number of features it may not be beneficial to blindly include all features. As Kordos (2016) explains, there are 3 main reasons for data selection: It limits the data set size and thus accelerates the model learning process, it removes noise from the data and thus improves the model predictive capabilities, and it can make the data interpretation easier by humans. As explained in Section 2.4.2 the number of features increases exponentially as you include additional predictors for the MF-ANN, especially if these are daily frequency predictors. I will therefore follow a simple forward selection wrapper method to determine which variables to include in the ANN, Reif et al (2014). The forward selection wrapper method will split the time series into the same time blocks used for the hyperparameter selection explained in Section 2.4.4. However, one question is what comes first; hyperparameter tuning or feature selection? These processes can be incorporated into the same algorithm, however this is extremely computationally expensive. I therefore carry out feature selection first using an ANN with one hidden layer with the number of nodes equal to 2/3 of the variables used in the model. For example, if I am testing an ANN with 3 monthly bond market factors I will include 2 neurons in the hidden layer. This is a generally excepted rule of thumb, see

Karsoliya (2012). Below is the feature selection algorithm.

1) Initially the model includes each possible predictor on its own and is trained using block 1 of the data. The accuracy of the model is then evaluated using block 2 of the data, and the predictor that produces the lowest forecast accuracy is kept.

2) New models are then trained on block 1 of the data with the predictor selected from the first stage and every other possible predictor. The accuracy of the model is again evaluated using block 2 of the data, and the feature that produces the lowest forecast accuracy is kept as an additional variable.

3) Steps 1 and 2 are repeated until there is no further improvement in the forecast error. These features will be used in the final predictive model on block 3 of the data.

4) The next stage is to repeat steps 1-3, but use blocks 1 and 1+2 of the data to train the model and block 2 and 3 of the data to evaluate the forecast error. Features found from this step with the lowest average forecast error will be used in the final predictive model on block 4 of the data.

5) As the data contains 6 blocks the process will continue until the final features are selected for all the blocks that will be used for the out of sample analysis (blocks 3-6).

Prior to the feature selection algorithm, data is first transformed to be stationary - see the data appendix for relevant transformations. Since real activity variables are usually available with some lag, I account for data availability at time t by using the last known value $x_{i,t-l}$, where l indicates the publication lag of variable i. Publication lags for US indicators are adopted from Katayama (2009) and for Canadian real activity indicators they are obtained from Statistics Canada. The lags are presented in the data appendix.

The data is then split into separate asset markets; (1) separate bond markets for the US and Canada, (2) separate exchange rate markets for the US and Canada, (3) separate stock markets for the US and Canada, (4) separate real activity markets for the US and Canada, and (5) a commodity market for Canada. Factor analysis is then run on each of these 9 markets to reduce the dimensionality and noise of the data, as opposed to including each variable as an individual predictor variable. This has the additional benefit of identifying which asset markets are key leading indicators for US recessions. The factor analysis model takes the following form

$$X = \Lambda F + \epsilon \tag{31}$$

Where X is a $(N \times T)$ matrix of observed data and Λ is an $(N \times M)$ matrix of factor loadings, where M corresponds to the number of common factors being estimated. F is a $(M \times T)$ matrix of latent factor scores and ϵ is the error term. There is no strict rule on deciding how many factors to use, but I choose as many factors needed to explain the majority of the data variance. There is therefore 9 potential features (which can be split further by frequency) to be chosen from using the forward selection wrapper method. I find that for each out of sample block of data the best features to use are; daily US bond market factors consisting of 3 factors, monthly US stock market factors consisting of 1 factor and monthly US real activity factors consisting of 2 factors. Each daily mixed frequency bond market factor for the neural network consists of 15 trading days worth of information for each month (K=15). In this paper I do not look at the optimal choice of K, and use the highest K possible given the data restrictions.

For direct comparison of the results from my MF-ANN I use a benchmark ANN model which aggregates all higher frequency data into the lowest frequency - monthly. I will also compare results from Mitchell (2021) who completes the same analysis of recession prediction using a mixed frequency probit model (MFP). This will give a good idea of the prediction advantage of machine learning over reduced form techniques when using mixed frequency data.

2.6 Results

2.6.1 Forecast Evaluation

The first method used to evaluate forecast accuracy of the MF-ANN and benchmark models is the Quadratic Probability Score (QPS). The QPS is equivalent to the mean square area when using classification models and is defined by

$$QPS = \frac{2}{T} \sum [\hat{p}_{t+h} - y_{t+h}]^2 \tag{32}$$

where T is the number of forecasts, $\hat{p}_{t+h}$ is the predicted probability of recession for month $t+h$ for a given model and y_{t+h} is the realized recession indicator in the month t+h. The QPS can take values from 0 to 2 with smaller values indicating more accurate predictions.

As discussed in the Introduction, the business cycle dating committee of the US delay their announcements of recession periods. Arguably the most important time for economic agents to know whether the economy is in a recession is the first few months, which is not always immediately obvious by observing single economic indicators. Therefore a model that is able to identify whether the economy is in a recession in the early stages is beneficial to economic agents. I devise an additional evaluation measure which slightly adapts equation (32). I call this measure QPS_{onset}, where *onset* refers to the onset of a recession which I define as the first 5 months of a recession period. In other words, I want to see if the MF-ANN has particularly good prediction power at the beginning of recession periods, versus that of the benchmark model.

$$QPS_{onset} = \frac{2}{T_{onset}} \sum [\hat{p}_{onset,t+h} - \mathbf{1}]^2 \tag{33}$$

T_{onset} refers to the number of forecasts during recession onsets, for example, if there are 3 recessions in the period being examined, $T_{onset} = 15$. $\hat{p}_{onset,t+h}$ is the predicted probability of a recession for month $t+h$, when period $t+h$ falls within the first 5 months of a recession. Finally, $\mathbf{1}$ is a vector indicating that the true state of the economy is a recession.

The third evaluation method I use is the log probability score (LPS) also called the cross entropy loss function. This is the same loss function used for training the network and is explained in 2.4.3. Compared to the QPS, the LPS score penalizes large errors more heavily.

For each of these first 3 evaluation methods I divide the statistic from the MF-ANN model by that of the benchmark model. A value less than 1 indicates that the MF-ANN is a better model. The amount below 1 can be interpreted as the % improvement from using the MF-ANN versus the benchmark model. For example, a value of 0.8 indicates a 20% improvement in the evaluation statistic.

The fourth evaluation method I use is the Diebold-Mariano test with a squared loss function. This test compares the predicted probabilities of the MF-ANN and the benchmark model with the actual values of Y that occurred. For the purpose of this paper the H_A is that the MF-ANN is more accurate than the benchmark model, and hence a low p-value is desired. I choose a squared loss function, as opposed to a linear loss function, as I deem incorrect recession predictions to be costly to the economy and therefore want to penalize these at a higher rate.

2.6.2 Hyper-Parameter Selection

Table 2.3 and 2.4 shows the results from the hyperparameter tuning method for the mixed and matched frequency ANN, displaying the final hyperparameters that will be used in the predictive model for that specific block and forecast horizon. The whole data set was split into 6 blocks and hyperparameter selection was focused on blocks 3, 4, 5 and 6 as these form the out of sample period. The grid search ranged from 1 to 2 for the hidden layers. Heaton (2008) states that neural networks with two hidden layers can represent functions with any kind of shape and therefore the benefit of going beyond this is extremely diminishing. The range for the number of nodes in each hidden layer was selected to be between 1 and 5. Again, Heaton (2008) lays out several rules of thumb to determine this number, of which this range satisfies. Additionally going beyond this was computationally expensive and at sometimes not available using the neural network algorithms in R. Finally the decay parameter was set between 0 and 0.4 and the number of iterations between 400 and 600. This was found to to be the best ranges in terms of the trade off between fitting the best hyperparameters and the computational time to run the algorithm. When selecting the hyperparameters the algorithm was repeated 30 times for each selection procedure to check that results were stable between runs, where the most common hyperparameter combination was chosen.

For all models 1 hidden layer was favored over 2 hidden layers, and interestingly the number of nodes selected within the hidden layer was consistently less for the mixed frequency network over the matched frequency network. This goes against the rule of thumb that states nodes in the hidden layer should be larger when you have more variables. I also find that the decay parameter is on average higher for the mixed model compared to the matched model. Intuitively this makes sense as the mixed model is more prone to overfitting due to the higher number of variables. Finally the number of iterations to train the model is similar between both networks.

2.6.3 In Sample

The in sample period runs from October 1971 to December 2020, for which I forecast recession probabilities at alternative horizons: $h = 1$, 2 and 3 months. I use the hyperparameters selected in Block 6 from Tables 2.3 and 2.4, as these represent the optimal parameters that should be used for the whole sample. Additionally I compare my models to a reduced form benchmark; a mixed frequency probit (MFP) model used to forecast US recessions from Mitchell (2021).

Figure 2.2 shows the in sample recession probabilities for forecast horizons 1, 2 and 3 months. The red line shows recession probabilities calculated from the MF-ANN, and the black line shows probabilities calculated from the matched frequency ANN. The in sample fit of the MF-ANN is visually better than the matched frequency model, giving stronger indications during recession periods and few false positives for all forecast horizons. The probability of a recession rarely rises above 20% during non recession periods. Although the matched frequency ANN still performs well in recession periods, it does not always give an as high signal of a recession as the MF-ANN. Additionally it produces more false positives during non recession periods. For example, in the period before the 1990 recession there are 3 false positives indicating a recession probability above 50% at the 1 month forecast horizon.

Tables 2.5 and 2.6 compare in sample forecast evaluations of the MF-ANN with the matched frequency ANN and the benchmark mixed frequency probit (MFP). For the QPS and QPS Onset metric the error rate for the MF-ANN is anywhere from 48% to 81% lower. The Diebold Mariano test concludes that the MF-ANN is a more accurate model than the matched ANN and MFP at the 1% level of significance. Table 2.7 compares the matched ANN with the MFP. The QPS Full suggests that the MFP marginally performs better than the matched ANN, but on average the matched ANN is better at predicting the onset of a recession as per the QPS Onset. The Diebold Mariano test suggests that neither model can be concluded as being more accurate in forecasting US recessions.

2.6.4 Out of Sample

The out of sample forecasting of recessions comes with some complications. The Y variable indicating whether the US economy is in a recession or not is observed with a lag. Nyberg (2010) assumes this 'publication lag' in the recession indicator is 9 months when forecasting US recessions. However, there is no exact answer as to what this lag should be. The Great Recession was not officially declared a recession until 12 months after the fact, whereas the most recent recession due to COVID-19 was announced 4 months after the fact. I incorporate the publication lag of Y into my out of sample forecasting the same way as Mitchell (2021) does when forecasting recessions using a mixed frequency probit model. They found changing the announcement lag did not have significant impacts on the results, and I find the same result when using neural networks.

It is also important that when carrying out the out of sample forecast I only use information known at time t to forecast recession probabilities in $t + h$. The real market factor contains data that is released with a publication lag. I follow Fossati (2015) assuming this is a 1 period lag, and adjust my data accordingly. The rest of the variables with publication lags are stated in the Data Appendix.

With this in mind, I use recursive estimation for the out of sample period, expanding the time window one month at a time. If the model uses factors then these are re-estimated every time an additional month of X data becomes available. Finally, I re-estimate the ANN model as I add each additional month of data adjusting the hyperparameters of the model in line with Tables 2.3 and 2.4. Rolling windows is another potential technique to use, however, because of the lack of recessions that have occurred in my US sample (7), I do not. The above methodology allows me to estimate real time h month ahead recession probabilities for the US.

The entire out of sample period runs from November 1988 to December 2020. Figure 2.3 shows the out of sample recession probabilities for forecast horizons 1, 2 and 3 months. The red line shows recession probabilities calculated from the MF-ANN, and the black line shows probabilities calculated from the matched frequency ANN. For all forecast horizons both models perform similarly in non recession periods with notable false positives before the 1990 recession and dotcom bubble, the latter possibly being attributed to 20% drop in the US stock market in 1998. In terms of performance during recessionary periods, results are mixed. The MF-ANN outperforms the matched ANN during the Great Recession in all forecast horizons, and also gives a larger probability for the recession caused by the COVID 19 pandemic. Both models perform similarly in the recession caused by the dotcom bubble, with the matched ANN slightly outperforming the MF-ANN at the 1 month forecast horizon. Finally, the matched ANN outperforms the MF-ANN at all forecast horizons for the 1990 recession.

Table 2.8 compares out of sample forecast evaluations of the MF-ANN with the matched frequency ANN. Although the MF-ANN performs better over the whole out of sample period compared to the matched model at the 1 month forecast horizon, shown by the QPS Full and LPS Full values being less than one, there seems to be no benefit in using mixed frequency data in neural networks at the longer forecast horizons of 2 and 3 months. In most cases the matched frequency network performs better than the MF-ANN, however the Diebold Mariano test indicates that we cannot conclude either model is more accurate at forecasting US recessions at the 5% significance level. One of the advantages Mitchell (2021) found when using mixed frequency data in a probit model to forecast US recession was the improved ability to detect the onset of a recession, however this is not apparent here.

Table 2.8 and 2.9 compare out of sample forecast evaluations of the MF-ANN with the benchmark mixed frequency probit (MFP). Results are interesting showing that when using the QPS Full evaluation metric the MF-ANN performs marginally better than the MFP at the 1 and 2 month forecast horizons. However, when using the LPS Full metric there are large improvements in forecasts for the MF-ANN between 28% and 54%. LPS penalizes larger errors more so than the QPS, and therefore suggests that the MF-ANN makes less extreme errors than the MFP. The Diebold Mariano test does not indicate that one method is more accurate than the other despite the large improvement in the LPS score for the MF-ANN. One reason is likely to be the fact that the error this test uses as an input is the same as the error used to calculate the QPS not the LPS.

Finally, Table 2.10 compares out of sample forecast evaluations of the matched frequency network with the benchmark mixed frequency probit (MFP). Overall the pattern of results are similar to Table 2.9 and the same intuition can be applied.

These out of sample results show promise of using machine learning over reduced form methods. Additionally they indicate that when applied to forecasting US recessions, mixed frequency data in a neural network can marginally improve performance at the short forecast horizons compared to matched frequency data in a neural network. However, the benefit of mixed frequency data diminishes at longer forecasting horizons.

2.7 Conclusion

This paper uses factors representing the bond, exchange rate, stock, commodity and real activity markets in the US and Canada estimated from panels of financial and macroeconomic variables. These factors that consist of daily, weekly and monthly frequencies are then used in a neural network to forecast US recession probabilities, as published by the NBER, at 1 to 3 month horizons. My main findings are that in sample a MF-ANN can drastically improve performance at all forecast horizons, in terms of forecasting accuracy, compared to a benchmark matched frequency ANN and a benchmark mixed frequency probit (MFP) model. These improvements can be as large as 79% when looking at the QPS and 81% when looking at the QPS onset.

Out of sample results also show improvement of neural networks over reduced form models. Over the whole out of sample period, and for all forecast horizons, the MF-ANN outperforms or performs similarly to the MFP according to the QPS and LPS. Improvements for the QPS are up to 9% and for the LPS as large as 54%. As the LPS puts a greater penalty on larger errors we can conclude that the MFP creates large forecasting errors at a higher rate than the MF-ANN. The story is similar when comparing the matched frequency ANN to the MFP. However, unlike the in sample results the benefit of the MF-ANN over the matched frequency ANN is not as evident. While at the 1 month forecast horizon, for the whole out of sample period, the MF-ANN reduced the QPS and LPS by 5%, all other metrics and forecast horizons favor or are indifferent between the matched and mixed frequency models. Although the reduction in performance is relatively low, with the MF-ANN increasing the forecast error by 5% at most over the full out of sample period, and 12% for the recession onset QPS.

Finally, I find that daily US bond market factors, monthly US stock market factors and monthly US real activity factors are the best leading indicators for US recession. Despite the availability of Canadian financial and macroeconomic data this was not selected in the best forecasting model. This could indicate that the US economy may be a leading indicator for the Canadian economy and not vice versa.

2.8 Data Appendix

For US data in Table 2.1, the Bond and Exchange rate market data are from FRED (St. Louis Fed) unless AC (authors calculation) is stated. All macroeconomic indicators are from FRED. The SP500 Industrial Index and SP500 PE ratio are from GFD (Global Financial Data). Finally, the S&P 500 and Dow Jones index closing price is from Yahoo Finance.

For Canadian data in Table 2.2, the Bond market data are from Statistics Canada, Exchange rate data is from FRED, Stock market data is from GFD, Commodity prices are from GFD and all Macroeconomic indicators are from FRED, apart from Housing Starts which is from Statistics Canada. AC indicates authors calculation.

Table 2.1: US Variables, Chapter 2

Variable	Daily	Weekly	Monthly	Description
Bond Market				
Fed Funds 2 (0)	Y	N	Y	Interest Rate: Federal Funds (Effective) (% per annum)
3m Tbill 2 (0)	Y	N	Y	Interest Rate: US Treasury Bills, 3-Mo. (% per annum)
6m Tbill 2 (0)	Y	N	Y	Interest Rate: US Treasury Bills, 6-Mo. (% per annum)
1y Tbond 2 (0)	Y	N	Y	Interest Rate: US Treasury, 1-Yr. (% per annum)
5y Tbond 2 (0)	Y	N	Y	Interest Rate: US Treasury, 5-Yr. (% per annum)
10y Tbond 2 (0)	Y	N	Y	Interest Rate: US Treasury, 10-Yr. (% per annum)
AAA bond 2 (0)	N	Y	Y	Bond Yield: Moody's AAA Corporate (% per annum)
BAA bond 2 (0)	N	Y	Y	Bond Yield: Moody's BAA Corporate (% per annum)
3m spread 1 (0)	Y	N	Y	3m Tbill - Fed Funds (AC)
6m spread 1 (0)	Y	N	Y	6m Tbill - Fed Funds (AC)
1y spread 1 (0)	Y	N	Y	1y Tbill - Fed Funds (AC)
5y spread 1 (0)	Y	N	Y	5y Tbill - Fed Funds (AC)
10y spread 1 (0)	Y	N	Y	10y Tbill - Fed Funds (AC)
AAA spread 1 (0)	N	Y	Y	AAA bond - Fed Funds (AC)
BAA spread 1 (0)	N	Y	Y	BAA bond - Fed Funds (AC)
Exchange Rate Market				
Ex. Rate Switzerland 3 (0)	Y	N	Y	Foreign Exchange Rate: Swiss Franc per US$
Ex. Rate Japan 3 (0)	Y	N	Y	Foreign Exchange Rate: Yen per US$
Ex. Rate UK 3 (0)	Y	N	Y	Foreign Exchange Rate: Cents per Pound
Ex. Rate Canada 3 (0)	Y	N	Y	Foreign Exchange Rate: Canadian$ per US$
Stock Market				
SP 500 3 (0)	Y	N	Y	S&P 500 Index Closing Price
DJ Index 3 (0)	Y	N	Y	Dow Jones Index, Closing Price
SP Industrials 3 (0)	N	N	Y	S&P 500 Industrials Index Closing Price
SP PE ratio 3 (2)	N	N	Y	S&P 500 Index: Price Earnings Ratio (%)
Macroeconomic Indicators				
IPI 3 (1)	N	N	Y	Industrial Production Index, Total Index
PILT 3 (1)	N	N	Y	Personal Income Less Transfer Payments
MTS 3 (1)	N	N	Y	Manufacturing and Trade Sales
Emp: Total 3 (1)	N	N	Y	Employees On Nonfarm Payrolls: Total Private
Housing Starts 3 (1)	N	N	Y	Total New Privately Owned Housing Units Started
Unemp claims: Weekly 3 (1)	N	Y	N	Unemployment Insurance Weekly Claims

1 = no transformation, 2 = first difference, 3 = log of first difference, number in parentheses = publication lag in weeks or months

Table 2.2: Canada Variables, Chapter 2

Variable	Daily	Weekly	Monthly	Description
Bond Market				
10y+ Gov Bond 2 (0)	N	Y	Y	Government marketable bonds average yield, 10 years
5-10y Gov Bond 2 (0)	N	Y	Y	Government marketable bonds average yield, 5-10 years
3-5y Gov Bond 2 (0)	N	Y	Y	Government marketable bonds average yield, 3-5 years
1-3y Gov Bond 2 (0)	N	Y	Y	Government marketable bonds average yield, 1-3 years
3m Prime Corp Paper 2 (0)	N	Y	Y	3 months prime corporate paper
2m Prime Corp Paper 2 (0)	N	Y	Y	2 months prime corporate paper
1m Prime Corp Paper 2 (0)	N	Y	Y	1 month prime corporate paper
10y+ spread 1 (0)	N	Y	Y	Yield Spread b/t 10-yr bond and 3-m prime (AC)
5-10y spread 1 (0)	N	Y	Y	Yield Spread b/t 5-10-yr bond and 3-m prime (AC)
3-5y spread 1 (0)	N	Y	Y	Yield Spread b/t 3-5-yr bond and 3-m prime (AC)
1-3y spread 1 (0)	N	Y	Y	Yield Spread b/t 1-3-yr bond and 3-m prime (AC)
Exchange Rate Market				
Ex. Rate Switzerland 3 (0)	Y	N	Y	Foreign Exchange Rate: Swiss Franc per Can$
Ex. Rate Japan 3 (0)	Y	N	Y	Foreign Exchange Rate: Yen per Can$
Ex. Rate US 3 (0)	Y	N	Y	Foreign Exchange Rate: US$ per Can$
Ex. Rate UK 3 (0)	Y	N	Y	Foreign Exchange Rate: Pound Sterling per Can$
Commodity Market				
Oil Price 3 (0)	Y	N	Y	West Texas Intermediate Oil Price
Gold Price 3 (0)	Y	N	Y	Gold Spot Price, London PM Fixing
Wheat Price 3 (0)	Y	N	Y	Wheat 2 Cash Price
Aluminum Price 3 (0)	Y	N	Y	Aluminum Spot Price
Stock Market				
TSX Index 3 (0)	Y	N	Y	Toronto Stock Exchange, composite index
TSX Value 3 (0)	N	N	Y	Toronto Stock Exchange Index, closing price
TSX Vol 3 (0)	N	N	Y	Toronto Stock Exchange, volume of shares traded
Macroeconomic Indicators				
Consumer Credit 3 (1)	N	N	Y	Consumer Credit, month-end, sa, Total outstanding balances
Manufacturing Prod 3 (1)	N	N	Y	Production in total manufacturing, sa
Emp: Total 2 (1)	N	N	Y	Employed population, total
Housing Starts 1 (1)	N	N	Y	Housing starts 12 month growth

1 = no transformation, 2 = first difference, 3 = log of first difference, number in parentheses = publication lag in weeks or months

Tables

Table 2.3: Hyperparameter Tuning, Mixed Frequency ANN

Hyperparameter		**Hidden Layers**	**Nodes**	**Decay**	**Iterations**
	h=1 month	1	1	0.10	600
Block 3	h=2 month	1	1	0.15	600
	h=3 month	1	1	0.40	600
	h=1 month	1	1	0.10	600
Block 4	h=2 month	1	1	0.15	600
	h=3 month	1	2	0.20	400
	h=1 month	1	2	0.15	600
Block 5	h=2 month	1	1	0.15	600
	h=3 month	1	1	0.25	600
	h=1 month	1	1	0.05	400
Block 6	h=2 month	1	2	0.10	600
	h=3 month	1	1	0.25	400

Table 2.4: Hyperparameter Tuning, Matched Frequency ANN

Hyperparameter		**Hidden Layers**	**Nodes**	**Decay**	**Iterations**
	h=1 month	1	5	0.05	600
Block 3	h=2 month	1	4	0.05	400
	h=3 month	1	3	0.10	400
	h=1 month	1	4	0.05	600
Block 4	h=2 month	1	5	0.05	400
	h=3 month	1	3	0.10	400
	h=1 month	1	2	0.05	600
Block 5	h=2 month	1	5	0.05	400
	h=3 month	1	3	0.05	600
	h=1 month	1	5	0.05	600
Block 6	h=2 month	1	3	0.05	400
	h=3 month	1	1	0.05	600

Hyperparameters chosen, for each time block of the data and forecast horizon, to be used for in sample and out of sample analysis for the MF-ANN and matched frequency ANN. Block 1: October 1971 - June 1981. Block 2: July 1982 - June 1990. Block 3: July 1990 - February 2001. Block 4: March 2001 - November 2007. Block 5: December 2007 - April 2014. Block 6: May 2014 - December 2020.

Table 2.5: In Sample: MF-ANN vs Matched ANN

Forecast Horizon		**1 month**	**2 month**	**3 month**
QPS Full	$\frac{QPS_{MF-ANN}}{QPS_{Matched}}$	0.21	0.41	0.23
QPS Onset	$\frac{QPS_{MF-ANN}}{QPS_{Matched}}$	0.26	0.52	0.26
Diebold Mariano	p value	0.01	0.01	0.01

Table 2.6: In Sample: MF-ANN vs Mixed Frequency Probit (MFP)

Forecast Horizon		**1 month**	**2 month**	**3 month**
QPS Full	$\frac{QPS_{MF-ANN}}{QPS_{MFP}}$	0.24	0.41	0.24
QPS Onset	$\frac{QPS_{MF-ANN}}{QPS_{MFP}}$	0.19	0.43	0.27
Diebold Mariano	p value	0.01	0.01	0.01

Table 2.7: In Sample: Matched ANN vs Mixed Frequency Probit (MFP)

Forecast Horizon		**1 month**	**2 month**	**3 month**
QPS Full	$\frac{QPS_{Matched}}{QPS_{MFP}}$	1.15	1.00	1.04
QPS Onset	$\frac{QPS_{Matched}}{QPS_{MFP}}$	0.73	0.81	1.06
Diebold Mariano	p value	0.56	0.29	0.42

Comparison of in sample results for the MF-ANN, matched frequency ANN and mixed frequency probit (MFP - from Mitchell (2021)) at various forecast horizons. All models include 3 bond market factors, 1 stock market factor and 2 real market factors. Note that the LPS is not reported as some predicted probabilities take the value of 0 or 1 and therefore the LPS is undefined.

Table 2.8: Out of Sample: MF-ANN vs Matched ANN

Forecast Horizon		**1 month**	**2 month**	**3 month**
QPS Full	$\frac{QPS_{MF-ANN}}{QPS_{Matched}}$	0.95	1.03	1.05
QPS Onset	$\frac{QPS_{MF-ANN}}{QPS_{Matched}}$	1.12	1.07	1.06
LPS Full	$\frac{LPS_{MF-ANN}}{LPS_{Matched}}$	0.95	1.04	1.00
Diebold Mariano	p value	0.30	0.82	0.93

Table 2.9: Out of Sample: MF-ANN vs Mixed Frequency Probit (MFP)

Forecast Horizon		**1 month**	**2 month**	**3 month**
QPS Full	$\frac{QPS_{MF-ANN}}{QPS_{MFP}}$	0.91	0.98	1.01
QPS Onset	$\frac{QPS_{MF-ANN}}{QPS_{MFP}}$	1.01	1.09	1.19
LPS Full	$\frac{LPS_{MF-ANN}}{LPS_{MFP}}$	0.46	0.62	0.72
Diebold Mariano	p value	0.17	0.59	0.87

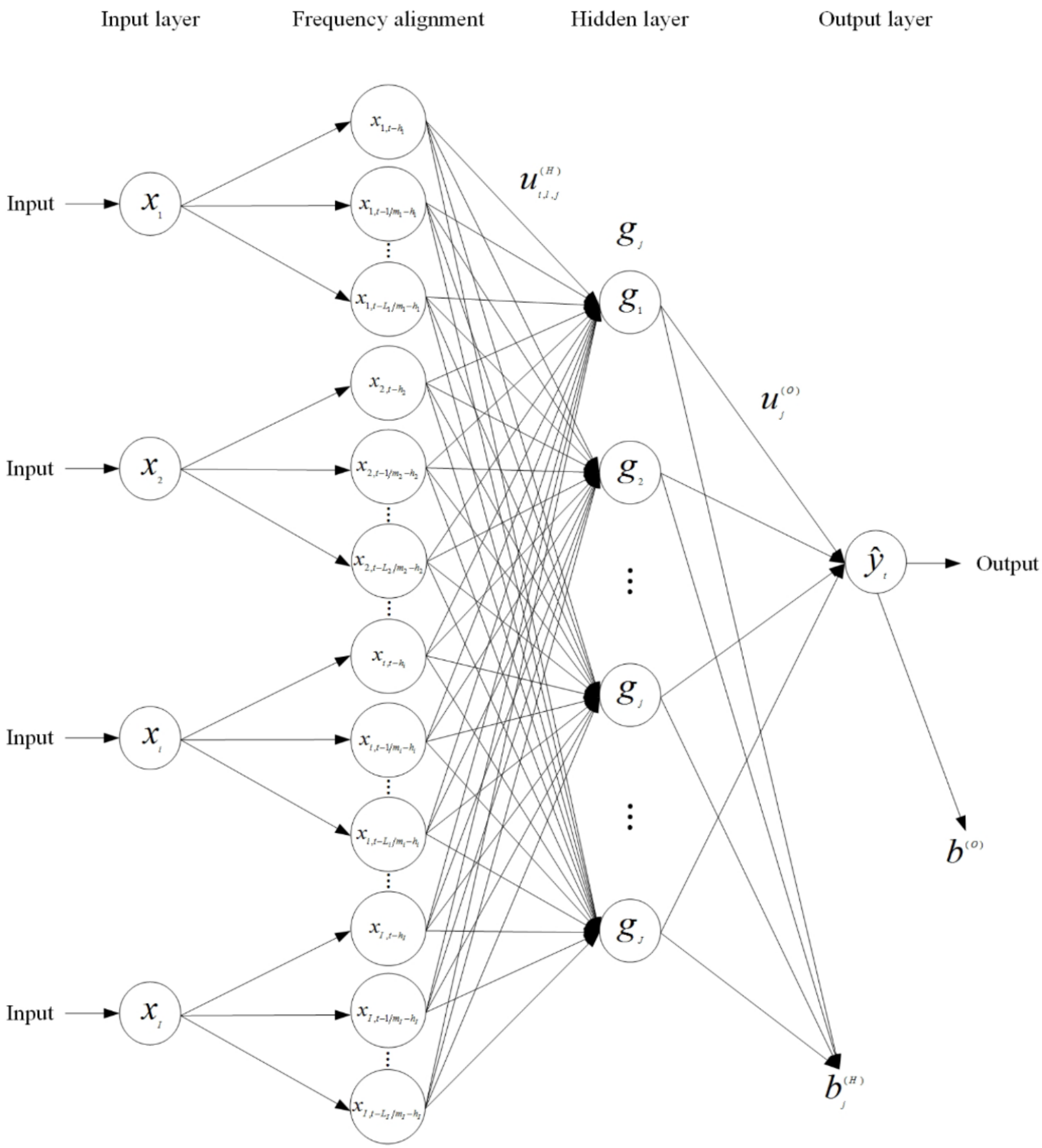
Table 2.10: Out of Sample: Matched ANN vs Mixed Frequency Probit (MFP)

Forecast Horizon		**1 month**	**2 month**	**3 month**
QPS Full	$\frac{QPS_{Matched}}{QPS_{MFP}}$	0.95	0.95	0.96
QPS Onset	$\frac{QPS_{Matched}}{QPS_{MFP}}$	0.90	1.02	1.13
LPS Full	$\frac{LPS_{Matched}}{LPS_{MFP}}$	0.48	0.62	0.71
Diebold Mariano	p value	0.34	0.29	0.41

Comparison of out of sample results for the MF-ANN, matched frequency ANN and mixed frequency probit (MFP - from Mitchell (2021)) at various forecast horizons. All models include 3 bond market factors, 1 stock market factor and 2 real market factors.

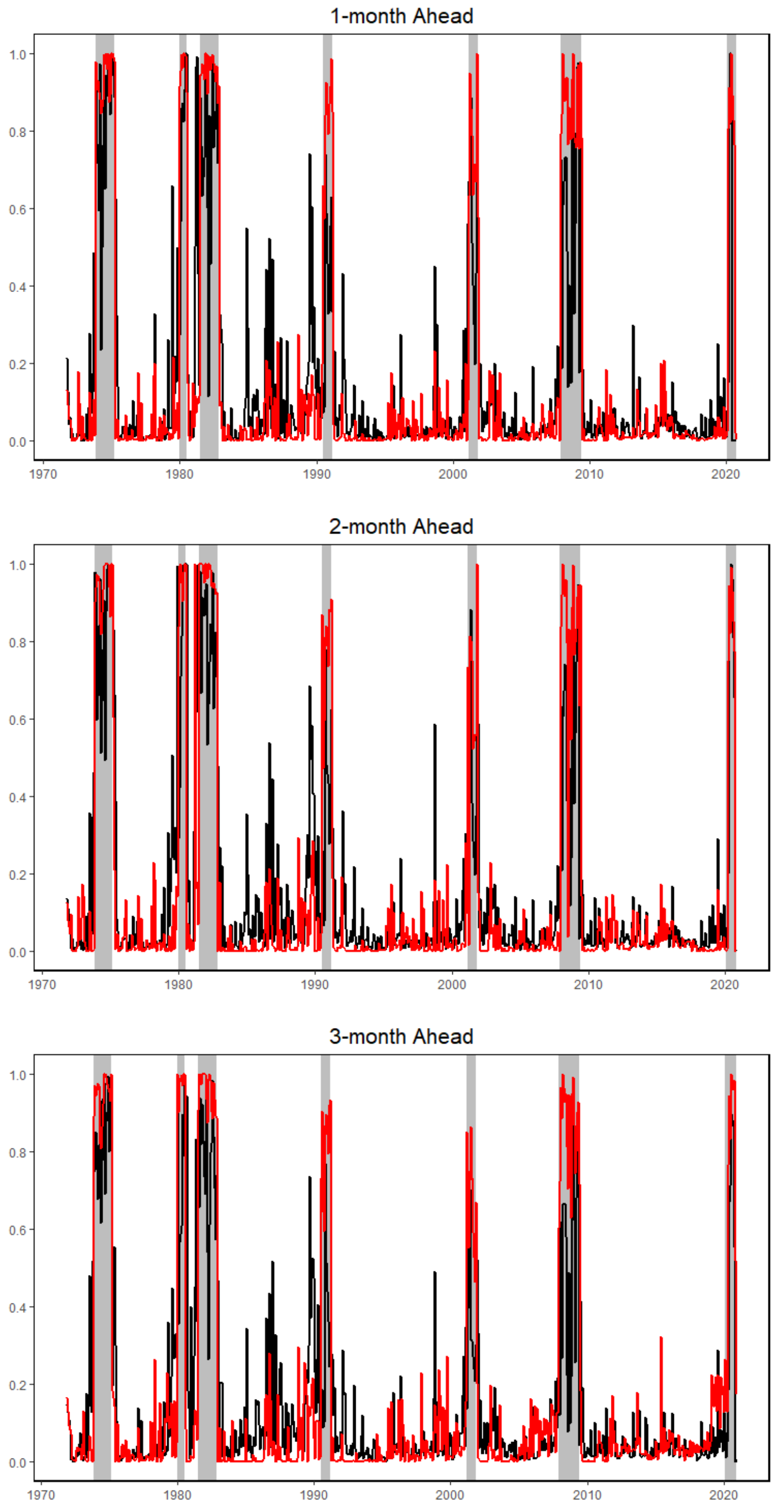
Figures

Figure 2.1: Unrestricted Mixed Frequency ANN



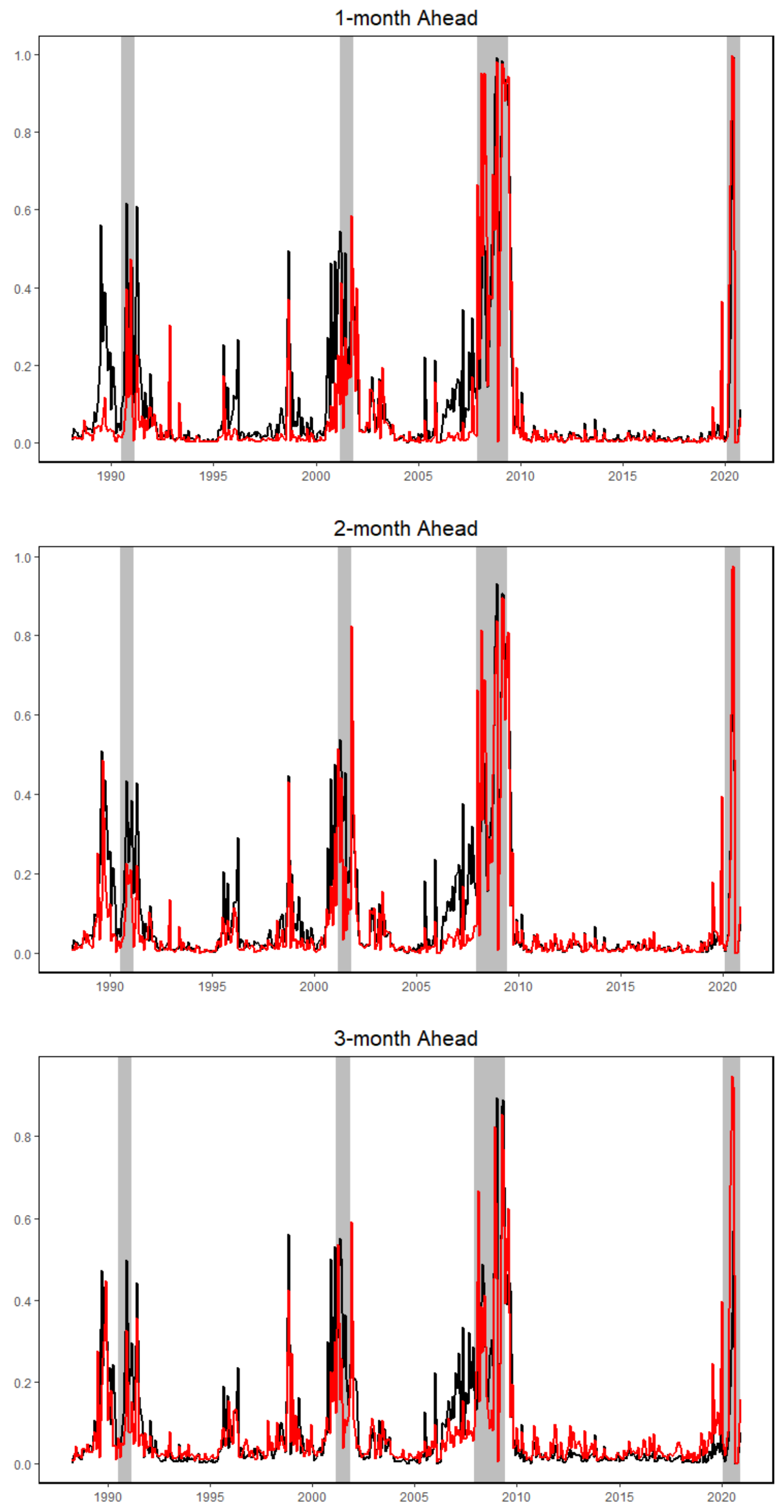
Unrestricted Mixed Frequency ANN, Xu et al (2019)

Figure 2.2: ANN In Sample



In sample US predicted probabilities of a recession for the best performing ANN model as chosen by the hyperparameter tuning and forward selection wrapper processes at different forecast horizons: MF-ANN (red); matched frequency ANN (black). Shaded areas show recessions as defined by the NBER.

Figure 2.3: ANN Out of Sample



Out of sample US predicted probabilities of a recession for the best performing ANN model as chosen by the hyperparameter tuning and forward selection wrapper processes at different forecast horizons: MF-ANN (red); matched frequency ANN (black). Shaded areas show recessions as defined by the NBER.

Forecasting GDP Using A Mixed Frequency Seemingly Unrelated Regression Approach

Authors: Yu-chin Chen, Matt Daniels and Ryan Mitchell

3 Chapter 3

3.1 Introduction

In this paper we use Seemingly Unrelated Regressions (SUR), an approach first introduced by Zellner (1962), to forecast GDP for 10 OECD countries at various forecast horizons. The common method adopted by past literature is to use a single regression equation with financial and macroeconomic predictors as leading indicators, chosen from past literature, to forecast GDP of each country separately. However, this approach usually ignores potential cross country influences and financial contagion of economic shocks from one country to another that may help forecast GDP of a particular country. Although cross country information can be included directly into the forecasting equations of GDP, certain issues arise, for example finding and limiting down the correct cross country data to be used among the many options. This can be especially difficult when only a relatively small number of variables can be used with the small sample sizes apparent with quarterly GDP.

SUR states that when the error terms in the different regression equations are correlated, and different independent variables appear in the equations, the regression equations are related, not unrelated hence the term "seemingly unrelated". Therefore Zellner (1962) says that the sample information in other regressions can be employed to improve the precision of estimation of parameters in any given regression equation when error terms across single forecasting regressions are highly correlated, Baltagi (2011). Taking account of the correlation across the error terms and using them in the estimation procedure can also lead to better out of sample prediction of the dependent variable. For example, Cadavez (2012) who shows that the SUR estimator performs better than the OLS estimator out of sample when predicting carcass composition of lambs. SUR is a form of feasible generalized least squares (FGLS) and under the conditions above Zellner (1962) shows it is the best linear unbiased estimator. This methodology can therefore complement existing panel and simultaneous equation techniques to forecast GDP by using the correlation across multiple country's error terms as a way to incorporate cross country information.

The key assumption behind SUR of correlation between the error terms of each forecasting equation is a feasible assumption with respect to GDP. Economic shocks such as recessions will usually hit multiple countries at the same time, especially when countries are highly dependent on one another through channels such as trade and monetary policy, as many OECD countries are. Hence, SUR seems a natural approach to forecast GDP.

This paper therefore adds to the literature in multiple ways. (1) We use SUR as a novel approach to include cross country information in GDP forecasting equations without having to directly select which cross country variables should be used. (2) We found SUR to be especially good compared to OLS when sample sizes are restricted, which is apparent due to data restrictions of many EU countries. Kmenta (1968) and Hoorsgate (2000) both have similar findings when comparing the two methodologies. This therefore allows us to forecast GDP for many countries which have not been extensively done by the past literature due to relatively short time series, with a more efficient estimator. (3) We expand the SUR model not just in terms of application purpose but to also allow for the inclusion of mixed frequency time series data. Since the introduction of mixed data sampling (MIDAS) approaches by Ghysels (2004) it has become popular to include higher frequency time series data to forecast lower frequency dependent variables, especially in the case of GDP with success. To our knowledge we are the first to incorporate mixed frequency data into the SUR model. As pointed out by Andreou, Ghysels and Kourtellos (2013), not using the readily available high frequency data such as daily financial and monthly macroeconomic predictors to perform the quarterly GDP forecasts may lead to important information being lost through temporal aggregation.

In summary in this paper we collect a range of bond, stock, exchange rate and macroeconomic data for 10 different OECD countries, found in Table 3.1, and use this to forecast their GDP using a seemingly unrelated regressions approach at various forecast horizons. When using matched frequency data aggregated at the quarterly level we find that SUR reduces the out of sample forecast error for 31/40 of the regressions run when compared to a benchmark OLS model. This improvement ranges from a reduction of 1% to 29% in the out of sample mean squared error. Results are similar when including mixed frequency data within the SUR model. In this case SUR also reduces out of sample forecast error for 31/40 of the regressions run when compared to a benchmark mixed frequency OLS model, with similar levels of magnitude with regards to error reduction. Monte Carlo experiments help us to conclude that the superior performance of SUR over OLS could be driven by the small sample sizes we have due to data restrictions.

Finally, when comparing a mixed frequency SUR with a matched frequency SUR, the mixed frequency model only reduces out of sample mean squared error for 6/40 of the regressions run. This indicates that the benefit of including mixed frequency data in an SUR model is diminished. Further exploration indicates a significantly lower level of correlation in the error terms for the mixed frequency model compared to the matched frequency model. A key assumption of SUR is high correlation between the forecasting equation error terms and we can conclude the reduced performance is due to this.

3.2 Literature Review

3.2.1 Matched and Mixed Frequency Literature

There is a vast literature forecasting various countries GDP with both matched and mixed frequency data with single regression equations, and therefore this section will give a brief overview. Stock and Watson (2003) undertook an empirical analysis of 38 quarterly indicators to forecast output for seven different OECD countries. Comparing forecasts of an autoregressive model with that of an additional indicator in the regression, they find small marginal predictive content for output at the two, four and eight quarter horizon using the additional indicator. While there is no stand out asset price that drastically improves predictive performance across all countries and time periods, the term spread comes closest to achieving this goal. While they do find instability of individual predictive regressors, they do see reliable improvements in forecasts when using forecast combinations, either by computing the trimmed mean or the median forecast.

Findings by Stock and Watson (2003), as well as others have led to much of the forecast literature to use factor analysis with relation to variable selection. For example, Ghysels et al (2013) estimate factors from a large panel of 1000 potential predictor variables. These factors are then used in the regression equation to forecast US GDP. Fossati (2015) also uses factor analysis as a data dimension reduction technique when forecasting recessions in the US. They find that using factors in their model reduces out of sample forecast errors compared to using individual indicators separately in the regression.

Since Ghysels (2004) introduced a mixed data sampling (MIDAS) to the forecasting literature, much of the GDP literature has looked to incorporate mixed frequency data, instead of aggregating all variables at the quarterly frequency. Ferrara (2013) forecasts GDP growth in different euro area countries using the MIDAS approach and found that higher frequency stock price data improved forecast accuracy. Ghysels et al (2013) used MIDAS with factors estimated from 1000 daily financial assets to forecast GDP in the US. Overall they found that MIDAS regression models using daily financial information via daily financial factors improve quarterly forecasts of US real GDP growth beyond the quarterly macroeconomic factors.

3.2.2 Forecasting GDP with Cross Country Data

The literature on combining multiple countries GDP regression equations through panel data to forecast is much more sparse. Hoorsgate (2000) aimed to forecast the GDP growth rate of 18 OECD countries. They found that the median mean squared error of OLS based pooled forecasts is found to be smaller than that of OLS based individual forecasts. A fairly large sample size, in terms of the time dimension, is needed for the OLS based pooled forecasts to be outperformed by a forecast based on unrestricted estimates. All results from this paper were based on in sample analysis, not out of sample analysis. Kholodilin (2008) used a dynamic panel model to make multi step forecasts of the annual growth rates of real GDP for each of the 16 German states. They find that by pooling and accounting for spatial effects, via fixed effects, they can improve on forecast accuracy by between 9% and 40% depending on the forecast horizon.

Although the literature of forecasting GDP in panel models is limited there is much literature on using cross country data directly to forecast economic conditions. For example, Fossati (2018) found improvements in forecasting performance for Canadian recessions when including US bond market data as a predictor variable. Lyu (2021) finds that cross country data helps produce more accurate forecasts of US GDP growth during economic downturns, but is less helpful in normal times. The reason they give for this is that foreign variables contain useful information about the spillover effects on the US economy, which are particularly important in economic downturns. Another natural methodology that may be used to include cross country information is vector autoregression (VAR) models. Given the small sample sizes we are faced with in this paper we do not consider this due to degrees of freedom issues, but combining SUR with Lasso variable selection techniques could be a natural extension and comparison for this paper in the future.

3.3 Methodology

3.3.1 Seemingly Unrelated Regressions (SUR)

Suppose there are m countries, each of whom we are interested in forecasting their GDP at some forecast horizon h. For each country we would run the following regression

$$y_{i,t+h} = \beta_{1,i}x_{1,i,t} + ... + \beta_{q,i}x_{q,i,t} + \epsilon_{i,t} \tag{34}$$

Where $i = 1, ..., m$ refers to each individual country and $t = 1, ..., T$ refers to the number of quarterly observations. The variable of interest to forecast, $y_{i,t+h}$, is the log first difference of GDP between quarter $t+h$ and quarter $t+h-1$ for country i, that is $y_{i,t+h} = ln(\frac{GDP_{i,t+h}}{GDP_{i,t+h-1}})$. Finally $x_{i,t}$ refers to the q covariates observed at time t useful in forecasting GDP for country i as determined by the past literature.

SUR assumes that the error terms are independent across time periods, but may have cross equation correlations within the same time period. That is $E[\epsilon_{i,t}, \epsilon_{i,t+h}] = 0$ whenever $h \neq 0$, but $E[\epsilon_{i,t}, \epsilon_{j,t}] = \sigma_{i,j}$. This is a fair assumption for GDP and OECD countries. Economic shocks are likely to hit related countries at similar times and hence standard OLS models are likely to over and under predict future GDP for these countries also at the same time, leading to high contemporaneous correlation between countries error terms that is required for SUR. Equation 34 can be estimated by OLS for each country individually, and would satisfy the Gauss Markov theorem. However, if we were to write Equation 34 as a system of equations across all the countries, then because of the assumption of correlated errors across countries, OLS will not be efficient and we should take into account this correlation of the error terms to improve our parameter estimations. In order to incorporate the contemporaneous correlation of errors between countries we can follow a 2 step SUR estimation procedure.

Firstly, we must convert Equation 34 to a system of equations across countries using matrix notation as follows

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_m \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_m \end{bmatrix} = X\beta + \epsilon \tag{35}$$

Where y_i is a $(T \times 1)$ vector of the transformed GDP variable for country i, X_i is a $(T \times q)$ matrix of covariates for country i, β_i is $(q \times 1)$ vector of coefficients for country i and ϵ_i is a

$(T \times 1)$ vector of error terms for country i. Given that we assume the errors are correlated between countries within the same time period let $\Sigma = \sigma_{i,j}$, which is an $(m \times m)$ covariance matrix for the countries error terms. The final covariance matrix of the stacked error terms will be equal to

$$\Omega = E[\epsilon\epsilon^\top | X] = \Sigma \otimes I_T \tag{36}$$

Where I_T is a $(T \times T)$ identity matrix and $\otimes$ is the Kronecker product, which gives the covariance matrix of the stacked errors the correct dimensions to be used in the second step of the SUR estimation procedure. Therefore the system of equations in Equation 35 exhibits both heteroskedasticity and non zero off diagonal values of the error covariance matrix, and therefore OLS will not be the best linear unbiased estimator.

The first step of the SUR estimation procedure is to run OLS on Equation 35 for each country individually. The second step saves the residuals $\hat{\epsilon}$ from the first step which can be used to estimate $\hat{\Sigma}$. Finally a feasible generalized least squares regression is run giving the following estimator

$$\hat{\beta} = \left(X^\top (\hat{\Sigma}^{-1} \otimes I_T) X \right)^{-1} X^\top (\hat{\Sigma}^{-1} \otimes I_T) y \tag{37}$$

Although both this SUR estimator and the OLS estimator are unbiased, the SUR estimator, under the conditions described above, will be the best linear unbiased estimator due to its increased efficiency.

3.3.2 Mixed Frequency Seemingly Unrelated Regressions (MF-SUR)

In this paper we extend the SUR model to allow for inclusion of higher frequency financial and macroeconomic data when forecasting GDP. The mixed frequency seemingly unrelated regressions (MF-SUR) model follows the methodology proposed by Chen and Tsay (2011), who use a polynomial to weight daily exchange rate data to forecast quarterly changes in commodity prices. To incorporate the higher frequency data there are two key changes to be made to the matched frequency SUR model.

Firstly, the higher frequency X data variables are transformed from vectors to matrices with each row corresponding to a different quarter. Columns then correspond to data points within the quarter. As an example, consider using daily financial data to forecast quarterly GDP. If there are K trading days in a quarter then we may have as many as K data points for each quarterly GDP observation. One individual variable would be a vector of length $T \times K$ that needs to be transformed into a matrix of dimensions $(T \times K)$, with the first column giving the most recent day's financial data, the second column giving the second most recent, and so on.

The below graphic gives a visual representation of the y vector and X matrix for a given country i assuming we are only using one covariate and forecasting one quarter ahead.

$$y_i = \begin{bmatrix} y_{Q2} \\ y_{Q3} \\ \vdots \\ y_T \end{bmatrix} X_i = \begin{bmatrix} x_{i,mar31st} & x_{i,mar30th} & \cdots & x_{i,jan1st} \\ x_{i,2-h} & x_{i,2-h-1/K} & \cdots & x_{i,2-h-(K-1)/K} \\ \vdots & \vdots & \cdots & \vdots \\ x_{i,T-h} & x_{i,T-h-1/K} & \cdots & x_{i,T-h-(K-1)/K} \end{bmatrix} \tag{38}$$

Secondly, each higher frequency covariate will be multiplied by a Vandermonde matrix of dimensions of $(K \times n)$, where n refers to the degree of polynomial that will be used to weight the higher frequency data. Visually the Vandermonde matrix will look like

$$V = \begin{bmatrix} 1 & 1^1 & 1^2 & \dots & 1^{n-1} \\ 1 & 2^1 & 2^2 & \dots & 2^{n-1} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & K^1 & K^2 & \dots & K^{n-1} \end{bmatrix} \tag{39}$$

This follows Almon (1965) that assumes each lag coefficient can be approximated by a polynomial of degree $n-1 < K$. Therefore, instead of having to estimate $1+Kq$ parameters (the intercept and weights on each K higher frequency observation of the q variables) the model is reduced to $1 + nq$ parameters to estimate. This ignores matched frequency coefficients that need to be estimated in addition. Let us take country i as an example. We can write the individual regression equation as

$$y_i = \beta_{0,i} + \beta_{1,i} X_{1,i} V \alpha_{1,i} + ... + \beta_{q,i} X_{q,i} V \alpha_{q,i} + \beta_{q+1,i} X_{q+1,i} + ... + \beta_{r,i} X_{r,i} + \epsilon_i \tag{40}$$

Where $X_{j,i}$ for $j = 1, 2, ..., q$ is a $(T \times K)$ matrix of higher frequency financial and macroeconomic data, sampled at daily, weekly and monthly frequencies. $X_{j,i}$ for $j = q + 1, ..., r$ is a vector of length T of matched frequency financial and macroeconomic data, sampled at a quarterly rate. y_i is a vector of length T of quarterly GDP data for country i. $\beta_{j,i}$ measures the aggregate impact of $X_{j,i}$ on y_i. V is as described in (39) and $\alpha_{j,i}$ is a $n \times 1$ vector of coefficients that form the polynomial of degree $n - 1$, which weights the higher frequency data for country i. This can be further simplified to

$$y_i = \beta_{0,i} + Z_{1,i} \gamma_{1,i} + ... + Z_{q,i} \gamma_{q,i} + \beta_{q+1,i} X_{q+1,i} + ... + \beta_{r,i} X_{r,i} + \epsilon_i \tag{41}$$

Where $Z_{j,i} = X_{j,i} V$ and $\gamma_{j,i} = \beta_{j,i} \alpha_{j,i}$ for $j = 1, 2, ..., q$. The parameter γ can then be estimated via the two step SUR procedure as explained in Section 3.3.1, and aggregate effects of $X_{j,i}$, $\forall j = 1, ..., q$, on y_i, as well as the higher frequency weighting functions, can be solved through manipulation of the $\gamma' s$.

3.3.3 Identification of β

By restricting the weights on the higher frequency data to sum up to 1, we can identify $\beta_{1,i}, ..., \beta_{q,i}$ for each country i. This means letting $\mathbf{1}^\top V \alpha_{j,i} = 1$ for $j = 1, ..., q$. We can then use the fact that $\gamma_{j,i} = \beta_{j,i} \alpha_{j,i}$ and therefore $V \gamma_{j,i} = \beta_{j,i} V \alpha_{j,i}$ to obtain:

$$\hat{\beta}_{j,i} = \mathbf{1}^\top (V \hat{\gamma}_{j,i}), \forall j = 1, ..., q \tag{42}$$

3.4 Data

Data for each country is collected via the Global Financial Data (GFD) API, please find a list of countries and data in the Data Appendix. The data poses multiple issues that must be addressed: 1) As the SUR methodology relies on collecting data for many countries we are restricted by the country that has the smallest sample size. For example, for our selected variables the US data on GFD goes back to 1975, whereas for other countries such as Germany full data for the required explanatory variables only go back to 1998. This small quarterly sample size of $Q1$ 1998 to $Q3$ 2020 means variable selection is crucial in order for us not to over-fit the model. 2) The MF-SUR leads to a large increase in the number of parameters to be estimated compared to a standard SUR model. For example, if we use daily data and decide to weight the data with a cubic polynomial then 4 parameters will need to be estimated for each daily variable. Whereas a matched frequency SUR would only require one parameter to be estimated. Again this issue highlights the importance of variable selection. 3) There is high multicollinearity between variables. For example, yields on 5 year Treasury Bonds are highly correlated with yields on 10 year Treasury Bonds. The same can be said for the macroeconomic indicators which vary with economic conditions and hence are correlated with one another.

To overcome each of these problems we use factor analysis to capture variability among observed, correlated variables in terms of a lower number of unobserved variables. Prior to estimation of the factors, data is first transformed to be stationary. The data is then split into 4 asset classes; (1) bond market (2) exchange rate market (3) stock market and (4) macroeconomic indicators. Factor analysis is then run separately on each asset class (1), (2), (3) and (4) to get a set of bond, exchange rate, stock and macroeconomic market factors for each country. This separation into asset markets allows for further intuition and conclusions to be developed from the results. The factor analysis model takes the following form for each country i

$$X = \Lambda F + \epsilon \tag{43}$$

Where X is a $(N \times T)$ matrix of observed data and Λ is an $(N \times L)$ matrix of factor loadings, where L corresponds to the number of common factors being estimated. F is a $(L \times T)$ matrix of latent factor scores and ϵ is the error term. There is no strict rule on deciding how many factors to use. These estimated factors are then used as explanatory variables in the regression equation 40.

Finally, one thing we would like to consider in the future is the impact that the zero lower bound could have on the predictive power of our bond market data and whether other more unconventional monetary policy data could be a good alternative to use.

3.5 Results

3.5.1 Forecast Evaluation

We compare the mixed and matched frequency SUR forecasts with mixed and matched frequency OLS forecasts, to allow us to examine the benefits of estimating GDP as a system of equations as opposed to separately for each country. To compare the methods we use two evaluation statistics. Firstly we calculate the mean squared error (MSE) for each country i, as shown below. For easy comparison of methodologies we will divide the MSE of each method by one another. For example, a number of 0.8 when comparing the the MF-SUR with the mixed frequency OLS means a 20% lower forecasting error for the MF-SUR.

$$MSE = \frac{1}{T} \sum [\widehat{GDP}_{i,t+h} - GDP_{i,t+h}]^2 \tag{44}$$

The second forecast evaluation method is the Diebold Mariano test with a squared loss function. This test compares the predicted values of two competing methods with the actual values of Y that occurred. For the purpose of this paper the H_A is that the first forecast method is more accurate than the second forecast method. We choose a squared loss function, as opposed to a linear loss function, as we deem incorrect GDP predictions to be costly to the economy and therefore want to penalize these at a higher rate.

3.5.2 Correlation Across Error Terms

One of the key assumptions for the seemingly unrelated regression methodology to improve forecasting performance versus OLS is that the correlation between residuals of the different countries GDP forecasting equations are correlated. The top panel of Graph 3.1 shows the correlation matrix for the residuals of the $1Q$ ahead matched frequency SUR model for the whole sample. The first thing to note is that the correlation between residuals is always positive, and secondly is relatively strong. The lowest correlation coefficient is 0.46 between Sweden and Spain, whilst the highest is 0.9 between Belgium and Italy.

The bottom panel of Graph 3.1 shows the correlation matrix for the residuals of the $1Q$ ahead MF-SUR model for the whole sample. Visually it is evident that while the correlation between residuals of the mixed frequency model is still positive, they are lower in magnitude compared to the matched model. The lowest correlation coefficient is 0.09 between Canada and Spain, whilst the highest is 0.74 between Belgium and Italy. The lower magnitude of correlation in the mixed frequency models is a common occurrence across all forecast horizons.

To compare correlation coefficients more broadly between the matched and mixed frequency models we first calculate the average correlation coefficient between each country and the other 9 countries in our sample for both models at all forecast horizons. We then divide this correlation average for the matched frequency SUR model by the MF-SUR model. Table 3.2 shows the results. On average, for all countries and forecast horizons, the residuals have higher correlation for the matched model over the mixed model, shown by a value greater than 1. In some cases the correlation of the residuals in the matched model is 98% higher than the mixed model, for example with Canada at the $1Q$ forecast horizon.

3.5.3 In Sample

By definition the mean squared error will be lower for the in sample analysis of the OLS model vs the comparable SUR model. Therefore I do not compare the in sample OLS with SUR, but instead compare the mixed frequency with the matched frequency within both of these model groups. Table 3.3 divides the in sample MSE for the MF-SUR by the MSE for the in sample matched frequency SUR. For all countries and forecast horizons, apart from the UK at the $2Q$ horizon, the MF-SUR has a lower or equal MSE. According to the Diebold Mariano test the MF-SUR is a more accurate model at the 10% level of significance for 18/40 regressions. The results are similar for the in sample OLS comparison between the mixed and matched frequency models, reported in Table 3.4, with all countries and forecast horizons producing a lower MSE in the MF-OLS. The Diebold Mariano test concludes that the MF-OLS is a more accurate model at the 10% level of significance for 14/40 regressions. However, this significant improvement in sample of using mixed frequency data cannot be conclusive of a superior model, but may instead indicate over-fitting due to the higher number of parameters estimated in the regression. Therefore out of sample analysis is required.

3.5.4 Out of Sample

As the sample size is relatively small we use recursive out of sample estimation to forecast GDP growth one to four quarters ahead, as opposed to rolling windows. The out of sample period runs from $Q2$ 2008 to $Q3$ 2020. Firstly, factors for each asset market are estimated, these are then used to estimate the model for each forecast horizon. Each time a new quarter of data is added the factors and model are re-estimated. Table 3.5 divides the out of sample MSE of the MF-SUR model by the MSE of the mixed frequency OLS model at various forecast horizons. The majority of forecasts see improvements in the out of sample MSE: 7/10 at $h = 1Q$, 9/10 at $h = 2Q$, 6/10 at $h = 3Q$ and 9/10 at $h = 4Q$. The largest improvements by the MF-SUR model are found at the $h = 1Q$, for example France sees a 29% forecast error improvement and Spain realizes a 28% improvement. As the forecast horizon extends from $2Q$ to $4Q$, improvements of the MF-SUR over the mixed frequency OLS model are reduced from anywhere between 1% and 13%. Additionally, all forecast evaluation metrics with a asterisk besides them indicate that the MF-SUR is more accurate than the mixed frequency OLS model at the 10% level of significance according to the Diebold Mariano test. Germany, Belgium, Netherlands, Sweden and Canada see the largest improvements as per the Diebold Mariano test with two or more out the four forecast horizons indicating the MF-SUR is more accurate.

Table 3.6 divides the out of sample MSE of the matched SUR model by the MSE of the matched frequency OLS model at various forecast horizons. Results are similar to the mixed frequency comparison with the majority of forecasts showing improvement when using the SUR model: 8/10 at $h = 1Q$, 9/10 at $h = 2Q$, 9/10 at $h = 3Q$ and 5/10 at $h = 4Q$. Again the largest improvements are seen at the shortest forecast horizon, for example Italy has a 29% improvement in the MSE, with MSE improvements reducing to anywhere between 1% and 10% as the forecast horizon extends from $2Q$ to $4Q$. With this comparison the Diebold Mariano test does not identify as many forecasts that have higher accuracy at the 10% significance level compared to Table 3.5.

The significant improvements of the SUR model over the OLS model, for both mixed and matched frequency data, highlights some important areas where this paper adds to the existing GDP forecasting literature. Due to the relatively short sample size that we are constrained by it is likely there could be a degree of over-fitting when estimating model parameters via OLS. The way that SUR estimates model parameters, by re-weighting data using the covariance matrix of the OLS residuals, helps in some way to reduce over-fitting of extreme data points by effectively weighting them less. This makes SUR especially effective when the sample size is short. This confirms similar findings by Kmenta (1968) who compared five different methodologies to estimate a set of linear regression equations with mutually correlated disturbances. From Monte Carlo experiments we ran in Section 3.6 we also confirm this finding of SUR outperforming OLS in small sample sizes, but not larger sample sizes. Additionally Hoorsgate (2000) found that forecasting OECD GDP using a panel data approach outperforms OLS when the sample size is small, but not when the sample size is large. The limited data availability we experienced for all countries except the UK, US and Canada, especially when using higher frequency data, may explain why there is limited literature for GDP forecasting for those remaining seven countries. Therefore using SUR for GDP forecasting contributes to the existing GDP forecasting literature as an alternative method to use when the time series is relatively short.

Table 3.7 divides the out of sample MSE of the MF-SUR model by the MSE of the matched frequency SUR model at various forecast horizons, to examine the benefit of using mixed frequency data in the regressions. Out of the 40 forecasts across all countries and forecast horizons the MF-SUR only outperforms the matched frequency SUR 6/40 times. The most notable improvements are at the $1Q$ forecast horizon for USA and Canada, with 10% and 12% improvements respectively. This confirms Mitchell (2021) who finds that mixed frequency data can improve recession forecasts in USA and Canada at the 1 month to 3 month forecast horizon. However, overall across all countries the mixed frequency data provides little to no benefits and actually creates larger MSE's in most cases.

Finally, with regards to seemingly unrelated regressions, we compare how the correlation magnitude between different countries residuals impact the out of sample forecasting improvements. The top panel in Figure 3.2 compares average residual correlation coefficients for each country with the forecast improvement for the matched frequency $1Q$ ahead SUR model. The forecast improvement is measured as it is in Table 3.6, for example a value of 0.9 represents a 10% improvement in the mean squared error of the SUR model over the OLS model. The residual correlation shows the average correlation of that specific country with every other country in our sample. There is a clear relationship indicating that the higher the average residual correlation coefficient the greater the out of sample forecast improvement at the $1Q$ forecast horizon, shown by the downward sloping blue line estimated by OLS. The bottom panel in Figure 3.2 is the same comparison but for the MF-SUR model, where the forecast improvement is as measured in Table 3.5. Although the relationship between correlation and forecast improvement is not as strong as the matched model, the blue line still indicates a weak relationship between the two.

Additionally we compare the MF-OLS to the matched frequency OLS in Table 3.8. It is apparent that in the majority of the cases the matched frequency model outperforms the MF-OLS. Part of this reason may be the short sample size we use and hence the mixed frequency model over-fitting.

3.6 Monte Carlo Experiments

This section examines the properties of SUR vs OLS estimators to evaluate certain conditions and criteria of when one methodology should be used over the other. Additionally we are able to check whether our empirical results are robust and not just sample dependent. We set up the following data generating process (DGP) for our Monte Carlo experiments:

$$y_{1,t} = \alpha_1 + \beta_1 x_{1,t} + \epsilon_{1,t} \tag{45}$$

$$y_{2,t} = \alpha_2 + \beta_2 x_{2,t} + \epsilon_{2,t} \tag{46}$$

where x_i and ϵ_i for $i = 1, 2$ is sampled T times from $Ni.i.d.(0, 1)$, with $corr(\epsilon_1, \epsilon_2)$ set to ρ depending on the experiment we are running. We then compare how well SUR and OLS perform out of sample using different combinations of sample size and error correlation for $T = (25, 50, 75, 100)$ and $\rho = (0.2, 0.4, 0.6, 0.8)$. Each run is trained on $T/2$ of the data and out sample forecast errors are calculated for the remaining $T/2$ of the data. Parameter values are set to be $\nu = [\alpha_1,\ \beta_1,\ \alpha_2,\ \beta_2] = [0, 1, 0, 1]$.

Monte Carlo simulations are ran 1000 times for each combination and results are reported in Table 3.9. The ✔ symbol indicates SUR has a lower out of sample error at the 5% level of significance according to a 2 sample t-test. The $\times$ symbol indicates no statistically significant difference in the SUR and OLS out of sample error according to a 2 sample t-test. These results help to explain why SUR performs well out of sample compared to OLS in our empirical results. Generally speaking SUR performs well with small sample sizes and not necessarily high correlation in the error terms. When the sample size is 25 SUR has superior out of sample performance even when the errors have a low correlation of 0.2. We see similar results with a sample size of 50 where the correlation of the error terms needs to be equal to or above 0.4 for SUR to have better out of sample accuracy then OLS. However, as the sample size increases the out of sample performance of SUR and OLS converge and are not statistically significantly different, regardless of the correlation of the error terms. Further analysis leads us to believe OLS over-fits in small sample sizes relative to SUR, especially when there are outliers present (as can be the case with GDP), and therefore can have poor performance when T is low.

3.7 Conclusion

This paper collects a range of bond, stock, exchange rate and macroeconomic data for 10 different OECD countries. Using factors estimated from each of these asset markets as leading indicators we forecast their GDP using a seemingly unrelated regressions approach at various forecast horizons. Overall we find significant improvement in out of sample forecasting accuracy when estimating GDP via SUR as opposed to a benchmark OLS model that estimates each countries GDP equation separately. When using matched frequency data aggregated at the quarterly level we find that SUR reduces out of sample forecast error for 31/40 of the regressions run when compared to a benchmark OLS model. This improvement ranges from a reduction of 1% to 29% in the out of sample mean squared error. Results are similar when including mixed frequency data within the SUR model. In this case SUR also reduces out of sample forecast error for 31/40 of the regressions run when compared to a benchmark mixed frequency OLS model, with similar levels of magnitude with regards to error reduction. Given that our time series are relatively short, and from our Monte Carlo experiments, we also conclude that this large improvement of SUR over OLS contributes to the existing GDP forecasting literature as an alternative and effective methodology to be used when the researcher is constrained with a short sample size to forecast a certain country's GDP. This short sample size could be driven by data restrictions or structural breaks in GDP and its relationship with potential predictors.

Finally, when comparing a mixed frequency SUR model with a matched frequency SUR model, the mixed frequency model only reduces the out of sample mean squared error for 6/40 of the regressions run. This indicates that the benefit of including mixed frequency data in an SUR model is diminished, and actually matched frequency data should be used for the majority of the countries in our sample. We conclude that the poorer performance on average by the mixed frequency SUR model compared to the matched SUR model can be attributed to the difference in the residual correlation matrices. As spoken about in Section 3.5.2 we found that correlation across residuals was consistently lower in the mixed model versus the matched model. Our results prove the importance of the assumption of high correlation in the residuals for improved forecasting performance in the seemingly unrelated regression methodology.

3.8 Data Appendix

Data for each country is collected via the Global Financial Data (GFD) API. The numbers reported in Table 3.1 show how many factors were used in the predictive regression for each asset market and country. Lagged GDP is quarterly frequency, real activity data is monthly frequency, stock data is a mix of daily and monthly frequency, bond and exchange rate date is daily frequency.

Table 3.1: Number of factors used for each country

Country	**Lagged GDP**	**Stock**	**Bond**	**Exchange Rate**	**Real Activity**
Belgium	✔	2	2	1	2
Canada	✔	2	2	1	2
France	✔	3	2	1	2
Germany	✔	3	2	1	2
Italy	✔	1	2	1	2
Netherlands	✔	2	2	1	2
Spain	✔	2	2	1	1
Sweden	✔	2	2	1	1
UK	✔	1	2	2	2
USA	✔	2	2	1	2

Tables

Table 3.2: Average Residual Correlation, Matched SUR vs MF-SUR Full Sample

	Forecast Horizon			
Country	1Q	2Q	3Q	4Q
Germany	1.76	1.16	1.06	1.03
France	1.90	1.20	1.12	1.12
Italy	1.52	1.16	1.07	1.06
Spain	1.64	1.09	1.05	1.06
Belgium	1.41	1.09	1.06	1.03
Netherlands	1.54	1.10	1.03	1.06
UK	1.46	1.07	1.05	1.07
Sweden	1.53	1.13	1.04	1.07
USA	1.76	1.13	1.04	1.08
Canada	1.98	1.22	1.04	1.04

The average residual correlation from the first step of the SUR estimation of the matched frequency model divided by the MF-SUR model for each country and forecast horizon.

Table 3.3: MF-SUR vs Matched SUR, In Sample Results

	Forecast Horizon			
Country	1Q	2Q	3Q	4Q
Germany	0.44*	0.82*	0.96	0.97
France	0.22*	0.67*	0.97	0.90
Italy	0.71	0.97	0.94*	0.97
Spain	0.43	0.93	1.00	1.00
Belgium	0.59*	0.91	0.99	1.00
Netherlands	0.53*	0.95	0.96	0.96*
UK	0.56	1.02*	0.97	0.94
Sweden	0.77	0.92*	0.91*	0.90*
USA	0.28*	0.91*	0.89*	0.90*
Canada	0.40*	0.80*	0.96	0.99

The in sample mean squared error (MSE) of the MF-SUR model divided by the matched frequency SUR model for each country and forecast horizon. The * indicates that the MF-SUR has more accurate forecasts than the matched frequency SUR at the 10% level of significance, according to the Diebold Mariano test.

Table 3.4: MF-OLS vs Matched OLS, In Sample Results

	Forecast Horizon			
	1Q	**2Q**	**3Q**	**4Q**
Country				
Germany	0.43*	0.63*	0.86	0.94
France	0.22*	0.55*	0.80	0.77
Italy	0.65	0.96	0.88	0.93
Spain	0.44*	0.80*	0.94	0.95
Belgium	0.53	0.77*	0.89	0.99
Netherlands	0.43	0.76	0.89*	0.88
UK	0.52	0.87	0.92	0.87
Sweden	0.73	0.76	0.86*	0.85*
USA	0.31*	0.80*	0.80	0.81
Canada	0.44*	0.67*	0.89	0.94

The in sample mean squared error (MSE) of the MF-OLS model divided by the matched frequency OLS model for each country and forecast horizon. The * indicates that the MF-OLS has more accurate forecasts than the matched frequency OLS at the 10% level of significance, according to the Diebold Mariano test.

Table 3.5: MF-SUR vs MF-OLS, Out of Sample Results

	Forecast Horizon			
	1Q	**2Q**	**3Q**	**4Q**
Country				
Germany	1.04	0.88*	1.01	0.87*
France	0.71	0.98	0.98*	0.98
Italy	0.79	0.92*	0.98	0.98
Spain	0.72	0.95	1.02	0.94
Belgium	0.82	0.96*	0.96*	0.96*
Netherlands	1.12	0.93*	0.93*	0.99
UK	0.76	0.97	1.01	0.98
Sweden	0.97	0.98	0.92*	0.90*
USA	1.10	0.98	0.97	1.00
Canada	0.91*	1.00	1.00	0.99*

The out of sample mean squared error (MSE) of the MF-SUR model divided by the MF-OLS model for each country and forecast horizon. The * indicates that the MF-SUR has more accurate forecasts than the MF-OLS at the 10% level of significance, according to the Diebold Mariano test.

Table 3.6: Matched SUR vs Matched OLS, Out of Sample Results

	Forecast Horizon			
	1Q	2Q	3Q	4Q
Country				
Germany	1.05	0.94	0.96	1.01
France	0.77	0.96	0.98	1.00
Italy	0.71	0.95*	0.97	0.99
Spain	0.75	0.95	0.99	1.00
Belgium	0.78	0.95	0.96*	0.98
Netherlands	0.91	0.98	0.99	0.99
UK	0.78	0.97	0.98	0.98
Sweden	0.95	1.03	0.98	0.98
USA	0.98	0.90*	1.02	1.01
Canada	1.00	0.99*	0.99	1.00

The out of sample mean squared error (MSE) of the matched frequency SUR model divided by the matched frequency OLS model for each country and forecast horizon. The * indicates that the matched SUR has more accurate forecasts than the matched OLS at the 10% level of significance, according to the Diebold Mariano test.

Table 3.7: MF-SUR vs Matched SUR, Out of Sample Results

	Forecast Horizon			
	1Q	2Q	3Q	4Q
Country				
Germany	1.13	0.99	1.06	1.17
France	0.96	1.14	1.02	1.01
Italy	1.08	1.06	1.01	1.02
Spain	1.06	1.10	1.02	1.04
Belgium	1.02	1.04	1.07	1.05
Netherlands	1.23	1.19	1.05	1.11
UK	1.17	1.06	1.04	0.95*
Sweden	1.26	1.18	1.07	1.63
USA	0.90	1.13	0.95	1.03
Canada	0.88*	1.01	1.00	1.07

The out of sample mean squared error (MSE) of the MF-SUR model divided by the matched frequency SUR model for each country and forecast horizon. The * indicates that the MF-SUR has more accurate forecasts than the matched SUR at the 10% level of significance, according to the Diebold Mariano test.

Table 3.8: MF-OLS vs Matched OLS, Out of Sample Results

	Forecast Horizon			
	1Q	**2Q**	**3Q**	**4Q**
Country				
Germany	1.13	1.06	1.00	1.36
France	1.03	1.12	1.02	1.04
Italy	0.97	1.09	1.00	1.03
Spain	1.11	1.11	0.99	1.10
Belgium	0.98	1.02	1.07	1.06
Netherlands	1.00	1.25	1.11	1.10
UK	1.20	1.06	1.01	0.95*
Sweden	1.24	1.24	1.14	1.76
USA	0.80*	1.04	0.99	1.04
Canada	0.97	0.99	1.00	1.08

The out of sample mean squared error (MSE) of the MF-OLS model divided by the matched frequency OLS model for each country and forecast horizon. The * indicates that the MF-OLS has more accurate forecasts than the matched OLS at the 10% level of significance, according to the Diebold Mariano test.

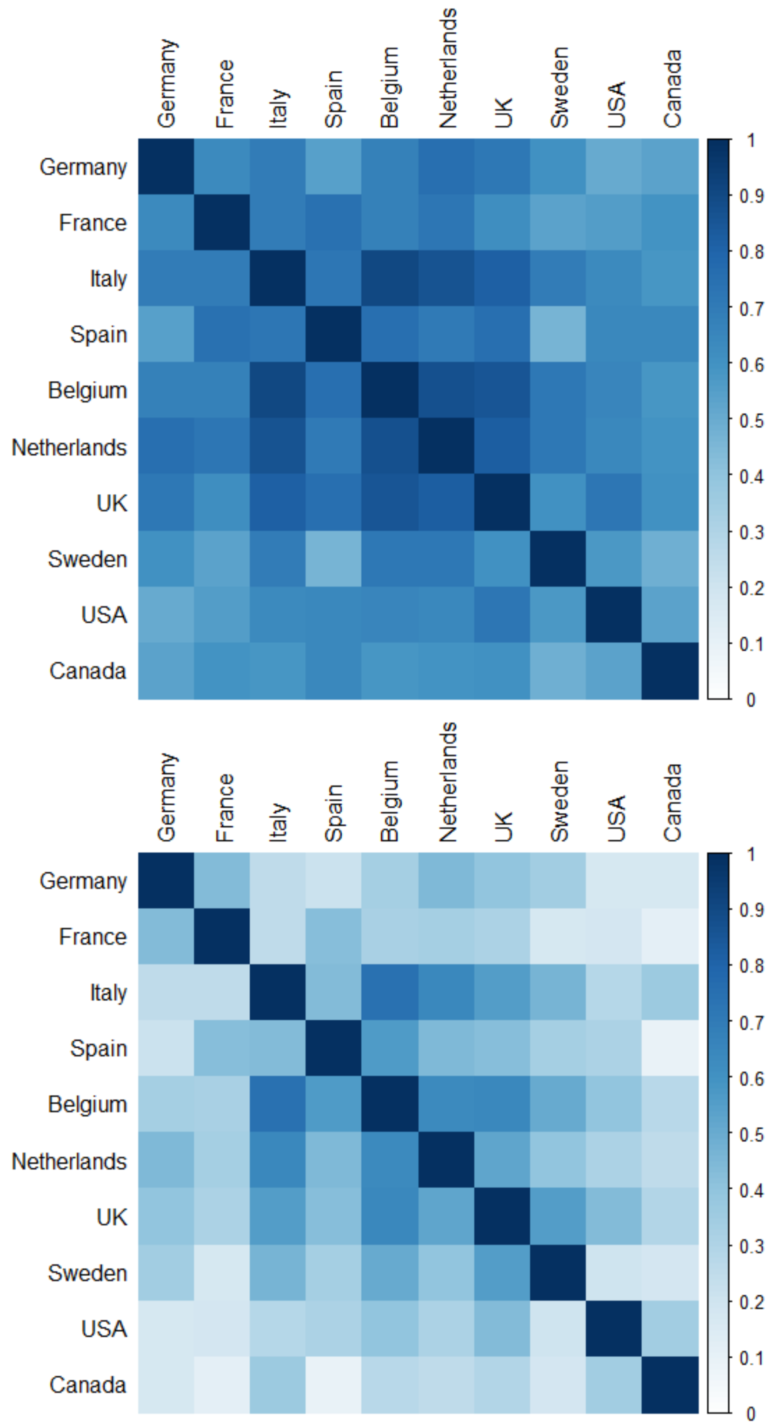
Table 3.9: Monte Carlo Experiments Results

	Error Correlation (ρ)			
	0.2	**0.4**	**0.6**	**0.8**
Sample Size (T)				
25	✔	✔	✔	✔
50	$\times$	✔	✔	✔
75	$\times$	$\times$	$\times$	✔
100	$\times$	$\times$	$\times$	$\times$

1000 Monte Carlo simulations are ran for each T and ρ combination to compare SUR to OLS. ✔ indicates SUR has a lower out of sample error at the 5% level of significance according to a 2 sample t-test. $\times$ indicates no statistically significant difference in the SUR and OLS out of sample error according to a 2 sample t-test.

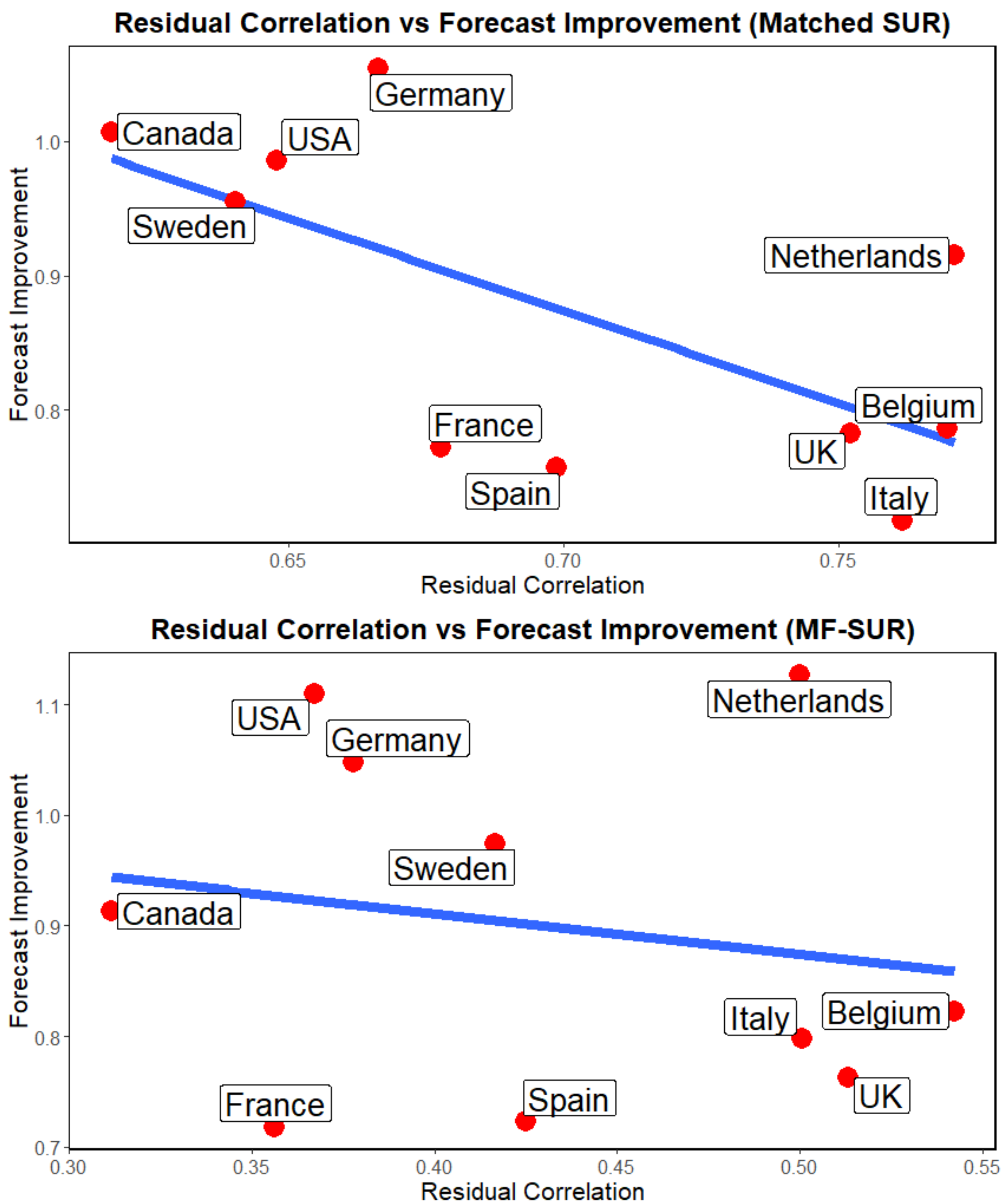
Graphs

Figure 3.1: Correlation matrix of residuals



Correlation matrix of residuals from the first step of the SUR estimation for the full sample $1Q$ ahead matched frequency SUR model (top) and MF-SUR model (bottom).

Figure 3.2: Out of sample results vs correlation matrix of residuals



The red dots show the average residual correlation between that country and all other countries (x axis) and the out of sample forecast improvement (y axis) at the $1Q$ ahead horizon for the matched frequency SUR model (top) and MF-SUR model (bottom). The forecast improvement is as measured in Table 3.6 for the top panel and Table 3.5 for the bottom panel. The blue line is estimated by OLS.

References

[1] Shirley Almon. The distributed lag between capital appropriations and expenditures. *Econometrica: Journal of the Econometric Society*, pages 178–196, 1965.

[2] Elena Andreou, Eric Ghysels, and Andros Kourtellos. Regression models with mixed sampling frequencies. *Journal of Econometrics*, 158(2):246–261, 2010.

[3] Elena Andreou, Eric Ghysels, and Andros Kourtellos. Should macroeconomic forecasters use daily financial data and how? *Journal of Business & Economic Statistics*, 31(2):240–251, 2013.

[4] Konstantin Arkadievich Kholodilin, Boriss Siliverstovs, and Stefan Kooths. A dynamic panel data approach to the forecasting of the gdp of german länder. *Spatial Economic Analysis*, 3(2):195–207, 2008.

[5] S Boragan Aruoba, Francis X Diebold, and Chiara Scotti. Real-time measurement of business conditions. *Journal of Business & Economic Statistics*, 27(4):417–427, 2009.

[6] Joseph Atta-Mensah, Greg Tkacz, et al. *Predicting Canadian recessions using financial variables: A probit approach*, volume 98. Bank of Canada, 1998.

[7] Francesco Audrino, Alexander Kostrov, and Juan-Pablo Ortega. Predicting us bank failures with midas logit models. *Journal of Financial and Quantitative Analysis*, 54(6):2575–2603, 2019.

[8] Mehmet Balcilar, Rangan Gupta, and Mawuli Segnon. The role of economic policy uncertainty in predicting us recessions: A mixed-frequency markov-switching vector autoregressive approach. *Economics: The Open-Access, Open-Assessment E-Journal*, 10(2016-27):1–20, 2016.

[9] Badi H Baltagi and Alain Pirotte. Seemingly unrelated regressions with spatial error components. *Empirical Economics*, 40(1):5–49, 2011.

[10] Artem Bequé and Stefan Lessmann. Extreme learning machines for credit scoring: An empirical evaluation. *Expert Systems with Applications*, 86:42–53, 2017.

[11] Travis J Berge and Òscar Jordà. Evaluating the classification of economic activity into recessions and expansions. *American Economic Journal: Macroeconomics*, 3(2):246–77, 2011.

[12] Henri Bernard and Stefan Gerlach. Does the term structure predict recessions? the international evidence. *International Journal of Finance & Economics*, 3(3):195–215, 1998.

[13] Marie Bessec and Othman Bouabdallah. Forecasting gdp over the business cycle in a multi-frequency and data-rich environment. *Oxford Bulletin of Economics and Statistics*, 77(3):360–384, 2015.

[14] R Bhansali. for multistep prediction of a time series: A review. *Asymptotics, nonparametrics, and time series*, 201, 1999.

[15] Vasco AP Cadavez and Arne Henningsen. The use of seemingly unrelated regression to predict the carcass composition of lambs. *Meat science*, 92(4):548–553, 2012.

[16] Maximo Camacho, Gabriel Perez Quiros, and Pilar Poncela. Green shoots and double dips in the euro area: A real time measure. *International Journal of Forecasting*, 30(3):520–535, 2014.

[17] Marcelle Chauvet and Jeremy Piger. A comparison of the real-time performance of business cycle dating methods. *Journal of Business & Economic Statistics*, 26(1):42–49, 2008.

[18] Yu-chin Chen and Wen-Jen Tsay. Forecasting commodity prices with mixed-frequency data: An ols-based generalized adl approach. *Available at SSRN 1782214*, 2011.

[19] Guillaume Chevillon and David F Hendry. Non-parametric direct multi-step estimation for forecasting economic processes. *International Journal of Forecasting*, 21(2):201–218, 2005.

[20] Eunsuk Chong, Chulwoo Han, and Frank C Park. Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications*, 83:187–205, 2017.

[21] Michael P Clements and David F Hendry. Multi-step estimation for forecasting. *Oxford Bulletin of Economics and Statistics*, 58(4):657–684, 1996.

[22] Arturo Estrella and Frederic S Mishkin. Predicting us recessions: Financial variables as leading indicators. *Review of Economics and Statistics*, 80(1):45–61, 1998.

[23] Laurent Ferrara and Clément Marsilli. Financial variables as leading indicators of gdp growth: Evidence from a midas approach during the great recession. *Applied Economics Letters*, 20(3):233–237, 2013.

[24] David F Findley. On the use of multiple models for multi-period forecasting. In *Proceedings of Business and Economic Statistics, American Statistical Association*, pages 528–531, 1983.

[25] David F Findley. Model selection for multi-step-ahead forecasting. *IFAC Proceedings Volumes*, 18(5):1039–1044, 1985.

[26] Claudia Foroni, Pierre Guérin, and Massimiliano Marcellino. Markov-switching mixed-frequency var models. *International Journal of Forecasting*, 31(3):692–711, 2015.

[27] Claudia Foroni and Massimiliano Giuseppe Marcellino. A survey of econometric methods for mixed-frequency data. *Available at SSRN 2268912*, 2013.

[28] Sebastian Fossati. Forecasting us recessions with macro factors. *Applied Economics*, 47(53):5726–5738, 2015.

[29] Sebastian Fossati, Rodrigo Sekkel, and Max Sties. Forecasting recessions in canada. 2018.

[30] Lennart Freitag et al. Default probabilities, cds premiums and downgrades: A probit-midas analysis. Technical report, 2014.

[31] Eric Ghysels, Pedro Santa-Clara, and Rossen Valkanov. The midas touch: Mixed data sampling regression models. 2004.

[32] Eric Ghysels, Arthur Sinko, and Rossen Valkanov. Midas regressions: Further results and new directions. *Econometric Reviews*, 26(1):53–90, 2007.

[33] Andrea Giusto and Jeremy Piger. Identifying business cycle turning points in real time with vector quantization. *International Journal of Forecasting*, 33(1):174–184, 2017.

[34] Anthony TC Goh. Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering*, 9(3):143–151, 1995.

[35] Luis M Gomez-Zamudio and Raul Ibarra. Are daily financial data useful for forecasting gdp? evidence from mexico. *Journal of the Latin American and Caribbean Economic Association*, 17(2):173–203, 2017.

[36] Lili Hao and Eric CY Ng. Predicting canadian recessions using dynamic probit modelling approaches. *Canadian Journal of Economics/Revue canadienne d'économique*, 44(4):1297–1330, 2011.

[37] Jeff Heaton. *Introduction to neural networks with Java*. Heaton Research, Inc., 2008.

[38] Andre J Hoogstrate, Franz C Palm, and Gerard A Pfann. Pooling in dynamic panel-data models: An application to forecasting gdp growth rates. *Journal of Business & Economic Statistics*, 18(3):274–283, 2000.

[39] Yu-Fan Huang and Richard Startz. Improved recession dating using stock market volatility. *International Journal of Forecasting*, 36(2):507–514, 2020.

[40] Bruce Jackson and Manjeet Rege. Machine learning for classification of economic recessions. In *2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI)*, pages 31–38. IEEE, 2019.

[41] Cuixia Jiang, Wei Xiong, Qifa Xu, and Yezheng Liu. Predicting default of listed companies in mainland china via u-midas logit model with group lasso penalty. *Finance Research Letters*, page 101487, 2020.

[42] Saurabh Karsoliya. Approximating number of hidden layer neurons in multiple hidden layer bpnn architecture. *International Journal of Engineering Trends and Technology*, 3(6):714–717, 2012.

[43] Munechika Katayama. Improving recession probability forecasts in the us economy, 2009.

[44] Khurshid Kiani. On forecasting recessions via neural nets. *Economic Bulletin*, 3(13):1–15, 2008.

[45] Jan Kmenta and Roy F Gilbert. Small sample properties of alternative estimators of seemingly unrelated regressions. *Journal of the American Statistical Association*, 63(324):1180–1200, 1968.

[46] Miroslaw Kordos. Data selection for neural networks. *Schedae Informaticae*, 25, 2016.

[47] Gitanjali Kumar. High-frequency real economic activity indicator for canada. Technical report, Bank of Canada Working Paper, 2013.

[48] Vladimir Kuzin, Massimiliano Marcellino, and Christian Schumacher. Midas vs. mixed-frequency var: Nowcasting gdp in the euro area. *International Journal of Forecasting*, 27(2):529–542, 2011.

[49] Jin-Lung Lin and Clive WJ Granger. Forecasting from non-linear models in practice. *Journal of Forecasting*, 13(1):1–9, 1994.

[50] Sydney C Ludvigson and Serena Ng. A factor analysis of bond risk premia. Technical report, National Bureau of Economic Research, 2009.

[51] Yifei Lyu, Jun Nie, and Shu-Kuei X Yang. Forecasting us economic growth in downturns using cross-country data. *Economics Letters*, 198:109668, 2021.

[52] Massimiliano Marcellino, James H Stock, and Mark W Watson. A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series. *Journal of econometrics*, 135(1-2):499–526, 2006.

[53] Henri Nyberg. Dynamic probit models and financial variables in recession forecasting. *Journal of Forecasting*, 29(1-2):215–230, 2010.

[54] Michael T Owyang, Jeremy Piger, and Howard J Wall. Forecasting national recessions using state-level data. *Journal of Money, Credit and Banking*, 47(5):847–866, 2015.

[55] Inske Pirschel. Forecasting euro area recessions in real-time. Technical report, Kiel Working Paper, 2016.

[56] Dadabada Pradeepkumar and Vadlamani Ravi. Forecasting financial time series volatility using particle swarm optimization trained quantile regression neural network. *Applied Soft Computing*, 58:35–52, 2017.

[57] Michael Puglia and Adam Tucker. Machine learning, the treasury yield curve and recession forecasting. 2020.

[58] Min Qi. Predicting us recessions with leading indicators via neural network models. *International journal of forecasting*, 17(3):383–401, 2001.

[59] Matthias Reif and Faisal Shafait. Efficient feature size reduction via predictive forward selection. *Pattern Recognition*, 47(4):1664–1673, 2014.

[60] James H Stock and Mark W Watson. Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, 41(3):788–829, 2003.

[61] James H Stock and Mark W Watson. New indexes of coincident and leading economic indicators. *NBER macroeconomics annual*, 4:351–394, 1989.

[62] Keshav P Vishwakarma. Recognizing business cycle turning points by means of a neural network. *Computational Economics*, 7(3):175–185, 1994.

[63] Halbert White. Learning in artificial neural networks: A statistical perspective. *Neural computation*, 1(4):425–464, 1989.

[64] Jonathan H Wright. The yield curve and predicting recessions. *Finance and Economics Discussion Series, Federal Reserve Board*, 2006.

[65] Qifa Xu, Xi Liu, Cuixia Jiang, and Keming Yu. Quantile autoregression neural network model with applications to evaluating value at risk. *Applied Soft Computing*, 49:1–12, 2016.

[66] Qifa Xu, Xingxuan Zhuo, Cuixia Jiang, and Yezheng Liu. An artificial neural network for mixed frequency data. *Expert Systems with Applications*, 118:127–139, 2019.

[67] Arnold Zellner. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American statistical Association*, 57(298):348–368, 1962.

[68] Dabin Zhang, Lean Yu, Shouyang Wang, and Haibin Xie. Neural network methods for forecasting turning points in economic time series: an asymmetric verification to business cycles. *Frontiers of Computer Science in China*, 4(2):254–262, 2010.