HEC MONTRÉAL

Predicting consumer spending before and during the pandemic: an analysis of the performance of machine learning methods and alternative data

par

Nicolas Neault

Sciences de la gestion
(Option Économie Appliquée)

Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences
(M. Sc.)

Décembre 2021

Abstract

The speed, severity and atypical nature of the Covid-19 pandemic made economic monitoring and forecasting particularly challenging. These challenges heightened the interest for larger and more timely sets of data, as well as for machine learning techniques able to properly handle them. This study investigates the performance of such techniques in understanding and predicting consumer spending both before and during the pandemic. We first construct a large panel of timely economic indicators which we supplement with alternative data in the form of Google Trends. We then pit a panel of theory-based models against machine learning models which namely differ on the number of predictors used and their specified form. Our results show that dense modelling techniques outperform sparse models during the pandemic. We also find a positive impact of using a large set of timely predictors and of supplementing it with Google data. Finally, despite improvements in forecast accuracy, forecast errors could possibly be reduced further by using forecast combination techniques or an intercept correction mechanism.

Keywords

Consumption, Covid, forecasting, machine learning, alternative data, Google Trends

Research methods

Time series analysis, vector autoregressions, machine learning, principal component regression, LASSO, random forest , forecasting horse race.

Sommaire

La rapidité et la force avec laquelle le virus Covid-19 s'est propagé ont créé plusieurs défis au niveau empirique. Ces défis ont suscité chez les économistes un fort intérêt pour les données alternatives et les procédures d'apprentissages statistiques. Cette étude vise à déterminer la contribution de ces techniques à la prédiction et à la compréhension des dépenses de consommation des ménages américains pendant la période précédant la pandémie ainsi que pendant la première année de la pandémie. Pour ce faire, nous construisons une base de données consistant de plusieurs indicateurs économiques et la supplémentons de données de recherche Google. Nous confrontons ensuite plusieurs modèles théoriques avec des modèles d'apprentissages statistiques qui diffèrent notamment sur le nombre de prédicteurs utilisés et sur la structure imposée au modèle. Nos résultats démontrent que les modèles utilisant un grand nombre d'indicateurs ont un plus grand pouvoir prédictif que les modèles parcimonieux pour l'année 2020. Nos résultats concluent aussi en faveur de l'utilisation de données alternatives. Finalement, malgré une amélioration notable de la précision, il serait probablement possible de réduire davantage les erreurs de prédictions à l'aide de combinaison de modèles ou de méthodes de correction basée sur la crise de 2008.

Mots-clés

Consommation, dépenses, contraintes de liquidités, Covid-19, pandémie, prédictions, apprentissage statistique, données alternatives, Google trends

Contents

Abstract **i**

Sommaire **ii**

List of Tables **v**

List of Figures **v**

List of acronyms **vi**

Acknowledgements **vii**

Introduction **1**

1 Literature review **4**

1.1 Understanding Consumption 4

1.2 Forecasting Models 9

1.3 Consumption during Covid-19 11

2 Data **13**

2.1 Target Variables 14

2.2 Economic Indicators 14

2.2.1 Measures of Income and Wealth 15

2.2.2 Other Economic Indicators 15

2.3 Google Trends Data 16

2.4 Data Treatment and Predictor Sets 18

3 Methodology **19**

3.1 Theory-Based Models 20

3.2 Selected Models 23

3.3 Prediction Periods and Estimation Procedure 28

3.4 Forecasting Approach and Horizons 29

4 Results **31**

4.1 Forecast Results - Models Performance 32

4.1.1 Quiet Period - One-Month Forecast 32

4.1.2 Quiet Period - Three-Month Forecast 34

4.1.3 Pandemic Year - One-Month Forecast 35

4.1.4 Pandemic Year - Three-Month Forecast 37

4.2 Forecast Results - Predictor Sets and Specifications 39

4.2.1 Predictor Sets 39

4.2.2 Specifications 41

4.3 Discussion 43

4.3.1 Results interpretation 43

4.3.2 Implications 45

4.3.3 Limitations 46

4.3.4 Recommendations 48

5 Conclusion **50**

Bibliography **52**

Appendix A i

Appendix B ii

Appendix C iv

List of Tables

4.1 Average Relative RMSPE for the Quiet Period - One-Month Forecast 33
4.2 Average Relative RMSPE for the Quiet Period - Three-Month Forecast 34
4.3 Relative RMSPE for the Pandemic Period - One-Month Forecast 35
4.4 Relative RMSPE for the Pandemic Period - Three-Month Forecast 37
4.5 RMSPE of the Best Models by Predictor Set 40
4.6 RMSPE of the Best Models by specification 41

1 List of Indicators and Data Treatment i
2 List of Models, Predictors Set and Specification ii
3 List of Principal Components and Their Association and (Explained Variance) iii
4 Predictor Importance for Select Regressions - QP1M iv
5 Predictor Importance for Select Regressions - QP1M vi
6 Predictor Importance for Select Regressions - QP3M vi

List of Figures

1 Regression result for Goods Consumption v

List of acronyms

LCH Life-Cycle Hypothesis

ML Machine Learning

MRF Macroeconomic Random Forest

OLS Ordinary Least Squares

PC Principal Component

PCR Principal Component Regression

PIH Permanent Income Hypothesis

POOS Pseudo Out-of-sample

RF Random Forest

RMSPE Root-Mean Square Prediction Error

VAR Vector Autoregression

Acknowledgements

I would like to thank my mentors and teachers at HEC Montréal whose teachings sparked my interest for economics and quantitative methods. In particular, I would like to thank Nicolas Vincent whose contagious passion for economics has been a great motivation in pursuing a master's degree in economics. Thank you for your availability and enthusiasm in answering my many questions over the years.

To my parents, Linda and Richard, I can't thank you enough for your support throughout my education. In many ways, this would not have been possible without you.

I also thank Alizée, whose unconditional support throughout this challenge has brought great comfort.

Finally, I would like to thank my friends and colleagues, without whom my time at HEC Montréal would not have been the same.

Introduction

When the Centers for Disease Control and Prevention (CDC) confirmed the first case of Coronavirus on U.S. soil on January 21st 2020, no one could have predicted the impact Covid-19 would ultimately have on the economy. After spreading throughout Asia, the virus spread within Europe and North America at an alarming speed and severity. Although rallying in early 2020, financial markets swiftly erased their annual gains and from February 14 until mid-March, the S&P500 suffered a 30 percent fall. The decline prompted the Federal Reserve (FED) to drastically reduce rates on two occasions: March 3 and March 13, the day a national emergency was declared by the U.S. government. In an effort to support the U.S. economy, the Federal Reserve announced further actions to provide liquidity in key markets. Extensions to the Fed's asset purchase program were announced later in the month. On the fiscal side, the Coronavirus Aid Relief and Economic Security (CARES) Act was announced providing more than 2 trillion dollars in aid.

As statewide lockdown measures started being imposed in late March and economists scrambled to make sense of the consequences of this unprecedented sanitary crisis, the incoming collapse in consumption expenditures, and mainly durable goods and services consumption, seemed to be a foregone conclusion. Survey and credit card transaction data showed how spending patterns quickly changed in the first months of the pandemic. After a brief period of stockpiling in early March, spending decreased drastically and, despite a bounce back in April, remained at a depressed level until the end of May. Survey respondents reported being highly concerned with their financial situation with more than half suffering a loss in income and wealth. Stimulus payments, however, led to a sharp increase in spending for liquidity constrained households.

Two-key features defined the early Covid-19 period. First, the pandemic spread rapidly and in turn, governments were swift to impose lockdown measures and to provide support to their citizens. Economists realized that traditional economic data, with its relatively low frequency and significant report lag, would not be enough to follow the impact of the pandemic on the economy in real time. Second, the economy was buffeted by many shocks. Lockdown measures and rising unemployment constituted a large supply shock. Stay-at-home orders also created a significant demand shock. Finally, oil prices plummeted at the same time contributing a third shock to the economy. In this context, generating forecasts proved to be a considerable task, even for short horizons.

Both these challenges encouraged market practitioners and economists to turn towards unconventional statistical methods; first by using alternative data sources, and secondly by basing their forecasts on empirical methods better suited to integrate large sets of variables. Although not quite as widely used as traditional empirical methods, it is now recognized that machine learning (ML) techniques may offer significant forecasting gains over traditional methods in times of recessions. Recent studies indicate that these gains also extend to the Covid-19 pandemic (Goulet Coulombe et al., 2021b). However, since these methods are aimed at producing forecasts rather than producing parameter estimates, some view ML as a black box with limited economic interpretation. One way of opening the black box is to compare a panel of models that differ with respect to their main features. Comparing the relative forecasting performance of many models highlights the key features that contribute to improved forecasts, and in this way, increases interpretability.

While there have been many descriptive accounts of changing spending patterns due to Covid-19 as well as a few forecasting case studies focused on the pandemic period, no research, to the best of our knowledge, has focused on forecasting personal consumption expenditures. This thesis bridges this gap by aiming to determine how machine learning and alternative data help in predicting, and understanding, American households' consumption behaviours both before and during the pandemic. By comparing a large panel of models, we intend to identify which model features contributed to greater forecasting accuracy and, in doing so, gain a better understanding of the variables driving consumer spending in the United States. Our work thus contributes to the dynamic field of machine learning forecasting. Using a large set of leading economic indicators and financial variables, supplemented by Google Trends data, we pit theory-based econometric models against statistical learning methods such as LASSO regressions, principal component regressions and random forest regressions. We find sizable forecasts improvements from using machine learning models during the pandemic over the best performing models during economic expansions. Such improvements are due to the combination of a large set of timely predictors and dense ML modelling techniques. We also find evidence of short-sightedness and financial frictions in consumer spending decisions during the pandemic.

This thesis begins by reviewing the relevant literature surrounding the theory of consumption, the use of machine learning in macroeconomic forecasting and a summary of the literature detailing consumption and forecasting during the Covid-19 pandemic in Section 1. Section 2 describes the data while Section 3 presents the methodology and describes the selected models. Section 4 presents and discusses the main empirical results. Finally, a conclusion is presented in Section 5.

1 Literature review

We separate the literature review in 3 sections. Section 1.1 reviews the main theories of consumption and serves as the basis for our panel of theory-based consumption models. Section 1.2 reviews the challenges that arise in estimating consumption with standard econometric models and present machine learning alternatives. Section 1.3 reviews the literature studying consumption and forecasting during the pandemic

1.1 Understanding Consumption

For most major economies, consumption is the main component of the demand for goods. Changes in consumption therefore lead to changes in production. Changes in production in turn lead to changes in income. Finally, households adjust their consumption expenditures in reaction to, first and foremost, changes in income (Keynes, 1936). This feedback mechanism can be expressed as:

$$C = c_0 + c_1 Y \tag{1.1}$$

$$Y = C + I + G \tag{1.2}$$

where c_0 represents autonomous consumption, i.e., the expenditures that must be undertaken by households regardless of the current economic situation (such as, among others, expenditure on food and shelter); and c_1 is the marginal propensity to consume and represents the marginal change in consumption for a marginal change in disposable income. Natural restrictions on c_1 are that it must be positive, and less than one. Thus, an increase in disposable income will lead to a less than one for one increase in consumption in the short run.

Substituting equation (1.1) into equation (1.2) and rearranging the terms highlights the importance of the marginal propensity to consume in amplifying short term economic fluctuations.

$$Y = \frac{1}{1-c_1}(c_0 + I + G) \tag{1.3}$$

Financial prudence embodied in greater savings (i.e., a lower propensity to consume or a lower value for autonomous consumption) dampens aggregate demand. Thus, Keynes argues that excessive savings, driven by psychological factors, can have a drastic impact on aggregate consumption and in turn on aggregate income.

While equation (1.1) fits consumption decisions across households quite well in microeconomic budget studies, two contradictory empirical facts stand out in aggregate data. First consumption is much smoother than current income. Second, aggregate consumption is essentially proportional to aggregate income in the long run.

$$C_t = c_1 Y_t \tag{1.4}$$

The theory of intertemporal consumption reconciles the theory with the empirical evidence by proposing that lifetime consumption decisions be based on lifetime resources rather than on current income. In its simplest form:

$$\sum_{t=1}^{T} C_t \leq A_0 + \sum_{t=1}^{T} Y_t \tag{1.5}$$

where the left-hand side of equation (1.6) represents lifetime consumption and the right-hand side represents lifetime resources (the sum of initial wealth A_0 and the present value of lifetime income). Individual preferences dictate the optimal fraction of lifetime resources to consume each year. For instance, a household willing to divide its lifetime resources equally across the T years of its life consumes the following amount:

$$C_t = \frac{1}{T}\left(A_0 + \sum_{t=1}^{T} Y_t\right) \tag{1.6}$$

Consumption then grows with permanent income $\frac{1}{T}\sum_{t=1}^{T} Y_t$ and thus appears to grow proportionally to current income over the long term. Since permanent income is a weighted average of lifetime income, it should be less volatile than income. Therefore, consumption is less volatile than current income.

Both the Life Cycle Hypothesis (LCH) of Modigliani and Brumberg (1954) and the Permanent Income Hypothesis (PIH) of Friedman (1957) are variations of the theory of intertemporal consumption. These theories propose that households desire a smooth consumption path throughout their life. Households therefore borrow or dis-save in periods of low income relative to their lifetime average (such as in early life and retirement) and accumulate capital when income is relatively high (such as in working years). The theory of inter-temporal choice demonstrates how rational behaviours, can account for patterns of aggregate savings: savings are simply future consumption.

An empirical complication of the LCH and PIH is the dependence of consumption on expected future income. Expected future income is itself not time invariant since changes in the economic environment cause rational households to adjust their expectations (Lucas, 1976). Therefore, to model consumption one must simultaneously model income and the dynamic process in which income expectations are generated by households.

Hall (1978) demonstrates the implications of rational expectations in the PIH framework. A consumer maximizes lifetime utility subject to a sequential budget constraint:

$$\max_{c_t, A_t} \; E_t \sum_{j=0}^{s} \beta^j u(c_{t+j}) \tag{1.7}$$

$$s.c. \;\; y_t + A_{t-1} = c_t + \frac{1}{1+r} A_t \tag{1.8}$$

where equation (1.7) is the present value of lifetime utility discounted using the household's subjective discount factor β; equation (1.8) stipulates that at every period, households have a choice to spend current income and accumulated assets on consumption expenditures or asset purchases. The Euler equation representing the optimal path from the consumer problem can be written as follows:

$$u'(c_t) = \beta(1+r)E_t u'(c_{t+1}) \tag{1.9}$$

Equalizing the subjective discount factor β and the market discount factor $\frac{1}{1+r}$ equalizes the marginal utility $u'(E_t[c_{t+j}])$ of all periods. Additionally, for linear specifications of $u'(c_t)$, such as the one arising from a quadratic utility function $u(c_t) = \left(C_t - \frac{a}{2}C_t^2\right)$, the marginal utility function exhibits certainty equivalence behaviour.

$$u'(c_t) = E_t u'(c_{t+1}) = u'(E_t c_{t+1}) \tag{1.10}$$

$$c_t = E_t c_{t+1} = E_t c_{t+2} = \cdots = c \tag{1.11}$$

In the presence of certainty equivalence, households value uncertain consumption as if it were certain (1.10). With $\beta(1+r) = 1$, the household desires a flat consumption path (1.11). For an infinitely lived household $T \to \infty$ (e.g., a dynasty) with no terminal assets ($A_T = 0$), the realized consumption plan which satisfies equation (1.6) is:

$$c_t = \frac{r}{1+r}\left[A_{t-1} + E_t \sum_{j=0}^{T} \frac{y_{t+j}}{(1+r)^j}\right] \tag{1.12}$$

And after a few manipulations expected change in consumption is:

$$\Delta c_{t+1} = r \sum_{j=1}^{T} \frac{(E_{t+1} - E_t) y_{t+j}}{(1+r)^j} \tag{1.13}$$

which states that the contemporaneous change in consumption is equal to a fraction of the revised expectation $E_{t+1} - E_t$ of the household's permanent income. Since households already incorporate all information concerning their income prospects in their consumption decision, only unanticipated changes in permanent income can impact consumption. Thus equation (1.13) implies that no past information can be used to forecast changes in consumption. Therefore, the rational-expectation permanent income hypothesis (RE-PIH) states that consumption follows a random walk.

$$c_{t+1} = E_t c_{t+1} + \varepsilon_{t+1} = c_t + \varepsilon_{t+1} \tag{1.14}$$

Hall then proceeds to empirically test the RE-PIH and finds it surprisingly hard to reject - detecting only weak evidence that past stock market prices help in predicting short-term consumption changes - sparking a large empirical literature in the process. Fortunately for forecasters, almost all subsequent tests in the literature reject the random-walk hypothesis.

Empirical evidence shows that consumption responds to predictable movements in income even after accounting for unexpected changes in income (Flavin, 1981). This finding is also robust to the inclusion of durable goods in total consumption (Bernanke, 1985). Furthermore, for many income processes exhibiting strong positive autocorrelations or non-stationarity, consumption varies less than the RE-PIH predicts to unanticipated changes in permanent income (Deaton, 1987).

Omitted information explains how consumption can be both excessively sensitive to predictable changes and excessively smooth to unpredictable changes in income (Campbell and Deaton, 1989). Households have access to more information than economists in generating their own income forecasts. Therefore, what appears unanticipated to the model builder might be partially known by the household. That is, a fraction α of the observed innovation is known before its realization, therefore the change in consumption is a weighted average of the first difference in labour income Δy_t and the expected revision in permanent income ε_{yp} (Flavin, 1993).

$$\Delta c_{t+1} = \alpha \Delta y_{t+1} + (1-\alpha) r \sum_{j=1}^{T} \frac{(E_{t+1} - E_t) y_{t+j}}{(1+r)^j} = \alpha \Delta y_{t+1} + (1-\alpha)\varepsilon_{yp} \tag{1.15}$$

The weight α can be interpreted as the "Keynesian" marginal propensity to consume out of current income. This excess sensitivity parameter merely reflects the household's contemporary adjustment to the income revision it had already anticipated. Thus, households are slower to fully adjust their consumption in reaction to innovations in income than predicted by the theory. This slowness explains both the departures from the RE-PIH hypothesis.

There are many factors that may explain such slow and gradual adjustment. First, households might be unable to borrow funds against their future earnings at their desired rate, liquidity problems constrain them to a lower level of consumption. Alternatively, households might form consumption habits that take time to adjust as new information is made available. Finally, after excluding certainty equivalence, there is a precautionary motive for saving: households may set aside current consumption as a buffer to absorb unexpected future shortfalls in income.

1.2 Forecasting Models

Sims (1980) heavily criticizes the one-equation-at-a-time specification of large models prevalent at the time. For instance, if consumption is to be modelled with a demand equation that is a function of income and income is in turn modelled by a labour supply equation, an equation by equation approach separates the impact of demand on supply, and likewise separates the impact of supply on demand. Such restrictions on the equations, to Sims, are an incredible assumption to impose on a model. Furthermore, such parameter restrictions, potentially exclude useful relations that are present in the data.

The author presents vector autoregressions (VARs) as a way to map such relations and as a reliable alternative for forecasting. A VAR is a N x 1 vector Y of variables of interest that is regressed on the p past values of itself.

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} \tag{1.16}$$

Albeit requiring a much smaller set of assumptions from the economist, a VAR requires the estimation of p x N^2 parameters. A technical problem is that as N grows larger, the model quickly loses degrees of freedom, which reduces the effective sample size. Models estimated on a relatively smaller sample are at a greater risk of modelling the idiosyncratic variations that are unique to the observations within the sample. Models which have "overfit" idiosyncratic errors are likely to have higher forecast variance. Thus, there is a trade-off that must be weighed between approximation and estimation error:

$$\underbrace{y_{t+h} - \hat{y}_{t+h}}_{\text{Forecast Error}} = \underbrace{f^*(X_t) - f(X_t)}_{\text{Approximation Error}} + \underbrace{f(X_t) - \hat{f}(X_t)}_{\text{Estimation Error}} + \underbrace{\varepsilon_{t+h}}_{\text{Irreducible error}} \tag{1.17}$$

where $f^*(X_t)$ is the true model; $f(X_t)$ is the specified model; and $\hat{f}(X_t)$ is the estimated model. Adding variables to a model potentially brings it closer to the true model, thus reducing bias, but increases the variance of the forecast due to the estimation error. $\hat{f}(X_t)$ can in turn be brought closer to the specified model by increasing sample size. ε_t is the irreducible error that contains the unmeasured variables and unmeasured variations.

The limited number of observations in macroeconomic time series and the unlimited number of predictors an economist might want to use gives rise to this "curse of dimensionality." Sims explains that structural models, despite their shaky foundations, may outperform VARs in forecasting due to their relatively fewer number of parameters. This highlighted the need to develop practical methods of limiting the number of parameters. There are two main approaches that seek to turn large predictor sets into more concise information.

Dense modelling techniques such as the principal component regression (Stock and Watson, 2002a) recognize that all variables might be jointly important for prediction although their individual contribution might be small. These techniques aim at extracting the important latent relations present in the data and using them as predictors. On the opposite end, sparse modelling techniques such as the LASSO (Tibshirani, 1996) focus on selecting variables with the highest predictive power out of all possible predictors. Some statistical algorithms, such as the random forest ensemble (Breiman, 2001), have the flexibility to be used as either a dense or, due to its variable selection feature, as a sparse modelling technique.

These methods, although able to partially correct for the curse of dimensionality, may in themselves introduce some bias. If the true model is a dense model, imposing a sparse statistical method may increase approximation error. Likewise, if the true data generating process is non-linear, imposing a linear model will increase the forecasting error.

Finding the best model specification thus often requires pitting models against each other in a pseudo-out-of-sample (POOS) forecasting horse race. Results are usually highly dependent on the time period and data series used. POOS horse races concluding in favour of dynamic factor models are presented by Stock and Watson (2002b), Kim and Swanson (2018), Smeekes and Wijler (2018), although the last authors also find evidence in favour of LASSO shrinkage methods for certain series as in Li and Chen (2014). Finally, Madeiros et al. (2019), Goulet Coulombe et al. (2021b) and Chen et al. (2019) obtain results in favour of using random forests ensemble methods. In sum, model uncertainty is pervasive (Giannone et al., 2021) and the best prediction is often obtained as a weighted average of several econometric models.

Reviewing the literature on forecast output, Chauvet and Potter (2013) identify that there are marked gains in using separate forecasting models for normal times and recessions. Dynamic factor models, such as the principal component regression, offer significant gains in accuracy during recessions. Siliverstovs and Wochner (2019) and Kotchoni et al. (2019) also present evidence that the use of a large dataset and machine learning techniques offer significant forecasting gains in recessions.

Chen et al. (2019) extend their dataset with alternative data sources, such as Google searches and credit card data, and test different model specifications out-of-sample. The authors find that alternative data offer increased accuracy albeit with diminishing returns.

1.3 Consumption during Covid-19

Given the fast spread of the virus and the swift reaction of the economy, understanding consumption spending during Covid-19, requires the use of non-standard economic data. Survey data collected from 10000 households by Coibion et al. (2020) detail how households adjusted to changes in income and expectations. Independent sets of debit and credit card transaction data used by Baker et al. (2020), Cox et al. (2020), and Chetty et al. (2020), show the evolution of consumer spending and liquidity holdings at a high frequency.

Half of respondents surveyed by Coibion et al. (2020) report an average income loss of 5293$ and an average wealth loss of 33482$ during the first months of the pandemic. Households expected both a higher unemployment rate and higher uncertainty. Households also anticipated the downturn to have a persistent negative effect. In sum, survey results pointed to a large future decline in aggregate demand.

Congruently, respondents also reported being highly concerned with their financial situation. In response households namely moved their assets from foreign stocks to more liquid forms of savings and postponed debt payments. Empirical evidence from Cox et al. (2020) show a year-over-year increase in liquid balance of 36% for the month of

May. Furthermore, households with low levels of liquidity had the largest decline in March spending (Baker et al., 2020). Finally, stimulus payment led to a sharp increase in spending for these same households, which is consistent with excess sensitivity due to financial frictions (Chetty et al., 2020).

Consistent with the high uncertainty associated to the virus, stockpiling behaviours from households led to an increase in spending in the first half of March. This initial increase was quickly followed by a sharp decrease in spending (Baker et al., 2020). Although expenditures started increasing with the stimulus in mid-April, they still remained depressed until the end of May (Cox et al., 2020).

Card transaction data also reveal many changes in spending patterns. First consumption was brought forward in the early days of the pandemic as households stockpiled on necessity (Hall et al., 2020; Baker et al., 2020). Second, declines in certain hard-hit sectors such as restaurants and accommodation were almost offset 1 for 1 with spending in food and beverage stores in March (Dunn et al., 2020; Carbajo, 2021). Third, contrary to what happens during most economic downturns, spending on luxury goods which do not require physical contact did not fall while services spending, both in essential and non-essential categories, fell sharply (Cox et al., 2020; Chetty et al., 2020). Finally, spending fell moderately more for high-income households, least affected by the pandemic, than for low-income households (Chetty et al., 2020). These changing patterns are primarily driven by health concerns rather than by the traditional linkages with income or wealth. Such new patterns are thus unlikely to be accurately predicted by models fit on past observations.

Foroni et al. (2021) test several forecast improvement methods in simple mixed-frequency model and identify a form of intercept correction, using forecasting errors from the great recession, as the best way to improve forecast accuracy. Finally, in a Covid-19 pandemic recession case study, Goulet Coulombe et al. (2021a) find substantial gains from using machine learning methods.

2 Data

As highlighted in the literature review, the main theoretical appeal of using machine learning models over standard statistical methods is their ability to sift through large quantity of data and extract the key elements useful for forecasting. This ability permits the use of larger predictor sets and thus facilitates the integration of timely information.

To build our predictor sets, we first select economic indicators for their timeliness and their relevance for consumption forecasting from a theory standpoint. All our predictors are reported on at least a monthly basis and are reported before personal consumption expenditures and disposable income. Indicators are namely chosen for their ability to explain shifts in income and wealth, consumer sentiment, expectations about future income and credit conditions.

We then supplement our selection of economic indicators with Google search data. Google data may offer some economic intuition by reflecting online purchases or by revealing future in-store purchase intentions. Google searches may also offer insights in general consumer sentiment. These variables are available on a daily basis.

This section details the data, the data treatment process, and the predictor sets. As is typical in the machine learning literature, we will refer to dependent variables as targets and refer to independent variables as predictors throughout this paper. A list of indicators and their transformation are presented in Appendix A.

Section 2.1 presents the values of consumption used as targets. Section 2.2 describes the economic indicators. Section 2.3 discusses Google data. Section 2.4 explains the data treatment process and defines our predictor sets.

2.1 Target Variables

The target variables are the personal consumption expenditure and its main categories: consumption of goods and services. We further disaggregate the consumption of goods in its two categories: durable goods and nondurable goods. Consumption of durables is composed of expenditures on motor vehicles and parts, furnishings and durable household equipment, recreational goods and vehicles, and other durable goods. Consumption of nondurables is composed of expenditures on food and beverages purchased for off-premise consumption, clothing and footwear, gasoline and other energy goods, and other nondurable goods. We chose to focus on goods consumption and services consumption is thus not disaggregated into its components.

The data is obtained from the Bureau of Economic Analysis (BEA), from the National Income and Product Account (NIPA), in the underlying Table 2.3.6. "Real Personal Consumption Expenditures by Major Type of Product and by Major Function." The data is already corrected for inflation and seasonally adjusted by the BEA.

2.2 Economic Indicators

Our predictors include measures of income and wealth, consumer sentiment and leading economic indicators. The inclusion of such measures helps in mapping revisions in expectations of permanent income. Our set of leading economic indicators is largely based on the index of leading indicators used by the Conference Board. We also use a range of monetary and financial indicators to capture the potential effects of financial frictions on the spending process. Unless otherwise mentioned, the data series are all obtained from the Federal Reserve Economic Data (FRED) database.

2.2.1 Measures of Income and Wealth

Personal disposable income is used as a measure of income. We opt to use personal disposable income over labour income because it includes government transfers and thus reflects the effects of fiscal policy. Government transfers play an important role in economic downturns. The measure is obtained from the BEA and is corrected for inflation and seasonality.

The value of the Standard and Poor's 500 index (S&P500) is used as a measure of financial wealth. No measures of housing wealth are used since they are reported after income and personal consumption expenditures.

2.2.2 Other Economic Indicators

As measures of consumer confidence, we use the OECD Consumer Confidence Index for the United States as well as the Index of Consumer Sentiment from the University of Michigan. We also use the "Current Index" and "Expected Index" subdivisions of the Consumer Sentiment index. These surveys detail consumer expectations for inflation, market prices and interest rates. Overall they give a good account of spending intentions.

To reflect the health of the labour market we use the average weekly hours of all employees in the manufacturing sector as well as the weekly initial claims for unemployment insurance. The New Private Housing Units Authorized by Building Permits measure is used as an early indicator of the health of the housing market.

Measures of financial risk are incorporated in the form of the BBB US Corporate Index Option-Adjusted Spread and the High Yield Index Option-Adjusted Spread to reflect perceived credit risk in financial markets. The CBOE Volatility Index (VIX) is used as a measure of the volatility of financial markets. The monetary base and the M2 money stock are added as general liquidity measures of the economy. Finally, measures of interest rates such as the effective Federal Funds Rates and the 10-Year Treasury Constant Maturity Rate are used to reflect the cost of liquidity.

The Trade Weighted U.S. Dollar Index: Broad Goods is included to reflect shifts in relative purchasing power. Commodity prices such as crude oil (WTI and Brent), copper and gold are also added for their role in affecting purchasing decisions and their role in reflecting cyclical variations. Namely, gold is generally seen as a safe haven by savers and its price movements as indicators of perceived risk in the economy. Similarly, positive copper price movements are seen as a bullish indicator by market participants.

2.3 Google Trends Data

Google search data have been made available for free on the Google Trends website since January 2004. Google Trends allow researchers to track the interest over time of categories of Google searches in different countries and regions of the globe. Albeit being free, there are many obstacles to using Google data in a quantitative setting. For instance, there are an unlimited number of keywords one can search, words can have multiple meanings, and there are several different languages representing the same search even though actual search queries are different. To correct for these problems, Google aggregates search queries into categories.

Selecting a category such as "vehicle brands (815)" with keyword "Saturn" allows you to obtain data for the search interest in the car brand Saturn exclusively (i.e., excluding searches for the planet). Categories also allow researchers to avoid having to specify keywords for their search. For instance, an empty search query with category "vehicle shopping" (473) will return the interest in all of the top searches falling within this category (e.g. "car dealership", "sales", "used cars" etc.)

To correct for the keyword issue, we exclusively download search categories without including any keyword in the search query. We preselect specific Google trends categories based on either their ability to help predict a subcomponent of consumption or their ability to reflect households' sentiment regarding their economic situation. As an example, search interest in category 11, "home and garden," might help predict expenditures in furnishings and durable household equipment. Search interest in category 60, "jobs," may offer insights into the labour market's trends and frictions.

Other technical issues in using Google trends data include:

1. The data is time frame sensitive for value: data is constrained between 0 and 100 where 100 is the date with the highest search interest within the time frame. Other values are set relative to this data point. This issue makes it hard to concatenate different time frames together.

2. The data is time frame sensitive for frequency: to obtain daily data one must specify a time frame no longer than 9 months, to obtain weekly data one must specify a time frame between 9 months and 5 years, otherwise monthly data are generated. To obtain a long time horizon of either daily or weekly observations, data must be downloaded in multiple datasets, with each dataset containing a different time frame, and concatenated. Each of these multiple datasets represents one search query and is therefore time intensive to download manually.

3. The data collection methodology has been upgraded in 2011 and in 2016. Due to these changes, most series exhibit a negative trend from 2004 to 2011 and an upward shift between December 2015 and January 2016.

To circumvent these problems, we use Python PyTrends API allowing us to download directly any Google Trends series directly into a Pandas dataframe. To obtain weekly data spanning the entire data series we download a set of 5 different dataframes per category. The data is transformed into weekly growth rates before being concatenated into one frame. To correct for the two methodology changes, we force the growth rate between the last week of December 2010 and first week of January 2011 to be of 0. A similar treatment is done to correct for the shift in 2015-2016. The first observation of the dataset is set equal to 100 and a new series is generated by computing the cumulative product of each weekly observation. This results in a time series displaying the cumulative growth of the search interest in the variable indexed at a 100 at its origin.

2.4 Data Treatment and Predictor Sets

Daily and weekly indicators are transformed into monthly indicators by taking their monthly average. We then adjust series that contain seasonal patterns and are not already corrected for seasonality using the X-13-ARIMA-SEATS seasonal adjustment program. In doing so, we provide untreated data with the same treatment used by the Federal Reserve Economic Data (FRED) from which we source most of our data.

We then test seasonally adjusted series for stationarity using a set of five advanced Dickey-Fuller (ADF) tests for integration. More specifically we compute ADF tests for the variable in level, first order difference (month over month), second order difference (month over month change in first order difference), seasonal difference (year over year) and first-order seasonal difference (month over month change in seasonal difference).

We confirm results from the ADF test using autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. The ACF and PACF plots display the decay of autocorrelations and partial autocorrelations at different lags for all variable specifications. A stationary variable should exhibit a rapid decay of autocorrelations towards 0 in the ACF plot and rapidly display a partial autocorrelation of 0 in the PACF plot.

Most data display first-order integration and are made stationary by taking the difference of the logged variable. By de-seasonalizing and de-trending data, we ensure that the cyclical component of each series is properly isolated and is ready for empirical work.

We group all of our data in two different predictor sets. First, our "Large" predictor set includes all of the macroeconomic indicators. Second, our "Large + Google" predictor set supplements the "Large" predictor set with Google searches.

Since all predictors are released before the targets, contemporaneous forecasts of the target variables are possible with the same-month observations of our predictors. Forecasts with same month observations can be done by ordinary least squares (OLS). Likewise forecasts using past observations can be done by vector autoregressions (VAR).

3 Methodology

We supply our two large predictor sets to our selection of models and compare its ability to forecast consumption against models inspired by the theory. Our selected models can be separated in two distinct categories: dense models and sparse models. Both categories attempt to partially correct for the curse of dimensionality but differ in their treatment of large predictor sets.

Given the timeliness of our predictors, these models can perform OLS forecasts with same month observations of our predictors and VAR forecasts with past observations of our predictors. That is, for each dense and sparse model, and each predictor set, we estimate an OLS and a VAR specification. A summary table of the models, their predictor sets and their specifications is presented in Appendix B.

As is common in the machine learning forecasting literature, we test our panel of models in a pseudo out-of-sample forecasting horse race. Each model is tasked with forecasting measures of consumption during two different prediction periods and at two different forecast horizons. The forecasting performance of every model is then measured using the root mean squared prediction error (RMSPE).

Section 3.1 presents the models inspired by the theory of consumption. Section 3.2 presents our selection of dense and sparse models. Section 3.3 presents the prediction periods under study and the estimation procedure. Finally, section 3.4 presents the forecasting approach and the two forecast horizons.

3.1 Theory-Based Models

Our limited information models are composed of relevant benchmarks and models inspired by the theory of consumption. The most commonly used benchmark in machine learning forecasting is a simple univariate autoregression. Other widely used benchmarks are the random walk and the random walk with drift models. These two models are also of theoretical relevance for our analysis due to the random walk result obtained by Hall (1978) and are included in our panel of theory-based models. We also add vector autoregression models with predictors inspired by the excess sensitivity hypothesis and the permanent income life cycle hypothesis to our panel.

AR(p)

A univariate autoregressive model of order p (AR(p)) is a linear combination of past values of the target variable up to p lags. It can be written as:

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \tag{3.1}$$

where y_t is the target; δ is a constant; and ε_t is a white noise error. ϕ_p is the p_{th} order coefficient representing the association between the contemporaneous observation of the target variable and its realization p lags in the past.

The parameter p is chosen to ensure that the errors ε follow a white noise process. A white noise process has an expected mean of 0, an expected variance that is constant and has uncorrelated realizations. Among the valid number of lag orders to use, we select the value which minimizes the Akaike Information Criterion (AIC). Once p is established, the $\phi_{1,2,\ldots,p}$ coefficients are chosen to minimize the sum of squared residuals.

Since a one-period-ahead forecast for an AR(p) is the regression equation led by one period, using an AR(p) model allows us to model a feedback loop such as the one described by Keynes (1936), where an increase in consumption in one period leads to higher consumption in the following period. For instance, using an AR(1):

$$\Delta y_t = \phi_1 \Delta y_{t-1} + \varepsilon_t \tag{3.2}$$

$$E_t \Delta y_{t+1} = E_t[\phi_1 \Delta y_t + \varepsilon_{t+1}] \tag{3.3}$$

$$E_t \Delta y_{t+1} = \phi_1 \Delta y_t \tag{3.4}$$

the white noise term ε_{t+1}, due to its random nature, is unknown and unpredictable when making a forecast. Thus, the term disappears once applying the expectation operator and can be interpreted as a prediction error:

$$\Delta y_{t+1} - E_t \Delta y_{t+1} = \varepsilon_{t+1} \tag{3.5}$$

Taking the square of each prediction error and computing the average across the prediction period returns the out-of-sample mean squared prediction error. Taking the root of this result returns the root mean squared prediction error (RMSPE) which is our out-of-sample measure of fit to evaluate the forecast accuracy of our models.

Random Walks

A random walk is characterized by the following process:

$$y_t = y_{t-1} + \varepsilon_t \tag{3.6}$$

$$y_t = y_0 + \sum_{j=0}^{t-1} \varepsilon_{t-j} \tag{3.7}$$

where ε is a white noise process. Each random realization of ε imparts a permanent change in the level of y_t. Furthermore, given the random stochastic nature of ε, the best forecast for the change in the target variable y_t is the expected mean of the white noise process, which is 0.

$$y_t - y_{t-1} = \varepsilon_t \tag{3.8}$$

$$E_t \Delta Y_{t+1} = E_t[\varepsilon_{t+1}] = 0 \tag{3.9}$$

One way of modelling the rational expectation permanent income hypothesis random walk result is thus by assuming no period-to-period change in the target variable.

Another possible route is to assume that the change in the target is equal to a constant, with the constant set as the average historical growth rate of the target. That is, the random walk can have either a positive or negative drift over the long term. Such a model takes the following form:

$$\Delta Y_t = \delta + \varepsilon_t \tag{3.10}$$

$$E_t \Delta Y_{t+1} = E_t[\delta + \varepsilon_{t+1}] = \delta \tag{3.11}$$

Both of these models are useful benchmarks against which to pit other models due to their simplicity and to their theoretical relevance.

VAR(p)

A vector autoregression model of order p, VAR(p), is an extension of an AR(p) in a multivariate setting. For instance, a VAR(1) with 2 variables takes the following form:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix} \tag{3.12}$$

where u_1 and u_2 are white noise processes; ϕ_{12} links the first variable to the second variable's first lag; ϕ_{11} links the first variable to its first lag.

A VAR(p) model can be redefined as an AR(p) model using matrix notation:

$$Y_t = \Phi Y_{t-1} + U_t \tag{3.13}$$

Therefore the parameter selection p, the estimation procedure and forecasting procedure are the same as for the AR(p) presented before. The two following specifications of Y_t are considered:

$$Y_t = \begin{bmatrix} Consumption \\ Disposable\ Income \end{bmatrix} \qquad Y_t = \begin{bmatrix} Consumption \\ Disposable\ Income \\ S\&P500 \end{bmatrix} \tag{3.14}$$

where the first makes uses of the observation by Flavin (1981) that contemporaneous consumption is more sensitive to past observations of consumption and disposable income than predicted by the rational expectation permanent income hypothesis. The second uses a measure of wealth along with income as predictors as is typical in the life cycle hypothesis of Modigliani and Brumberg (1954) and permanent income hypothesis of Friedman (1957).

3.2 Selected Models

Our selected models all use our two large predictor sets. This panel of models is divided into two categories: dense information models and sparse information models.

Dense models consider all of the available predictors when making a prediction. This family of model includes our fat linear regressions (which include all the available predictors), the principal component regressions (which reduce the entire set of predictors into a few common factors) and the random forest ensemble.

Sparse information models select only a subset of the entire predictor set to make predictions. This category is composed of models such as the LASSO regression (which penalizes and removes predictors with less explanatory power) and linear specifications of models such as the random forest (which uses predictors selected by the algorithm in a linear regression).

Finally, to set apart the importance from using the timeliest observations from using past observations, each model has two different specifications. First an ordinary least squares (OLS) specification which produces forecasts using only contemporaneous same-month observation of our predictors. Second, a vector autoregression (VAR) specification which produces forecasts based only on past values of our predictors and of the target variable. The random forest ensemble has an OLS and VAR specification only in its sparse linear form, not in its dense form.

Fat Regressions

The Fat VAR follows the same VAR(p) methodology as theory-based models but uses all of the available predictors. One caveat of using a large number of predictors in a linear model involves the loss of degrees of freedom. A VAR requires the estimation of p x N^2 parameters. A large number of predictors N restricts the maximum number of lags p that can be considered. For instance, the "Large" predictor set is limited to a maximum of 2 lags. The "Large + Google" predictor set is limited to 1 lag.

The Fat OLS projects y_t on the entire predictor set X_t and takes the following form:

$$y_t = \beta_0 + \beta X_t + \varepsilon_t \tag{3.15}$$

By using all available predictors, such "kitchen sink" regressions partially correct some of the bias from omitting potentially important information. But, due to the number of parameters that need to be estimated, these regressions have a high potential of modelling the noise unique to the data sample on which it is fit. Thus, it tends to "overfit" idiosyncratic error which can increase out-of-sample forecast variability. Given the number of parameters, another caveat is the difficulty of performing inference from the coefficients of the regression.

Despite such problems, these regressions are good models against which to compare the predictive ability of the other large predictor set models. Furthermore, a Fat regression is a special case of a Lasso regression in which predictors are not penalized and are therefore not removed from the regression.

Principal Component Regressions

Principal Component Analysis transforms the data by projecting it onto a set of orthogonal axes. A large number of predictors are then reduced to a limited set of principal components (PC).

$$X_t = \Lambda F_t + u_t \tag{3.16}$$

A principal component can be likened to a factor F_t: a latent variable that is responsible for the co-movements of all predictor series. The first PC captures the direction along which the data vary the most. The second PC is set orthogonal to the first PC and captures the second direction along which the data vary the most, and so forth for further principal components. Using this method, a few principal components are able to explain a large percentage of the variation present in the data.

For estimation, all predictor series are first transformed into z-scores by subtracting each observation by the series' sample mean and then dividing by the series' sample variance. Standardizing the entire set of predictors is necessary to moderate the importance high-variance predictors tend to have in the final principal components obtained.

We then compute and select a number of principal components which explains at least 90% of the variance in the data. This restriction results in four principal components. These principal components are then used as explanatory variables in an OLS and VAR(p) model as before. The model takes the following form:

$$X_t = \beta F_t + \varepsilon_t \tag{3.17}$$

$$y_t = \delta + \phi(L)y_t + \beta(L)F_t + \varepsilon_t \tag{3.18}$$

where (L) is the lag operator specifying the number of lags to use; and F_t are the extracted components (factors) from our set of predictors X_t.

Although not a predictor selection method, we are able to infer how each predictor influence the different principal components both by looking at correlation between the predictors and the extracted PC, and by looking at the magnitude of each predictor's eigenvalue for each PC. We find that our first PC is mostly associated with credit measures (high yield spread, investment grade spread), our second PC is mostly associated with unemployment measures (mainly initial claims), our third PC with the 10 year bond rate and our fourth PC with the volatility index (VIX). Our four factors can thus be hypothesized as capturing credit risk, employment uncertainty, liquidity of money and financial markets uncertainty. That is, all four factors represent either credit or future income uncertainty. Furthermore, these can be loosely associated with financial frictions.

Random Forest Regressions

Dense

Tree-based methods segment the set of possible values into J distinct and non-overlapping regions. For each distinct region R_j, the predicted outcome is the mean of all observations (y_i) found within that region.

$$R_j^- = i : x_j < \theta \tag{3.19}$$

$$R_j^+ = i : x_j \geq \theta \tag{3.20}$$

Segmentation thresholds θ are chosen to minimize the sum of squared residuals given by:

$$SSR = \sum_{i \in R^-} (y_i - \bar{y}^-)^2 + \sum_{i \in R^+} (y_i - \bar{y}^+)^2 \tag{3.21}$$

where $\bar{y}^-$ and $\bar{y}^+$ represent the mean of all outcomes within region R_j^+ and R_j^-. Each region R_j is then further partitioned until the fit of the model is no longer improved with additional splits. What results from the algorithm is a prediction that follows a set of splitting rules which can be summarized in a tree.

A caveat of tree-based methods is their sensitivity to small changes in data. To reduce dependence on a particular set of data (i.e. to minimize overfitting), two improvements can be made. The first involves fitting a large number of trees (we use 1000 trees in our model) with different subsets of the dataset. Each subset of data is generated via a bootstrap algorithm. Tree predictions are then aggregated and the "forest" prediction is set equal to the average of all predictions produced by the individual trees. For B bootstraps the prediction takes the form:

$$\hat{y}_i = \frac{1}{B} \sum_{b=1}^{B} f_B(x_i) \tag{3.22}$$

The second improves on the first by de-correlating the trees. De-correlation is accomplished by adding randomness in the number of predictors considered at each split. In our case, we set the number of predictors to be considered at each split to be equal to the square root of the number of predictors. That is for e.g. 36 indicators, only 6 will be selected at random to generate a splitting rule at each different node. The use of many trees and the addition of randomness at each split results in a random forest regression.

Sparse

As for principal component analysis, we are able to infer how important individual predictors are in generating the random forest forecasts. In regression trees, each split reduces the sum of squared residuals (SSR). An important predictor is thus one that generates a large decrease in the SSR once used in a split. As with forecasts, we can aggregate and average the impact each predictor has in reducing the SSR in each of their respective tree. We then rank the variables in terms of importance and use the top 4 variables as predictors in simple OLS and VAR(p) regressions. The choice of using four variables is based on a desire to put this regression on equal footing with the four summary variables generated by the PCA approach and close to the three variables used by the PIH/LCH VAR.

LASSO

The LASSO regression is a least squares regression that is supplemented by a penalty.

$$\hat{\beta} = (X'X + \lambda I)^1 X'Y \tag{3.23}$$

First, as for the principal component regression, all variables are standardized by subtracting the mean and dividing by the variance. This reduces the influence of both the variance and scale of predictors on the estimated coefficients $\hat{\beta}_i$. Parameter estimates are then obtained by minimizing the sum of squared residuals supplemented by the penalty which is the l_1 norm in the case of a Lasso:

$$PSS = \sum_{i=1}^{n}(y_i - \sum_{j=1}^{J} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{J} |\beta_j| \tag{3.24}$$

$$PSS = SSR + l_1 \tag{3.25}$$

The l_1 penalty has the effect of shrinking all estimated coefficients toward zero and in doing so, removes variables with lesser explanatory power from the regression. The Lasso regression therefore results into a sparse model by selecting only predictors with strong explanatory power over the target variable. This process reduces the out-of-sample variance of forecasts that is induced by using a large number of predictors at the cost of potentially introducing some bias in an otherwise unbiased least squares regression.

The parameter λ determines the importance of the penalty on the resulting model and the resulting shrinkage that is performed. A parameter λ of 0 returns the least squares fitted model (a Fat regression) whereas, a sufficiently large λ returns a regression without predictors (an AR for the VAR specification or a random walk for the OLS specification). The λ parameter must thus be finely tuned using k-fold cross-validation on the dataset. k-fold cross-validation is performed by dividing the dataset into k groups, using $k-1$ subsets to train the model and testing on the kth left out model. The procedure is repeated k times and each time forecasting errors are collected. A grid search is performed to find the value of λ which minimizes the cross-validation error (CVE):

$$CVE = \frac{1}{k}\sum_{i=1}^{k} MSE \tag{3.26}$$

LASSO regressions are then performed. As for the random forest, if available, the top 4 variables are selected and introduced in both an OLS regression and a VAR(p) regression.

3.3 Prediction Periods and Estimation Procedure

It is suggested in the literature that the gains from using machine learning and a large predictor set are concentrated in periods of recession. To assess the difference between an expansionary and a recessionary economic environment we consider two prediction periods. First, our "quieter" period - exempt of the excess noise which is characteristic of a recessionary episode - extends from January 2015 to December 2019. Second, our recessionary episode is the first year of the Covid-19 pandemic shock and spans from January 2020 to December 2020. Comparing the first year of the pandemic to a baseline "quiet" period helps us confront the theory of consumption with the data in a time period where uncertainty is high and consumer confidence is low.

For each prediction period, the model is fit on data preceding the prediction year. For instance, forecasts for the year 2015 are produced with a model fit on the data spanning from January 2006 - the beginning of our sample - to December 2014. Likewise, predictions for the year 2016 are produced with a model fit on the data spanning from January 2006 to December 2015. And so on for the other years composing our two prediction periods.

Once model parameters are estimated, model forecasts are generated for each month of the prediction year. Each monthly forecast is performed with all the data available up to that month. Since consumption expenditure and disposable income are released with a significant lag, same-month observations of all our other predictors are available for generating the consumption forecasts. For instance, January observations of our predictors are available to generate the January prediction of our target variable. Likewise, February observations are available to generate the February prediction. And so on for the other months of the prediction year.

3.4 Forecasting Approach and Horizons

Given the recursive nature of autoregressive models, forecasting errors tend to compound over the length of the forecast horizon. Using an AR(1) as an example:

$$E_t y_{t+1} = \delta + \phi y_t \tag{3.27}$$

$$E_t y_{t+2} = \delta + \phi E_t y_{t+1} \tag{3.28}$$

$$E_t y_{t+3} = \delta + \phi E_t y_{t+2} = \delta + \phi(\delta + \phi E_t y_{t+1}) \tag{3.29}$$

such that the errors in forecasting y_{t+1} are integrated into the forecast of y_{t+2} and in turn errors in forecasting y_{t+2} are integrated into the forecast of y_{t+3}. Such an approach to forecasting is named the iterative approach. Thus short term forecast errors result in long-term forecast errors.

To avoid this issue from the iterative approach, the standard practice in machine learning applications is to use direct predictive modelling (Goulet Coulombe, 2020). This alternative approach redefines the target variable y_{t+h} as a weighted average of its h future realizations. y_{t+h} is then projected on the predictor set, and the forecast is made using the most recent observations.

Given that our independent variables are first-order integrated, following Stock and Watson (2002b) we define y_{t+h} as the average growth rate over the period $[t+1, t+h]$:

$$y_{t+h} = (1/h)ln(Y_{t+h}/Y_t) \tag{3.30}$$

then perform the following regression:

$$y_{t+h} = \delta + f(X_t) + \varepsilon_t \tag{3.31}$$

and compare the model performance for one-step and three-step ahead forecasts $h = [1,3]$. Setting $h = 3$ is equivalent to predicting the average over the next three months. This procedure tests the various models' ability to forecast the length and severity of downswings/upswings during the pandemic. This also allows us to test the different models' performance once some of the variation in the target variable has been smoothed.

As per Goulet Coulombe (2020) we limit the forecast horizon to a maximum of three months. A practical concern for this horizon choice is the speed of the decline and the recovery. Most of the fluctuations happened over a course of three months and most consumption variables were quick to stabilize after the first three months. Considering longer horizon, with our forecasting methodology, would further smooth out the shock we are ultimately trying to predict.

From a theoretical perspective, as emphasized by Keynes, even though long-term dynamics are usually of interest for forecasters, short-term fluctuations may have a drastic impact on aggregate consumption and income in the long term. During periods of high uncertainty, very short-term forecasting is then of primary importance for formulating stabilization policies.

4 Results

Our panel of models are tasked with forecasting consumption for two different prediction periods and at two different forecast horizons. Given the number of models and the number of consumption subcategories to be predicted, there is a substantial number of results to be presented. To ease up the presentation we separate the forecasts first by prediction period and then by forecast horizon. Thus, we present four different tables each representing a forecasting task (a prediction period and a forecast horizon).

Our main justifications for the use of sparse and dense models are to integrate a large predictor set and to use the timeliest information available for forecasting. To further ease up the presentation, we investigate the importance of our selected predictor sets in two summary tables in a separate section. The first table compares the best models for each predictor set: theory, Large, and Large+Google. The second table compares the best performing model for each model specification: OLS and VAR.

Finally, given the density of our results, we reserve most of their interpretation in a separate discussion where we also highlight the main implications and limitations of our results and issue our recommendations.

We present the results in three sections. Section 4.1 presents the forecast performance of each model for all subcategories of consumption, prediction period and forecast horizon. Section 4.2 presents the summary tables studying the impact of our choice of predictor sets. Finally, section 4.3 interprets and discusses the results. Tables supporting our results are presented in Appendix B.

4.1 Forecast Results - Models Performance

This section reviews the best performing models for the quiet prediction period and the pandemic period at a one-month and three-month forecast horizon. For each prediction period and forecast horizon, we identify the best models and the main features which seem to have contributed to their success. Thus, relative outperformance from a model allows us to infer about the driving forces behind household behaviours. In turn, comparing results between the quiet period and the pandemic period helps uncover the idiosyncrasies originating from the impact of the pandemic. A deeper analysis of the role played by our large sets of timely predictors is presented in Section 4.2.

As is typical in the forecasting literature, we use the root-mean squared prediction error of a simple AR(p) as the baseline forecast. For interpretability, the root mean squared prediction error (RMSPE) of all other models are presented relative to the baseline's RMSPE. Given the nature of the Covid-19 shock and its asymmetric impact on different industries, presenting the results relative to a baseline allows the forecasts of different subcategories to be compared to each other. The best performing model for each spending subcategory is highlighted in bold in the results table.

4.1.1 Quiet Period - One-Month Forecast

Table 4.1 displays the average of the yearly RMSPE values for the five years preceding 2020, which we refer to as the Quiet Period (QP). These results are for a one-month-ahead forecast. Relative RMSPE values are all close to one which shows it is hard to improve over the AR(p) baseline for this forecasting task. This suggests that past values of consumption are key indicators.

Both theory-based VARs perform equally or better than the benchmark for most categories of consumption. However, improvements are slim and usually range between 0 to 5%. Personal disposable income appears to be the main contributor of such improvements. However, the inclusion of the S&P500 contribute positively to the PIH/LCH VAR forecast of nondurables, gas, and other durables with marginal increases over the Excess Sensitivity VAR of 2%, 6%, and 5% respectively.

Table 4.1: Average Relative RMSPE for the Quiet Period - One-Month Forecast

	Total	**Services**	**Goods**	**NDur**	**Food**	**Cloth**	**Gas**	**Other**	**Dur**	**Vehic**	**Furn**	**Rec.**	**Other**
AR Baseline (RMSPE)	0.26	0.18	0.65	0.52	0.6	1.16	1.25	0.62	0.99	1.76	1.11	1.55	1.15
Panel A: Theory													
Random Walk	1.15	1.17	1.06	1.13	1.12	1.10	1.22	1.06	1.08	1.07	1.01	1.10	1.07
Random Walk Drift	1.00	**1.00**	1.02	1.12	1.12	1.09	1.22	1.02	1.01	1.07	0.95	0.99	1.03
Excess Sensitivity VAR	**0.96**	**1.00**	**0.95**	0.98	0.98	0.98	1.01	**0.98**	**0.97**	**0.98**	1.00	**0.95**	0.98
PIH / LCH VAR	1.00	**1.00**	0.97	**0.96**	0.97	0.97	0.95	1.00	1.00	1.02	0.98	0.95	**0.93**
Panel B: Dense													
Fat VAR	1.08	1.11	1.03	1.08	1.05	1.15	1.03	1.03	1.08	1.32	1.01	0.95	1.03
Fat VAR (Google)	1.38	1.56	1.20	1.31	1.12	1.24	1.37	1.31	1.28	1.49	1.31	1.03	1.26
Fat OLS	1.04	1.06	1.03	1.13	1.15	1.12	1.28	1.08	1.13	1.31	1.05	1.06	1.15
Fat OLS (Google)	1.38	1.56	1.48	1.44	1.15	1.47	2.02	1.45	1.83	2.94	1.27	1.32	1.44
PCA VAR	1.15	1.06	1.09	1.04	0.97	1.10	1.02	**0.98**	1.03	1.41	1.03	0.99	1.05
PCA VAR (Google)	1.19	1.06	1.03	1.04	0.97	1.10	1.02	**0.98**	0.99	1.30	1.06	0.97	1.07
PCA OLS	1.15	1.17	1.05	1.13	1.12	1.12	1.24	1.06	1.08	1.10	1.00	1.10	1.09
PCA OLS (Google)	1.15	1.17	1.06	1.13	1.13	1.12	1.24	1.06	1.08	1.10	1.01	1.10	1.09
Random Forest	1.00	**1.00**	1.02	1.15	1.13	1.10	1.26	1.05	1.07	1.15	**0.95**	1.00	1.03
Random Forest (Google)	1.00	**1.00**	1.02	1.17	1.13	1.10	1.29	1.08	1.00	1.07	**0.95**	1.01	1.04
Panel C: Sparse													
Lasso VAR	**0.96**	**1.00**	1.08	1.00	0.97	1.00	1.02	**0.98**	1.11	1.22	0.97	0.97	1.06
Lasso VAR (Google)	**0.96**	**1.00**	1.05	1.00	0.97	**0.96**	1.00	1.03	1.08	1.15	1.00	0.97	1.08
Lasso OLS	1.12	1.17	1.03	1.13	1.13	1.10	1.24	1.05	1.11	1.24	1.05	1.10	1.10
Lasso OLS (Google)	1.08	1.17	1.12	1.13	1.12	1.16	1.22	1.13	1.05	1.12	1.05	1.10	1.08
Random Forest VAR	1.08	1.11	1.06	0.98	0.97	1.04	0.94	1.02	1.17	1.23	0.96	0.96	0.97
Random Forest VAR (Google)	1.00	1.06	1.03	1.04	**0.95**	1.03	**0.93**	1.05	1.12	1.26	0.97	0.97	0.99
Random Forest OLS	1.15	1.17	0.98	1.13	1.15	1.11	1.26	1.06	1.05	1.22	1.05	1.10	1.10
Random Forest OLS (Google)	1.19	1.44	1.02	1.19	1.17	1.15	1.26	1.27	1.13	1.49	1.06	1.10	1.08

Sparse machine learning models using a VAR specification offer some improvements over the AR benchmark for the subcategories of nondurable goods and a few select subcategories of durable goods. The Lasso VAR and Random Forest VAR outperform by small amounts in forecasting food (5%), clothing (4%) and gasoline consumption (3%). These results can namely be attributed to better predictor selection and to the use of alternative data. Google category 951 "kitchen and dining", category 263 "sporting goods", and category 273 "motorcycle" are all selected as important predictors.

The random forest ensemble outperforms for furniture spending (5%) by prioritizing initial claims, building permits and the investment grade credit spread in its forecast. However, it underperforms for all other subcategories. Dense models and OLS specifications generally perform poorly for almost all subcategories of spending.

4.1.2 Quiet Period - Three-Month Forecast

Table 4.2 displays the average of the yearly RMSPE values for the quiet period at a three-month-ahead horizon. As for the one-month-ahead forecast, very few models are able to improve upon the AR(p) benchmark. Among theory-based models, the random walk with drift is the best performing specification with improvements of up to 20%, which supports the rational expectations permanent income hypothesis.

Table 4.2: Average Relative RMSPE for the Quiet Period - Three-Month Forecast

	Total	**Services**	**Goods**	**NDur**	**Food**	**Cloth**	**Gas**	**Other**	**Dur**	**Vehicles**	**Furn**	**Rec.**	**Other**
AR Baseline (RMSE)	0.33	0.42	0.3	0.31	0.9	1.44	0.85	0.41	0.79	1.35	0.59	0.76	1.02
Panel A: Theory													
Random Walk	1.03	**1.00**	1.13	1.03	0.99	1.00	0.99	1.05	1.04	1.01	1.19	1.22	1.04
Random Walk Drift	**0.97**	**1.00**	**0.90**	0.97	**0.98**	1.01	0.96	**0.90**	**0.95**	1.01	**1.00**	0.80	1.03
Excess Sensitivity Var	1.00	**1.00**	1.00	1.03	1.00	1.01	1.00	1.00	1.00	**1.00**	1.05	1.00	**1.00**
PIH / LCH VAR	1.00	1.02	1.00	1.03	1.00	1.01	1.00	1.00	1.00	1.02	1.08	1.03	1.01
Panel B: Dense													
Fat VAR	1.03	1.02	1.13	1.03	1.01	1.03	1.01	0.98	1.10	1.22	1.14	0.97	1.06
Fat VAR (Google)	1.12	1.05	1.93	1.26	1.03	1.13	1.42	1.12	1.71	2.01	1.41	1.22	1.20
Fat OLS	**0.97**	**1.00**	1.03	1.06	0.99	1.00	1.01	0.98	1.01	1.17	1.15	0.91	1.03
Fat OLS (Google)	1.06	**1.00**	1.60	1.13	1.02	1.01	1.34	1.07	1.53	1.56	1.90	1.41	1.25
PCA VAR	1.03	1.02	1.57	1.03	1.00	1.10	1.39	1.02	1.51	1.81	1.25	1.04	1.19
PCA VAR (Google)	1.03	1.02	1.60	1.06	1.01	1.10	1.40	1.02	1.53	1.84	1.25	1.04	1.18
PCA OLS	1.03	**1.00**	1.10	1.03	0.99	1.01	0.98	1.02	1.03	1.04	1.15	1.21	1.05
PCA OLS (Google)	1.03	**1.00**	1.10	1.03	0.99	1.01	0.99	1.02	1.04	1.04	1.15	1.21	1.05
Random Forest	**0.97**	**1.00**	**0.90**	1.00	0.99	1.01	**0.94**	0.93	0.99	1.07	1.07	0.82	1.05
Random Forest (Google)	1.00	**1.00**	**0.90**	**0.94**	**0.98**	1.01	0.96	**0.90**	0.99	1.08	1.02	**0.79**	1.04
Panel C: Sparse													
Lasso VAR	1.03	1.02	1.60	1.06	1.00	1.02	1.40	1.02	1.53	**1.00**	1.25	1.04	1.18
Lasso VAR (Google)	1.06	1.02	1.47	1.03	1.00	1.10	1.41	0.98	1.24	**1.00**	1.31	1.08	1.17
Lasso OLS	1.00	**1.00**	1.10	1.10	0.99	1.00	1.01	1.05	1.03	1.01	1.17	1.24	1.02
Lasso OLS (Google)	1.00	**1.00**	1.10	1.06	0.99	0.99	0.99	1.07	1.04	1.01	1.17	1.21	1.02
Random Forest VAR	1.03	1.02	1.43	1.00	1.00	**0.99**	1.61	1.05	1.47	1.77	1.31	1.20	1.19
Random Forest VAR (Google)	1.03	1.02	1.50	1.03	0.99	1.01	1.62	1.00	1.43	1.77	1.32	1.12	1.20
Random Forest OLS	1.03	**1.00**	1.17	1.10	1.00	1.00	0.99	1.07	1.06	1.08	1.19	1.20	1.04
Random Forest OLS (Google)	1.00	**1.00**	1.13	1.06	**0.98**	0.99	1.01	1.07	1.08	1.08	1.20	1.21	1.04

The random forest ensemble model exhibits strong performance, namely in its dense ensemble form but also sometimes in its sparse feature selection form. As an ensemble, it rivals the random walk as the best forecasting method for total (3%), goods (10%), recreational goods (21%), and for the majority of subcategories of nondurable goods with improvements of up to 10%. This relative outperformance over other dense models indicates the algorithm's superior ability to sift through large quantities of data without modelling too much noise in the process. The inclusion of Google data also tends to improve most of its forecast, which further supports this claim.

4.1.3 Pandemic Year - One-Month Forecast

Table 4.3 displays the set of results for the one-month-ahead forecast during the Covid-19 year. As opposed to the quiet time period, models based on theory no longer outperform other models. The PIH/LCH VAR outperforms other theory-based models for most subcategories and most notably durable goods. This can be attributed to the inclusion of the S&P500 as an indicator. Where the PIH/LCH VAR fails to improve, the random walk is usually the best model for the group of theory-based models.

Table 4.3: Relative RMSPE for the Pandemic Period - One-Month Forecast

	Total	**Services**	**Goods**	**NDur**	**Food**	**Cloth**	**Gas**	**Other**	**Dur**	**Vehic**	**Furn**	**Rec.**	**Other**
AR Baseline (RMSPE)	5.95	6.06	5.6	4.68	6.6	18.33	13.71	3.2	9.2	12.25	8.93	6.44	15.26
Panel A: Theory													
Random Walk	0.87	0.86	1.03	1.08	1.13	0.94	0.80	1.10	0.98	0.99	0.91	0.99	0.96
Random Walk Drift	0.87	0.86	1.03	1.08	1.13	0.94	0.80	1.09	0.98	0.99	0.91	0.97	0.96
Excess Sensitivity VAR	0.96	1.00	0.94	0.93	1.00	0.98	1.00	0.88	0.92	1.01	0.93	0.87	1.00
PIH / LCH VAR	0.91	1.01	0.86	0.88	0.96	0.95	1.05	0.83	0.87	0.93	0.88	**0.85**	1.01
Panel B: Dense													
Fat VAR	0.67	0.85	0.69	0.57	**0.92**	0.87	0.74	0.70	1.04	1.36	0.77	1.29	0.77
Fat VAR (Google)	**0.62**	0.85	**0.63**	**0.46**	0.93	**0.83**	**0.66**	**0.62**	1.06	1.48	**0.67**	1.23	**0.74**
Fat OLS	0.90	0.92	1.02	1.09	1.17	0.88	0.80	1.12	0.93	1.05	0.73	0.92	1.13
Fat OLS (Google)	0.90	0.86	1.09	1.12	1.11	0.94	0.81	1.12	1.08	1.50	0.78	0.88	1.11
PCA VAR	0.94	1.06	0.94	1.07	0.99	0.88	1.56	1.16	0.99	1.41	0.87	1.45	0.85
PCA VAR (Google)	0.95	1.03	0.95	1.09	0.93	0.92	1.34	1.25	0.98	1.36	0.90	1.44	0.88
PCA OLS	0.79	**0.82**	0.94	1.07	1.18	0.92	0.82	1.15	0.78	0.74	0.71	0.86	0.90
PCA OLS (Google)	0.79	**0.82**	0.94	1.07	1.17	0.91	0.82	1.13	**0.77**	**0.72**	0.70	0.87	0.89
Random Forest	0.85	0.86	1.00	1.07	1.14	0.93	0.81	1.09	0.93	0.93	0.86	0.93	0.92
Random Forest (Google)	0.86	0.86	1.01	1.08	1.14	0.93	0.81	1.10	0.90	0.83	0.88	0.94	0.93
Panel C: Sparse													
Lasso VAR	0.88	1.00	0.94	1.00	1.00	1.00	1.06	0.95	1.00	1.07	0.88	1.00	1.01
Lasso VAR (Google)	0.87	1.00	0.92	1.00	0.99	1.00	1.00	1.00	0.95	0.97	0.88	1.02	1.01
Lasso OLS	0.85	0.86	0.99	1.08	1.16	0.92	0.82	1.10	0.95	0.94	0.86	0.99	0.92
Lasso OLS (Google)	0.86	0.86	1.02	1.08	1.11	0.94	0.80	1.10	0.84	0.73	0.86	0.96	0.92
Random Forest VAR	0.93	0.99	0.99	0.92	0.94	1.03	1.18	0.88	0.97	0.82	1.46	0.95	1.04
Random Forest VAR (Google)	0.94	0.88	0.93	0.92	0.96	1.03	1.14	0.91	0.95	0.78	1.46	1.01	1.02
Random Forest OLS	0.86	0.87	1.02	1.07	1.16	0.91	0.83	1.04	0.95	0.90	0.74	**0.85**	0.91
Random Forest OLS (Google)	0.86	0.87	1.02	1.07	1.15	0.93	0.85	1.06	0.86	0.77	0.74	0.96	0.88

In contrast to the quieter prediction period, the best performing models are almost exclusively dense models this time. The accuracy improvements over the AR benchmark are also much larger, ranging from 8% to 54%. The unanimous results in favour of dense models demonstrate that the use of a wide variety of predictors offers great gains in forecast accuracy during the Covid-19 prediction period. The inclusion of Google data also increases forecast accuracy for these models, which further favours the use of many indicators.

The best performer for the majority of subcategories is the Fat VAR. This surprising result suggests that many predictors included in the regression have individual explanatory power over consumption spending. For instance, initial claims for unemployment, the three measures of consumer sentiment from the University of Michigan, the OECD consumer confidence index, M2, and lagged consumption all Granger-cause goods consumption even after controlling for all the other variables.

Forecast accuracy improvements by the PIH/LCH VAR (14%), the Random Forest VAR (7%), and Lasso VAR (8%), also suggest that a variety of predictors seem to have individual importance in forecasting goods consumption. These models select, the S&P500, the high yield credit spread, the FED funds rate, the consumer sentiment index and Google category 89 "Vehicle Parts and Accessories" as important predictors.

Another strong performer is the principal component OLS regression which outperforms for services (18%), durable goods (23%) and vehicles (28%). This suggests that the factors driving the co-movements in our predictor series play an important role in forecasting durable goods. As mentioned in the methodology section, these factors can be loosely associated with financial frictions. Moreover, the outperformance is particularly striking for vehicles and furniture, two types of goods that can usually be financed by credit.

The Lasso OLS and Random Forest OLS also outperform for vehicles (22% and 23%) and furniture spending (14% and 16%). Both these algorithms select mainly credit and sentiment indicators as their best predictors. This result also suggests an important role played by financial variables. Finally, Google category 473 "vehicle shopping", is an important predictor for vehicles spending for both models.

4.1.4 Pandemic Year - Three-Month Forecast

Table 4.4 displays the set of three-months-ahead forecast results for the sub-components of consumption expenditures during the pandemic. Since it is a three-month forecast in a noisy period, we chose to start the prediction period at the third month of the year. Signs of a Covid-19-related economic recession started showing in the economy in February and the impact of the pandemic on consumer spending started showing in the months that followed. Starting the forecasts earlier would be using quiet observations to forecast very noisy observations. Thus we chose to start the prediction period in March for this forecasting task.

A key observation is that the baseline forecast errors are now much less than for the one-month-ahead forecast with reductions in the AR RMSPEs ranging between 32 to 72% depending on the subcategory. This is due to our direct forecasting methodology: taking the three-month average smooths out the economic shock.

Table 4.4: Relative RMSPE for the Pandemic Period - Three-Month Forecast

	Total	**Services**	**Goods**	**NDur**	**Food**	**Cloth**	**Gas**	**Other**	**Dur**	**Vehic**	**Furn**	**Rec.**	**Other**
AR Baseline (RMSE)	2.52	1.71	3.03	1.91	2.01	7.94	7.16	1.22	5.38	5.97	5.18	4.4	8.48
Panel A: Theory													
Random Walk	0.77	0.97	0.92	0.96	0.87	1.03	0.53	1.07	0.93	1.00	0.83	0.93	0.92
Random Walk Drift	0.76	0.96	**0.89**	0.94	0.87	1.02	0.53	1.00	**0.89**	**0.98**	**0.80**	**0.85**	**0.89**
Excess Sensitivity VAR	1.03	1.06	0.99	1.00	1.03	0.99	1.04	1.02	0.99	1.06	1.00	0.97	1.03
PIH / LCH VAR	1.19	1.19	1.02	1.00	1.04	**0.99**	0.97	1.07	1.01	1.09	0.99	0.97	1.08
Panel B: Dense													
Fat VAR	1.51	1.89	1.38	1.48	1.11	1.47	1.01	1.83	1.36	1.46	1.39	1.50	1.16
Fat VAR (Google)	1.63	1.67	1.38	1.40	1.00	1.72	1.10	1.57	1.51	1.72	1.52	1.59	1.18
Fat OLS	0.92	1.25	0.99	0.95	1.08	1.28	0.71	0.87	1.10	1.19	1.03	1.32	0.97
Fat OLS (Google)	0.87	1.20	0.98	**0.83**	1.16	1.38	0.65	**0.81**	1.18	1.29	1.11	1.42	0.94
PCA VAR	1.98	1.85	1.71	1.96	0.96	1.74	2.19	1.85	1.73	2.10	1.48	1.54	1.47
PCA VAR (Google)	1.93	1.87	1.65	1.91	1.01	1.64	1.94	1.90	1.69	2.02	1.45	1.56	1.40
PCA OLS	0.87	0.99	1.12	1.04	**0.77**	1.08	0.53	1.32	1.20	1.17	1.17	1.47	0.97
PCA OLS (Google)	0.86	0.99	1.09	1.01	0.84	1.09	0.52	1.30	1.20	1.20	1.13	1.43	0.97
Random Forest	0.93	1.24	0.98	1.00	0.85	1.16	0.63	1.06	1.00	1.09	0.90	0.92	1.05
Random Forest (Google)	0.94	1.29	0.97	0.97	0.89	1.18	0.64	1.01	1.01	1.11	0.89	0.92	1.06
Panel C: Sparse													
Lasso VAR	1.93	1.87	1.65	1.91	1.01	1.00	1.00	1.90	1.69	1.00	1.45	1.56	1.40
Lasso VAR (Google)	1.00	1.00	1.05	1.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Lasso OLS	0.81	**0.92**	0.95	1.03	0.81	1.03	0.53	1.01	0.94	1.00	0.94	0.93	0.94
Lasso OLS (Google)	0.77	0.97	0.92	0.93	0.87	1.03	0.53	1.07	0.93	1.00	0.83	0.93	0.92
Random Forest VAR	3.29	3.88	1.86	2.59	1.16	2.73	2.61	2.62	2.02	1.93	1.65	1.64	2.78
Random Forest VAR (Google)	3.11	4.33	1.86	2.20	1.12	2.65	2.55	2.21	2.16	2.02	1.77	1.79	2.48
Random Forest OLS	0.81	1.05	1.15	1.07	0.95	1.04	0.54	1.50	1.29	1.29	1.14	1.71	0.94
Random Forest OLS (Google)	**0.75**	0.95	1.12	0.99	0.83	1.00	**0.51**	1.10	1.28	1.28	1.14	1.71	0.90

The forecasting accuracy of our models also varies largely depending on the target subcategory. For instance, forecast accuracy improvements for clothing (1%) and vehicles (2%) are relatively insignificant compared to improvements in total (25%) and food (23%) spending. These first two subcategories have been particularly hard to forecast as demonstrated by the AR RMSPE (7.94% and 5.97% respectively).

For this forecasting task, our panel of theory-based models performs well due to the performance of the random walk with drift (RWD) model. The RWD namely performs well for total (24%) and goods consumption (11%). It also dominates all other models for the subcategories of durable goods with improvements of up to 20%. Thus, the best three-months-ahead forecast of durable goods is its historical average.

Dense models continue to be the best forecasting models for nondurable goods (17%), food (23%), and other nondurables (19%). These subcategories are among the least variable over the prediction period and also offer some of the largest relative forecast improvements. However, dense models are no longer the best at forecasting the relatively variable clothing and gas subcategories where all of our models struggle to rival the RWD model.

Outside of nondurable goods, only a few models manage to offer small improvements over the RWD. The random forest OLS offers the biggest improvement in predicting total consumption (25%) by selecting credit spreads, initial claims for unemployment, stock market prices and the federal funds rate. This once again suggests an important role played by financial variables. The LASSO OLS is best at predicting services consumption (8%). It selects only one indicator: the monetary base.

In contrast to the pandemic one-month-ahead forecast, the random forest method outperforms its sparse linear version for most of the subcategories of durable goods. This may be explained by the smoothing effect of using a three-month average of consumption spending. Random forests tend to perform better when the fluctuations in the target variable are smaller.

4.2 Forecast Results - Predictor Sets and Specifications

This section investigates the importance of using a large number of predictors and the importance of using the timeliest observations of such predictors. First, we compare the performance of the best models for each set of predictors used. In this way, we assess the marginal importance of our economic indicators and Google search data for forecasting. Second, we compare the performance of the best performing OLS models relative to the best performing VAR models. In doing so, we observe the impact of current and lagged predictors on forecast accuracy. We also provide the performance of random walk models as a naive benchmark against which to evaluate both specifications.

Results are presented for each category of consumption, prediction period and forecast horizon. Forecast accuracy is evaluated using the root mean squared prediction error (RMSPE). The RMSPE roughly measures the average prediction error in the month-over-month change in the target variable. Results are not presented relative to a benchmark to better display the difference in the size of forecasting errors between categories of goods and forecasting tasks.

4.2.1 Predictor Sets

Table 4.5 compares the best models for each predictor sets used. Results from section 4.1 show that theory-based models are hard to beat during the quiet period. Congruently, it is hard to improve upon theory-based predictors for this period. However, the larger predictor sets manage to rival theory-based predictors in terms of accuracy for the headline measures and the subcategories of nondurable goods at both forecast horizons during the quiet period. For the one-month-ahead forecasts, Google variables selected by the Lasso and the Random Forest algorithm contribute positively to the forecasts. At a three-month horizon, the random forest ensemble method performs on par with the other best model.

For the pandemic prediction period at a one-month-ahead forecast, results show that dense modelling techniques outperform sparse and theory-based models. Consistent with

Table 4.5: RMSPE of the Best Models by Predictor Set

	Total	**Services**	**Goods**	**NDur**	**Food**	**Cloth**	**Gas**	**Other**	**Dur**	**Vehic**	**Furn**	**Rec.**	**Other**
Quiet, H = 1													
Theory	**0.25**	**0.18**	**0.62**	**0.50**	0.58	1.12	1.19	**0.61**	**0.96**	**1.73**	1.06	**1.47**	**1.07**
Large	**0.25**	**0.18**	0.64	0.51	0.58	1.16	1.17	**0.61**	1.02	1.93	**1.05**	1.48	1.12
Large + Google	**0.25**	**0.18**	0.66	0.52	**0.57**	**1.11**	**1.16**	**0.61**	0.98	1.88	**1.05**	1.50	1.14
Quiet, H = 3													
Theory	**0.32**	**0.42**	**0.27**	0.30	**0.88**	1.44	0.82	**0.37**	**0.75**	**1.73**	**0.59**	0.61	**1.02**
Large	**0.32**	**0.42**	**0.27**	0.31	0.89	**1.42**	**0.80**	0.38	0.78	1.93	0.63	0.62	1.04
Large + Google	0.33	**0.42**	**0.27**	**0.29**	**0.88**	1.43	0.82	**0.37**	0.78	1.88	0.60	**0.60**	1.04
Covid, H = 1													
Theory	5.17	5.23	4.82	4.11	6.35	17.19	11.00	2.65	7.99	11.36	7.87	**5.46**	14.59
Large	3.99	**4.99**	3.86	2.65	**6.06**	15.92	10.11	2.23	7.13	9.05	6.34	**5.46**	11.80
Large + Google	**3.68**	**4.99**	**3.53**	**2.16**	6.14	**15.30**	**9.05**	**1.98**	**7.06**	**8.86**	**5.98**	5.61	**11.30**
Covid, H = 3													
Theory	1.91	1.64	**2.69**	1.79	1.74	**7.83**	3.76	1.22	**4.81**	**5.85**	**4.12**	**3.72**	**7.58**
Large	2.03	**1.57**	2.87	1.82	**1.54**	7.94	3.76	1.06	5.04	5.96	4.64	4.04	7.94
Large + Google	**1.88**	1.63	2.79	**1.58**	1.67	7.93	**3.65**	**0.99**	5.03	5.96	4.30	4.03	7.64

this result, our large predictor set including Google data is preferred for the majority of subcategories. However, most of the improvement from using our largest predictor set comes from the use of our economic indicators. For instance, the addition of Google data improves the forecasts RMSE for goods consumption by 0.33 and improves the forecast of total consumption by 0.31. Although non-negligible, this improvement is slim compared to the marginal improvement from using the large predictor set of 1.18 and 0.96 respectively for the same categories.

For the three-month-ahead forecast during the pandemic, the largest predictor set offers some improvement in total (0.03) and subcategories of nondurable goods but the improvements are smaller than at a one-month horizon. Moreover, the random walk with drift is the best model for durable goods. Thus, durables are best predicted without any of our predictors.

Finally, an interesting observation is that Google searches fail to improve the forecast accuracy for services in each prediction period and forecast horizon. This suggests that Google searches improve the forecast of total consumption by enabling better predictions of goods consumption. In particular, forecasts of nondurable goods, gas and other nondurables tend to benefit from the inclusion of Google data.

4.2.2 Specifications

Table 4.6 compares the best VAR and OLS models for each subcategory of consumption. VAR models based on theory are the top performers for the one-month horizon during the quiet period. Also, models with OLS specifications fail to improve over the random walk model for all categories of spending except goods (0.02%). This shows that current information offers close to no predictive power during quiet times.

These results indicate that when uncertainty is low, spending appears mainly determined by its past value and grows following income or wealth growth. In other words, there is very little gain from having a wider variety of current economic indicators.

Table 4.6: RMSPE of the Best Models by specification

	Total	Services	Goods	Ndur	Food	Cloth	Gas	Other	Dur	Vehic	Furn	Rec	Other
Quiet, H = 1													
VAR	**0.25**	**0.18**	**0.62**	**0.50**	**0.57**	**1.11**	**1.16**	**0.61**	**0.96**	**1.73**	1.07	**1.47**	**1.07**
OLS	0.27	0.19	0.64	0.59	0.67	1.28	1.53	0.65	1.04	1.93	1.11	1.64	1.24
Random Walk	0.26	**0.18**	0.66	0.58	0.67	1.27	1.53	0.63	1.00	1.88	**1.06**	1.54	1.19
Quiet, H = 3													
VAR	0.33	**0.42**	0.30	0.31	0.89	**1.42**	0.85	0.40	0.79	**1.73**	0.62	0.74	**1.02**
OLS	**0.32**	**0.42**	0.31	0.32	**0.88**	1.43	0.83	0.40	0.80	1.93	0.68	0.69	1.04
Random Walk	**0.32**	**0.42**	**0.27**	**0.30**	**0.88**	1.44	**0.82**	**0.37**	**0.75**	1.88	**0.59**	**0.61**	1.05
Covid, H=1													
VAR	**3.68**	5.14	**3.53**	**2.16**	**6.06**	**15.30**	**9.05**	**1.98**	7.99	9.55	**5.98**	**5.46**	**11.30**
OLS	4.68	**4.99**	5.26	4.99	7.32	16.13	11.01	3.32	**7.06**	**8.86**	6.24	**5.46**	13.44
Random Walk	5.17	5.23	5.74	5.05	7.46	17.19	11.00	3.50	9.03	12.07	8.11	6.27	14.59
Covid, H = 3													
VAR	2.52	1.71	3.01	1.91	1.92	**7.83**	6.98	1.22	5.31	5.97	5.15	4.28	8.48
OLS	**1.88**	**1.57**	2.79	**1.58**	**1.54**	7.93	**3.65**	**0.99**	5.03	5.96	4.30	4.08	7.64
Random Walk	1.91	1.64	**2.69**	1.79	1.74	8.09	3.76	1.22	**4.81**	**5.85**	**4.12**	**3.72**	**7.58**

During the quiet period at a three-month horizon, neither of the VAR or OLS specification dominates the other. The performance of the random walk shows that a naive method which uses no predictors offers the best prediction accuracy. Thus, the information content of both past and current observations is quite limited.

For the one-month-ahead forecast during the Covid-19 prediction period, the relative performance of the VAR against the OLS specification is highly variable. The best performing model is the Fat VAR which outperforms its OLS counterpart for a majority of subcategories. A possible explanation for such strong performance is in its limited number of lags. For instance, the FAT VAR with Google data includes only one lag of each predictor. In an episode like the pandemic, where variables decline and recover quickly, shorter lag lengths might allow for faster adjustments. However, the principal component regression and sparse models tend to offer better forecast accuracy with their OLS specification for most subcategories.

The variable results between the OLS and VAR specification for this forecasting task may simply indicate that the best model is one that uses a same-month measure and a one-month-lagged measure of all predictors. This preference for two months of observation is likely explained by the fact that most of the decline in consumption happened over a period of two months, from mid-March to mid-April. It is therefore unclear whether this advantage would extend to other, more typical recessionary periods.

Finally, once we extend the forecast to a three-month horizon for the pandemic period, most machine learning models offer better forecasting performance using the OLS rather than the VAR specification. Moreover, VAR specifications underperform the random walk models for all categories except clothing. Past observations offer little information content for three-month-ahead predictions. But current observations increase forecast accuracy for all subcategories of nondurable goods as well as total and services consumption. However, this result does not extend to durable goods.

4.3 Discussion

4.3.1 Results interpretation

The results indicate that both machine learning and Google search data offer clear gains in forecast accuracy during the Covid-19 pandemic. In rarer occasions, significant gains are also possible during the quieter prediction period. Despite these positive results, the best forecasting models vary depending on the forecasting task. Thus there is no clear winner and model uncertainty is pervasive. Nonetheless, comparing many models with different sets of attributes and predictors offers valuable insight in understanding consumer spending under different economic environments.

In our quieter period at a one-month forecast horizon, improving upon an AR(p) benchmark is difficult; vector autoregressions backed by theory offer the most consistent improvements. Other sparse machine learning models, such as the Lasso VAR and the random forest VAR, also offer improvements for some categories of nondurable goods. These findings namely suggest that, past values of consumption are the key determinants of household spending. That is, spending is determined by simple rules based on habits, income and wealth. This result justifies the bias from theory in favour of using only a few predictors during quiet periods.

At a three-month horizon during the quiet period, the random walk with drift is the most consistent model for all subcategories of spending. As demonstrated by Hall (1978), the rational expectation permanent income hypothesis states that only revisions in permanent income cause changes in consumption. But future revisions depend on future changes in the economic environment and are therefore unpredictable. The longer the forecast horizon is extended, the more of such revisions are then omitted when forecasts are generated. Thus, the best forecast quickly becomes the historical average as we extend the forecast horizon.

At a one-month horizon during the pandemic, the principal component OLS regression is the strongest performer for durable goods. The outperformance from our factor model suggests that durables are mainly driven by the co-movements between our predictor series. Such co-movements can all be loosely associated with financial frictions. Strong results from the Lasso OLS and Random Forest OLS regressions also suggest that financial variables play an important role in shaping spending decisions for durable goods. These findings are consistent with a reduction or a delay in nonessential consumption due to either liquidity constraints or precautionary saving motives.

The fat VAR model surprisingly outperforms every other model in one-month-ahead forecasting of nondurable goods during the pandemic. As this result runs contrary to the bias in favour of using a limited number of predictors in a linear regression, it is unclear if this result is due to a superior model choice or simply due to sheer luck. On the one hand, perhaps the curse of dimensionality is a blessing and many of the individual predictors included in the regression have unique explanatory power over the target variables. Indicators such as initial claims for unemployment, the three measures of consumer sentiment from the University of Michigan, the OECD consumer confidence index, and the M2 money supply all have statistically significant impact over goods consumption. The regression results are presented in Appendix C. On the other hand, perhaps the added noise on which the model is fit, due to the lower effective sample size, is useful in a noisy period such as the pandemic. However, a large variety of different predictors are also used in other good performing models such as the PIH/LCH VAR and the Random Forest VAR. This observation supports the former explanation.

Extending the pandemic forecast horizon to 3 months brings the consumption of durable goods and its subcategories close to a random walk with drift. However, dense forecasting methods using their OLS specifications are able to offer improvements in forecasting less volatile categories of consumption such as nondurable goods, food and other nondurables. The preference for current over past observations show that when the economic situation declines and recovers quickly, the information content of economic indicators also expires quickly. For instance, the RWD outperformance for durable goods shows that current information offers close to no useful information at this forecast horizon as is consistent with the rational expectation permanent income hypothesis.

4.3.2 Implications

The variability of our results between both prediction periods shows that consumer expectations are not static (Lucas, 1976) and highlights the role of the current economic environment in shaping households' expectations. As proposed by Chauvet and Potter (2013), our results show that there are substantial gains from using separate forecasting models for normal times and our period of abrupt changes. By integrating large predictor sets including timely information, our panel of dense models offer accuracy improvement during the pandemic over the best performing models in normal times for at least two reasons.

First, households are likely to base their spending decisions on a more complex set of information when uncertainty is high, and use simple rules when uncertainty is low. As in Goulet Coulombe et al. (2021b) our analysis shows that there are substantial gains of using a large number of predictors during the Covid-19 pandemic. This result further corroborates findings reported by Siliverstovs and Wochner (2019) and Kotchoni et al. (2019) which demonstrate benefits from using a large set of predictors in recessionary episodes. Thus, imposing sparsity on the number of predictors, as is best in normal times, would omit useful information for forecasting.

This result extends to the inclusion of Google data which also improves forecasts during the pandemic. Google searches may be useful predictors for many reasons: they may reflect consumer confidence, may reveal spending intention in subcategories of durable goods, or may directly indicate spending on items which are purchasable online. Thus, as a bigger percentage of consumption is displaced online, Google searches could become increasingly relevant. Data collected by Hall et al. (2020) namely show that online sales volumes increased by up to 25% over the previous year during the pandemic. This may explain why Google data offer better forecast improvements during the pandemic than during the quiet period. For now, however, our results show that most of the improvements in forecast accuracy from using a large predictor set still come from traditional economic indicators which echo the conclusion from Chen et al. (2019).

Second, in periods of high uncertainty, households are more likely to closely track economic developments than in quiet periods. During the pandemic, our OLS vs VAR comparison indicates that households determine their consumption by putting relatively more weight on current economic indicators rather than on their past values. Specifically, current developments in credit markets appear to play an important role in shaping spending decisions. Our results support evidence collected by Baker et al. (2020) which show that households with low levels of liquidity had the largest decline in spending. Chetty et al. (2020) in turn show that these same households drastically increased their spending after receiving stimulus payments. These findings are all consistent with the theory of excess sensitivity where households faced with a liquidity constraint have a greater Keynesian marginal propensity to consume out of current income. The ability of our large set of timely indicators to closely track current developments in financial markets allows forecast improvements during the pandemic.

4.3.3 Limitations

The main limitation of our analysis is due to the atypical nature and swiftness of the economic shock from the pandemic. Spending patterns have differed from those of more "traditional" recessions. This might have biased results in favour of noisier models such as the Fat VAR or in favour of naive status quo models such as the random walk with drift. Thus, it is unclear whether all of our results apply to other recessions.

Furthermore, some consumption measures are likely affected due to lockdown measures and, given the lack of precedent, we don't expect models fit on historical data to have integrated such an effect. One way of increasing forecast accuracy is through a form of intercept correction based on forecast errors from the great recession as proposed by Foroni et al. (2021). However, such improvements would likely increase forecast performance for all models in similar ways. Given that model comparison is our main objective, considering relative RMSPE is a way to mitigate the effects of lockdown measures.

Similarly, certain states were disproportionately impacted by the Covid-19 pandemic and lockdown severity varied between states. Such heterogeneity is lost when considering country aggregates. Thus a limitation of our results is that they do not necessarily generalize to every state. However, there are a few practical reasons that make a more granular approach impractical for the Covid-19 episode. First, detailed consumption expenditures by states are available only at an annual basis. Second, measures of consumer confidence are not available on a state-by-state basis. Third, the quality of Google Trends data also varies widely from state to state, with some state having no recorded data for certain search categories. Despite such limitations, future research focusing on a longer time horizon might be able to uncover unique spending patterns for each region with a similar methodology.

Another limit of our analysis is in the simplicity of the models we use. We opt to use models in their simplest form in order to ease the interpretation of the results and comparison between models. However, evidence from Giannone et al. (2021) indicates that the best forecast is generated by a weighted average of many models while Li and Chen (2014) and Kim and Swanson (2018) find that forecast combination techniques outperform other forecasts. It is possible that these more complex techniques might have offered better forecast accuracy and different results.

For instance, we note that the random forest ensemble method, while also being a dense forecasting method, performs poorly during the pandemic. This is most likely due to the mechanics of the algorithm. The algorithm makes forecasts based on the mean observation in each region, it cannot predict out-of-sample values that are larger than its in-sample maximum. Thus, its performance is affected when fluctuations in the target variable are much larger than their average, such as in recessions.

One way to allow the algorithm to extrapolate is to use a linear part within each leaf as proposed by Goulet Coulombe (2020). So-called macroeconomic random forests (MRFs) improve upon random forests by being able to predict large bounce backs (Goulet Coulombe et al., 2021a). Furthermore, Goulet Coulombe et al. (2021b) show that non-linear models are true game changers in periods of high macroeconomic uncertainty and financial stress. Thus, the MRF is a model combination technique which might have potentially offered better forecasts during the pandemic.

Another avenue that wasn't explored is to use a simple OLS regression with lagged values of a few predictors. As discussed for the one-month-ahead period during the pandemic, this combination of both OLS and VAR specifications might have yielded better forecasting results. However, keeping both specifications separated allows us to better understand the relative roles played by current and past observations of our indicators on forecast accuracy.

Finally, since we limit our predictors to a set of variables that are likely known by households, we note that the number of indicators used in this research is inferior to the number typically used in out-of-sample forecasting horse races. It is possible that some omitted predictors might have contributed additional predictive power over the dependent variable. However, using a sparser predictor set facilitates results interpretation.

Similarly, it is also likely that other alternative data sources such as credit and debit card purchases may lead to better predictions. However, as pointed by Chen et al. (2019), the cost, quality and availability of private data change over time which makes their use unstable. For instance, credit and debit card purchases are available only from private sources, and in the rare cases where they are made publicly available, have only a very limited time series. We thus choose to focus exclusively Google Trends since, to the best of our knowledge, they are the most accessible alternative data source with a sufficient number of observations for time series analysis. Future research focusing exclusively on high frequency alternative data might be able to use a combination of Google Trends and transaction data to uncover weekly patterns in consumption.

4.3.4 Recommendations

In light of our research, we highlight that machine learning models are important forecasting tools during periods of high uncertainty such as the first year of the Covid-19 pandemic. During this period, our results show that the best forecast accuracy is obtainable by using dense models with a large predictor set including alternative data. Our results also suggest the use of the timeliest observations available.

Future research is needed to establish the best performing empirical model during Covid-19. Apart from large scale structural models, more complex forecast combination techniques, such as the MRF, and intercept correction techniques have the potential to offer further forecast improvements over the methods presented in this paper. However, given the uniqueness of the pandemic period in both its speed and severity, large forecast errors are likely to remain endemic for most categories of consumption.

Finally, despite the "black box" nature of machine learning techniques, economic intuition is obtainable by understanding the strengths and weaknesses of each algorithm. Meaningful insights can then be gained by considering a large panel of models. One such insight is the importance played by financial variables in shaping durable goods spending decisions during the pandemic. This evidence is consistent with excess sensitivity due to financial frictions and highlights the importance of governments supplying liquidity in credit markets to support consumption, and in turn production, in times of crisis.

5 Conclusion

This research aimed to determine how machine learning and alternative data help in forecasting and consumer spending both before and during the Covid-19 pandemic. Based on the results of a pseudo-out-of-sample forecasting horse race of theory based VAR as well as principal component regressions, LASSO regressions and random forest regressions, it can be concluded that machine learning modelling techniques offer both greater forecast accuracy and valuable insights in predicting household personal expenditures during the pandemic. However, these improvements are mostly limited to a one-month-ahead forecast.

First, during the pandemic, purchases of durable goods are primarily driven by co-movements in our predictor series. Such co-movements are loosely associated with financial frictions. In accordance with macroeconomic theory, this provides evidence that financial frictions accelerated the decline in durable goods spending; relaxation of the constraints and motives helped the subsequent recovery.

Next, many individual variables were important in forecasting nondurable goods purchase during the pandemic, this warns against the a priori bias in favour of sparse models. Sparse models omitted important information such as measures of consumer confidence and measures of labour market health.

These favourable results in favour of dense models help us conclude that using a large set of timely economic indicators and including Google trends data was instrumental in increasing forecast accuracy during the pandemic.

Two limitations of our analysis are in the number of models considered and the number of predictors included. Conclusions from the literature namely point in the direction of using a combination of models to obtain the greatest forecast accuracy. Moreover, some private data sources including credit card transaction data might have offered greater explanatory power, but they do not offer the same ease of access and stability over time as Google search data.

Our analysis demonstrates that model uncertainty is pervasive and no single model systematically outperforms other models for all prediction periods and forecast horizons. Practitioners should thus consider a wide variety of models in conducting their experiments. Models considered best during quiet periods should not necessarily be considered best for recessionary episodes. In periods of high uncertainty specifically, machine learning techniques using large predictor sets supplemented by a careful selection of alternative data may offer improvements in forecast accuracy over simpler models. Apart from improving forecast accuracy, observing the performance of many models also offers valuable insights in understanding how consumption variables behave in different economic environments.

This thesis contributes to the pseudo out-of-sample horse race forecasting literature by validating it as a valuable tool in forecasting and understanding consumer spending in a period of high uncertainty. First, our results challenge the theoretical bias in favour of using only a few predictors; our models using a large predictor set offer improvements in forecast accuracy during the pandemic. Second, our results corroborate the common finding that machine learning models are useful, but this result is mostly limited to recessionary episodes. Finally, our results challenge the common "black box" criticism of machine learning models by showing that valuable insights can be gained from a careful analysis. For instance, our analysis provides some evidence of the importance of financial variables in determining durable goods spending during the pandemic.

Bibliography

Baker, S. R., Farrokhnia, R., Meyer, S., Pagel, M., and Yannelis, C. (2020). How does household spending respond to an epidemic? consumption during the 2020 covid-19 pandemic. *The Review of Asset Pricing Studies*, 10(4):834–862.

Bernanke, B. (1985). Adjustment costs, durables, and aggregate consumption. *Journal of Monetary Economics*, 15(1):41–68.

Breiman, L. (2001). Random forests. *Machine Learning*, 45:5–32.

Campbell, J. and Deaton, A. S. (1989). Why is consumption so smooth? *The Review of Economic Studies*, 56(3):357–373.

Carbajo, D. (2021). Consumer spending and the covid-19 pandemic. *Page One Economics Federal Reserve Bank of St. Louis*.

Chauvet, M. and Potter, S. (2013). *Handbook of Economic Forecasting*. Elsevier.

Chen, J. C., Dunn, A., Hood, K., Driessen, A., and Batch, A. (2019). *Off to the races: a comparison of machine learning and alternative data for predicting economic indicators.* University of Chicago Press.

Chetty, R., Friedman, J. N., Hendren, N., Stepner, M., , and the Opportunity Insights Team (2020). The economic impacts of covid-19: Evidence from a new public database built using private sector data. *NBER Working Papers*, 27431.

Coibion, O., Gorodnichenko, Y., and Weber, M. (2020). The cost of the covid-19 crisis: Lockdowns, macroeconomic expectations, and consumer spending. *NBER Working Papers*, 27141.

Cox, N., Ganong, P., Noel, P., Vavra, J., Wong, A., Farrell, D., Greig, F., and Deadman, E. (2020). Initial impacts of the pandemic on consumer behavior: Evidence from linked income, spending, and savings data. *Brookings Papers on Economic Activity*, v2020(2):35–82.

Deaton, A. S. (1987). *Life-cycle models of consumption: is the evidence consistent with the theory?* Cambridge University Press.

Dunn, A., Hood, K., and Driessen, A. (2020). Measuring the effects of the covid-19 pandemic on consumer spending using card transaction data. *BEA Working Paper Series*, WP2020(5).

Flavin, M. A. (1981). The adjustment of consumption to changing expectations about future income. *Journal of Political Economy*, 89(5):974–1009.

Flavin, M. A. (1993). The excess smoothness of consumption: Identification and interpretation. *The Review of Economic Studies*, 60(3):651–666.

Foroni, C., Marcellino, M., and Stevanovic, D. (2021). Forecasting the covid-19 recession and recovery: Lessons from the financial crisis. *International Journal of Forecasting*.

Friedman, M. (1957). *A Theory of the Consumption Function*. Princeton University Press.

Giannone, D., Lenza, M., and Primiceri, G. E. (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica*, 89(5):2409–2437.

Goulet Coulombe, P. (2020). The macroeconomy as a random forest. Available at SSRN: https://ssrn.com/abstract=3633110 or http://dx.doi.org/10.2139/ssrn.3633110.

Goulet Coulombe, P., Leroux, M., Stevanovic, D., and Surprenant, S. (2021a). How is machine learning useful for macroeconomic forecasting? *Working Paper*.

Goulet Coulombe, P., Marcellino, M., and Stevanovic, D. (2021b). Can machine learning catch the covid-19 recession? *National Institute Economic Review*, 256:71–109.

Hall, M. C., Prayag, G., Fieger, P., and Dyason, D. (2020). Beyond panic buying: consumption displacement and covid-19. *Journal of service management*, 32(1):113–128.

Hall, R. E. (1978). Stochastic implications of the life cycle-permanent income hypothesis: Theory and evidence. *Journal of Political Economy*, 86(6):971–987.

Keynes, J. M. (1936). *The General Theory of Employment, Interest, and Money*. Palgrave Macmillan.

Kim, H. H. and Swanson, N. R. (2018). Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods. *International Journal of Forecasting*, 34(2):339–354.

Kotchoni, R., Leroux, M., and Stevanovic, D. (2019). Macroeconomic forecast accuracy in a data-rich environment. *Journal of Applied Econometrics*, 34(7):1050–1072.

Li, J. and Chen, W. (2014). Forecasting macroeconomic time series: Lasso-based approaches and their forecast combinations with dynamic factor models. *International Journal of Forecasting*, 30(4):996–1015.

Lucas, R. J. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, 1(1):19–46.

Madeiros, M., Vasconcelos, G., Veige, A., and Ziberman, E. (2019). Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business and Economic Statistics*, 39(1):1–45.

Modigliani, F. and Brumberg, R. (1954). Utility analysis and the consumption function: an interpretation of cross-section data. *PostKeynesian Economics*, pages 388–436.

Siliverstovs, B. and Wochner, D. (2019). Recessions as breadwinner for forecasters state-dependent evaluation of predictive ability: Evidence from big macroeconomic us data. *KOF Working papers*, 463.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1):1–48.

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

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

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

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, 58(1):267–288.

Appendix A

Table 1: List of Indicators and Data Treatment

Macroeconomic Indicators	Treatment	Source	Release Frequency
Personal Consumption Expenditure	Log first Diff	Bureau of Economic Analysis	Last friday of following month
Personal Income	Log first Diff	Bureau of Economic Analysis	Last friday of following month
Consumer Confidence Index	Log first Diff	OECD	Middle of month
Consumer Sentiment Index	Log first Diff	University of Michigan	Middle of month
Average Weekly Hours (mfg)	Log first Diff	Bureau of Labor Statistics (FRED)	First Friday of following month
Initial Claims	Log first Diff	US Emp. and Train. Adm. (FRED)	Every following Thursday
Building Permits	Log first Diff	Census Bureau (FRED)	12th workday of following month
Fed Funds Rate	First Diff	Federal Reserve (FRED)	End of Month
10 yr	First Diff	Federal Reserve (FRED)	Every following monday
Monetary Base	Log First Diff	Federal Reserve (FRED)	4th Tuesday of following month
M2	Log First Diff	Federal Reserve (FRED)	4th Tuesday of following month
Investment Grade	First Diff	Ice Data Indices (FRED)	Daily
High Yield	First Diff	Ice Data Indices (FRED)	Daily
USD	First Diff	Federal Reserve (FRED)	Daily
WTI	Log First Diff	FRED	Daily
Brent	Log First Diff	FRED	Daily
Gold	Log First Diff	FRED	Daily
Copper	Log First Diff	IMF (FRED)	Monthly
VIX	Log First Diff	CBOE (FRED)	Daily
Google Categories			
(11) Home and Garden	Log First Diff	Google Trends	Daily
(29) Real Estate	Log First Diff	Google Trends	Daily
(48) Construction and Maintenance	Log First Diff	Google Trends	Daily
(60) Jobs	Log First Diff	Google Trends	Daily
(89) Vehicle Parts and Accessories	Log First Diff	Google Trends	Daily
(124) Clothing Accessories	Log First Diff	Google Trends	Daily
(158) Home Improvement	Log First Diff	Google Trends	Daily
(229) TV and Video Equipment	Log First Diff	Google Trends	Daily
(263) Sporting Goods	Log First Diff	Google Trends	Daily
(270) Home Furnishings	Log First Diff	Google Trends	Daily
(273) Motorcycles	Log First Diff	Google Trends	Daily
(361) Audio Equipment	Log First Diff	Google Trends	Daily
(473) Vehicle Shopping	Log First Diff	Google Trends	Daily
(650) Building Materials and Supplies	Log First Diff	Google Trends	Daily
(815) Vehicle Brands	Log First Diff	Google Trends	Daily
(899) Game Systems and Consoles	Log First Diff	Google Trends	Daily
(951) Kitchen and Dining	Log First Diff	Google Trends	Daily

Appendix B

Table 2: List of Models, Predictors Set and Specification

Name	Predictors X_t	Specification	Values of X_{t}
Panel A: Theory			
AR Baseline (RMSE)	Consumption	Autoregression	$L(C_t)$
Random Walk	None	Random Walk	C_{t-1}
Random Walk Drift	None	Random Walk	$\Delta\bar{C}$
Excess Sensitivity VAR	Cons, Income	VAR	$L(C_t, Inc_t)$
PIH / LCH VAR	Cons, Inc, SP500	VAR	$L(C_t, Inc_t, SP_t)$
Panel B: Dense			
Fat VAR	Large	VAR	X_{t-1}, X_{t-2}
Fat VAR (Google)	Large + Google	VAR	X_{t-1}
Fat OLS	Large	OLS	X_t
Fat OLS (Google)	Large + Google	OLS	X_t
PCA VAR	Large	VAR	$L(X_t)$
PCA VAR (Google)	Large + Google	VAR	$L(X_t)$
PCA OLS	Large	OLS	X_t
PCA OLS (Google)	Large + Google	OLS	X_t
Random Forest	Large	Random Forest	X_t
Random Forest (Google)	Large + Google	Random Forest	X_t
Panel C: Sparse			
Lasso VAR	Large	VAR	$L(X_t)$
Lasso VAR (Google)	Large + Google	VAR	$L(X_t)$
Lasso OLS	Large	OLS	X_t
Lasso OLS (Google)	Large + Google	OLS	X_t
Random Forest VAR	Large	VAR	$L(X_t)$
Random Forest VAR (Google)	Large + Google	VAR	$L(X_t)$
Random Forest OLS	Large	OLS	X_t
Random Forest OLS (Google)	Large + Google	OLS	X_t

Table 3: List of Principal Components and Their Association and (Explained Variance)

First Principal Component (81.20%)										
Variable	**HY**	**IG**	**VIX**	**10yr**	**Claims**	**FED**	**BRENT**	**WTI**	**COPPER**	**SP500**
Correlation	0.9970	0.9539	0.5946	-0.5079	0.4506	-0.5539	-0.6273	-0.6137	-0.5891	-0.6701
Eigenvalue	0.9084	0.3346	0.1244	0.1082	0.1016	0.0933	0.0747	0.0736	0.0433	0.0305
Second Principal Component (4.55%)										
Variable	**Claims**	**FED**	**HY**	**WTI**	**VIX**	**BRENT**	**899**	**IG**	**10yr**	**60**
Correlation	0.8422	-0.3559	-0.0647	-0.4005	0.2212	-0.3877	0.4160	0.0899	-0.1201	-0.3102
Eigenvalue	0.8057	0.2545	0.2502	0.2039	0.1965	0.1959	0.1812	0.1338	0.1086	0.0844
Third Principal Component (2.89%)										
Variable	**10yr**	**VIX**	**WTI**	**Claims**	**BRENT**	**FED**	**HY**	**GOLD**	**270**	**UMICHCur**
Correlation	-0.8394	-0.1098	-0.1889	-0.0992	-0.1734	0.1126	-0.0161	0.3265	0.2209	0.1256
Eigenvalue	0.9513	0.1223	0.1205	0.1190	0.1098	0.1009	0.0779	0.0715	0.0524	0.0402
Fourth Principal Component (2.5%)										
Variable	**VIX**	**FED**	**IG**	**899**	**SP500**	**10yr**	**Permits**	**MB**	**263**	**270**
Correlation	-0.7349	-0.3537	0.0926	0.1964	0.4409	0.0759	-0.2045	0.2936	0.2668	0.2614
Eigenvalue	0.8744	0.3387	0.1846	0.1146	0.1140	0.0919	0.0755	0.0704	0.0681	0.0663

Appendix C

Quiet Period - One-Month Forecast

Table 4: Predictor Importance for Select Regressions - QP1M

Food, QP1M - Feature Selection, Random Forest VAR (Google)						
Importance	**Prediction 2015**	**Prediction 2016**	**Prediction 2017**	**Prediction 2018**	**Prediction 2019**	**Prediction 2020**
1	951	951	UMICH	UMICH	UMICH	229
2	UMICH	UMICH	UMICHCur	UMICHCur	UMICHCur	SP500
3	UMICHCur	UMICHCur	UMICHexp	UMICHexp	SP500	VIX
4	UMICHexp	UMICHexp	SP500	AWH	10yr	VIX

Clothing, QP1M - Feature Selection, LASSO VAR (Google)						
Importance	**Prediction 2015**	**Prediction 2016**	**Prediction 2017**	**Prediction 2018**	**Prediction 2019**	**Prediction 2020**
1	815	815	815		263	
2	HY	HY	273		89	
3	Permits	273	HY		M2	
4	M2	PERMITS	PERMITS		HY	

Gasoline, QP1M - Feature Selection, Random Forest VAR (Google)						
Importance	**Prediction 2015**	**Prediction 2016**	**Prediction 2017**	**Prediction 2018**	**Prediction 2019**	**Prediction 2020**
1	WTI	IG	IG	IG	IG	IG
2	48	USD	263	263	263	OECD
3	263	263	OECD	OECD	OECD	UMICH
4	273	UMICH	UMICH	UMICH	UMICH	UMICHCur

Furniture, QP1M - Feature Selection, Random Forest Ensemble						
Importance	**Prediction 2015**	**Prediction 2016**	**Prediction 2017**	**Prediction 2018**	**Prediction 2019**	**Prediction 2020**
1	Permits	Claims	Claims	Claims	Claims	Claims
2	Claims	Permits	Permits	Permits	Permits	Permits
3	IG	IG	IG	IG	SP500	SP500
4	UMichCur	AWH	UMICHexp	AWH	IG	IG

Quiet Period - Three-Month Forecast

Pandemic Period - One-Month Forecast

Summary of Regression Results

Model:	VAR
Method:	OLS
Date:	Wed, 08, Dec, 2021
Time:	20:48:22

No. of Equations:	39.0000	BIC:	-258.683
Nobs:	163.000	HQIC:	-276.271
Log likelihood:	16035.6	FPE:	9.32774e-126
AIC:	-288.291	Det(Omega_mle):	1.78956e-129

Results for equation Goods

	coefficient	std. error	t-stat	prob
const	0.000087	0.001144	0.076	0.939
L1.Goods	**-0.313984**	**0.088152**	**-3.562**	**0.000**
L1.Income	0.054143	0.072149	0.750	0.453
L1.OECD	**0.923859**	**0.373567**	**2.473**	**0.013**
L1.UMICH	**-1.394806**	**0.425006**	**-3.282**	**0.001**
L1.UMICHCur	**0.620837**	**0.189399**	**3.278**	**0.001**
L1.UMICHexp	**0.737767**	**0.235387**	**3.134**	**0.002**
L1.AWH	0.046027	0.167450	0.275	0.783
L1.Permits	0.000000	0.011482	0.000	1.000
L1.FED	0.008228	0.004832	1.703	0.089
L1.MB	-0.007521	0.019147	-0.393	0.694
L1.M2	**0.419890**	**0.197083**	**2.131**	**0.033**
L1.COPPER	0.000169	0.011717	0.014	0.989
L1.Claims	**-0.059328**	**0.016115**	**-3.682**	**0.000**
L1.SP500	0.006628	0.029361	0.226	0.821
L1.10yr	-0.006459	0.003811	-1.695	0.090
L1.IG	0.002861	0.005494	0.521	0.603
L1.HY	-0.000878	0.002090	-0.420	0.674
L1.USD	-0.048915	0.067333	-0.726	0.468
L1.WTI	-0.010865	0.018569	-0.585	0.558
L1.BRENT	0.022616	0.019373	1.167	0.243
L1.GOLD	-0.024315	0.017660	-1.377	0.169
L1.VIX	0.002154	0.005174	0.416	0.677
L1.11	-0.004672	0.045772	-0.102	0.919
L1.29	0.015310	0.035564	0.430	0.667
L1.48	-0.007829	0.042672	-0.183	0.854
L1.60	-0.001880	0.023637	-0.080	0.937
L1.89	0.022964	0.037837	0.607	0.544
L1.124	0.000937	0.016334	0.057	0.954
L1.158	-0.019363	0.038815	-0.499	0.618
L1.229	0.006100	0.022702	0.269	0.788
L1.263	-0.004625	0.020893	-0.221	0.825
L1.270	0.034676	0.028775	1.205	0.228
L1.273	-0.010214	0.017169	-0.595	0.552
L1.361	-0.013027	0.021412	-0.608	0.543
L1.473	0.014660	0.017438	0.841	0.401
L1.650	-0.008664	0.035300	-0.245	0.806
L1.815	0.004118	0.009637	0.427	0.669
L1.899	0.004451	0.007778	0.572	0.567
L1.951	0.037857	0.030706	1.233	0.218

Figure 1: Regression result for Goods Consumption

Table 5: Predictor Importance for Select Regressions - QP1M

Goods, PP1M - Feature Selection, 2021				
Importance	**1st Variable**	**2nd Variable**	**3rd Variable**	**4th Variable**
RF Var Google	High Yield	LagGoods	SP500	UMICH
Lasso VAR (Google)	SP500	FED	High Yield	89

Vehicles, PP1M - Feature Selection, 2021				
Importance	**1st Variable**	**2nd Variable**	**3rd Variable**	**4th Variable**
RF Var Google	473	SP500	815	UMICH
Lasso VAR (Google)	473	HY	SP500	MB

Furniture, PP1M - Feature Selection, 2021				
Importance	**1st Variable**	**2nd Variable**	**3rd Variable**	**4th Variable**
RF Var Google	Claims	Permits	SP500	IG
Lasso VAR (Google)	SP500	Permits	FED	UMICHexp

Pandemic Period - Three-Month Forecast

Table 6: Predictor Importance for Select Regressions - QP3M

Total, PP3M - Feature Selection, 2021				
Importance	**1st Variable**	**2nd Variable**	**3rd Variable**	**4th Variable**
RF Var Google	124	Claims	IG	263

Services, PP3M - Feature Selection, 2021				
Importance	**1st Variable**	**2nd Variable**	**3rd Variable**	**4th Variable**
Lasso VAR (Google)	MB			