A Nowcasting Model for Time Series with Ragged-Edge Data

Model Run Date: March 01, 2022

charles@cmefi.com

Contents

List of Tables

List of Figures

1 Motivation

Nowcasting is the prediction of the present, the near future, and the near past. Nowcasting is important in economics because many important macroeconomic statistics are released with a lengthy delay. For example, the Bureau of Economic Analysis releases quarterly GDP data typically two months after the quarter has already ended - a significant delay for any companies or individuals who need the data for planning and forecast models.

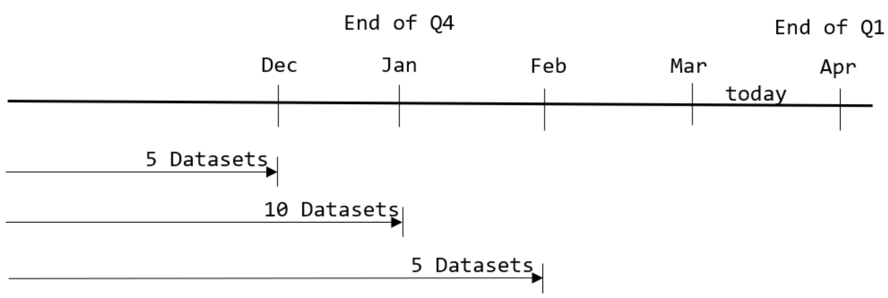
This delay is particularly salient during times of high volatility. During the first few months of the COVID-19 pandemic in the United States, many companies attempted to use high-frequency indicators to attempt nowcasts of the state of the macroeconomy. For example, a major investment bank forecasted Q2 GDP growth of -3% on March 3rd; the estimate was revised down to -14% by March 21st, -25% by March 25th, and -40% by April 10th. Yet many such models were ad-hoc and only able to use a small number of predictive indicators, such as jobless claims or traffic data.

Nowcasting is about deciphering key information about the state of the economy before official data is released. Because of the fundamentally urgent nature of nowcasting, it is important that nowcast models exploit any latest, high-frequency data available. Nowcasts should be able to generate constantly rolling forecasts, updating these numbers in response to any new data releases.

For example, suppose the date is early March, and the variable we want to predict is Q1 GDP. The simplest way to predict Q1 GDP would be to use historical quarterly data from various economic variables. But this data would only go up to Q4 of last year, and would fail to capture the critically important predictive power that could be provided by monthly and daily data released throughout January and February. This lag in publication dates of different data series is known as the **ragged-edge** problem.

Suppose instead, we used monthly data as our predictors of Q1 GDP. Again, we will soon run in to a problem. As an example, imagine that we had imported 20 monthly data series. Suppose 5 of these series ended in December, 10 ended in January, and 5 ended in March.

Figure 1: Example of ragged-edge data



Traditional modeling methods would require us to either throw out variables or throw out months - for example, we could truncate all our data series at January and lose out the information provided by the 5 February data points. Alternatively, we could completely remove the 15 variables with data releasing before March. Both methods result in a serious loss of useful data and are unappealing.

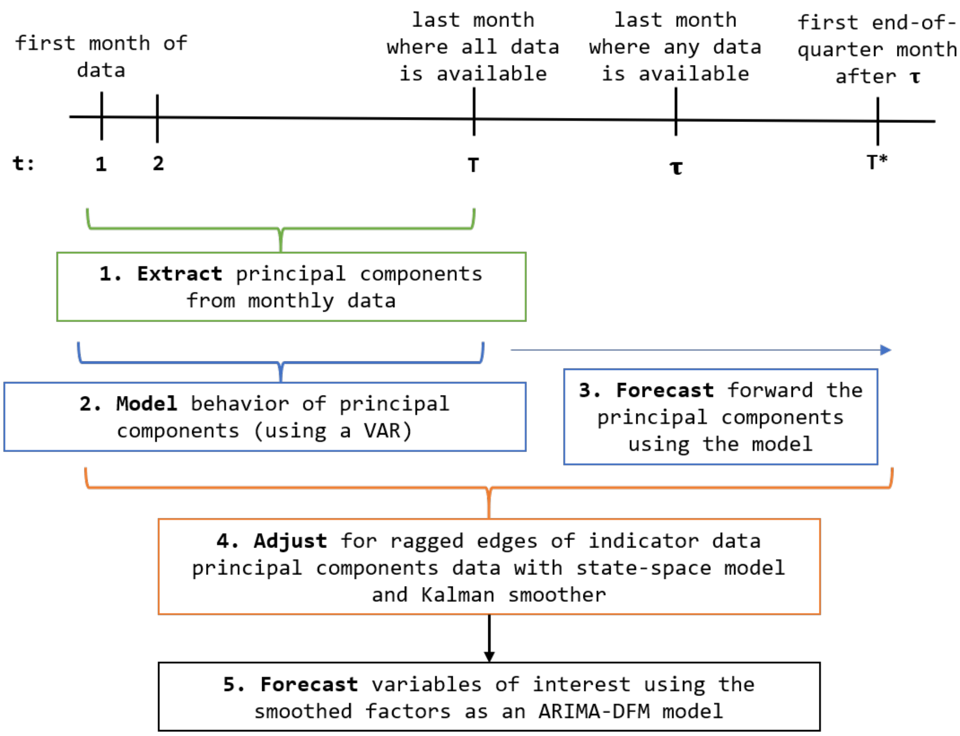
In this paper, we will develop and utilize a methodology that will allow us to use the information from

all variables at **any dates**. This model will give an updated forecast in response to any new data releases. Additionally, the model can be generalized to nowcast any time series variable, not just GDP. The methodology for the model will be described in the next section.

2 Methodology Overview

Our goal will be to take monthly-frequency leading economic predictor variables — industrial production, consumer sentiment, vehicle sales, and so on — and use these to predict our quarterly variables. However, we will need a way to adjust for the ragged edges of the data.

Figure 2: Overview of methodology (more detail in later sections)



To do this, we will use a methodology that relies heavily on principal components. It is well known that most macroeconomic variables are highly correlated with one another (see, for example, Bernanke et al 2005).

term	GDP	VehicleSales	ConsumerSentiment	AdvSales	IndustrialProd	VIX	SP500
GDP							
VehicleSales	0.73						
ConsumerSentiment	0.12	0.06					
AdvSales	0.82	0.78	0.10				
IndustrialProd	0.92	0.73	0.11	0.78			
VIX	-0.36	-0.29	-0.64	-0.40	-0.40		
SP500	0.45	0.50	0.27	0.65	0.48	-0.65	

Table 1: Correlation Between Major Leading Predictors

As a result, we can take a very high-dimensional dataset of indicator variables and extract a few time series, *factors* or *principal components*, using principal components analysis. These factors will be able to contain the majority of information within our larger, high-dimensional dataset. Then we will use a time series model of the factor behavior — a vector autoregression, or VAR — to forecast the factors forward in time.

Finally, we will deal with the ragged edges of the data by casting our model dynamics in state-space form and using a Kalman filtration and smoothing process. This will, in essence, adjust our factor forecasts by whether or not the indicator variables used in the construction of the factors have been released. This procedure is a modified version of the two-step dynamic factor model utilized in Giannone et al (2008) and Doz et al (2011).

Finally, we will aggregate the smoothed and forecasted time series of factors up to a quarterly level. These now-quarterly forecasts of the factors can be used as covariates in an ARIMA model of the quarterly economic variables. The usage of factors as regressors of these economic variables is known as a *dynamic factor model* (DFM). We additionally regularize our selection of regressors using a dynamic factor model.

The next section will go over the data and estimation procedure in detail. While this model can be generalized to nowcasts of other economic variables, we will use the example of nowcasting GDP and components of GDP.

3 Estimation Process

The model is run every weekday; this documentation will use actual data and estimates from the most recent model run (2022-03-01) to illustrate the procedure. The final nowcasted output is located in the results section.

3.1 Data

We begin by importing monthly data of various leading indicators from the St. Louis Federal Reserve Database (FRED). We choose data of at least monthly frequency and with historical data available as of at least 2010. Data are transformed for stationarity as listed below; *dlog* refers to the natural log of the first difference, whereas *base* means that no transformation was necessary for the variable to be stationary.

Variable	Stationary Form
PCE	dlog
Personal Savings Rate	base
Disposable Personal Income	dlog
Bank Commercial & Industrial Loans	dlog
Bank Real Estate Loans	dlog
Case-Shiller National Housing Price Index	dlog
Housing Starts	log
Houses Sold	log
New Private Housing Permits	log
1-Year PCEPI Inflation	base
1-Year CPI Inflation	base
Producer Price Index: All Commodities	dlog
Unemployment Rate	log
Total Nonfarm Job Openings	dlog
Total Unfilled Job Vacancies	pchg
Average Hourly Earnings	dlog
KC Fed Labor Market Index	base
Labor Force Participation Rate	base
Industrial Production index	dlog
Industrial Production: Final Products	dlog
Industrial Production: Consumer Goods	dlog
Industrial Production: Fuels	dlog
Manufacturers New Orders: Durable Goods	pchg
Manufacturers New Orders: Consumer Goods	pchg
Total Construction Spending	dlog
Inventory to Sales Ratio	base
Capacity Utilization	base
Federal Funds Rate	base
30-Year Mortgage Rate	base
Corporate Aaa Bond Yield	base
Corporate Baa Bond Yield	base
Vehicle Sales	dlog
Retail & Food Sales	dlog
Consumer Confidence (OECD)	base
New York Fed: Future Capex Manufacturing Survey	base
Dallas Fed: Current General Business Activity	base
Philadelphia Fed: Future Capital Expenditures Survey	base
S&P 500	dlog
Nasdaq Composite	dlog
VIX	base
Agricultural Commodities (DBA)	dlog
Metal Commodities (DBB)	dlog
Crude Oil Price (West Texas Intermediate)	base
Ex-US Equities (SPDW)	dlog

Table 2: Imported Monthly Data - Leading Economic Variables

We additionally import quarterly data for our variables of interest. While this model can be used for nowcasting other data, here we will use import GDP, its components, as well as several other major macroeconomic variables of interest.

```
## Error in (function (cond) : error in evaluating the argument 'x' in selecting a method
for function 'print': Problem while computing `..1 = freq == "q"`.
## Caused by error:
## ! object 'freq' not found
```

Finally, we import additional monthly data that may be of interest to forecast, but are not leading indicators that we will use in our principal components analysis.

```
## Error in (function (cond) :  error in evaluating the argument 'x' in selecting a method
for function 'print':  Problem while computing `..1 = (is.na(nc_dfm_input) | nc_dfm_input
== 0)
## & freq == "m"`.
## Caused by error:
## !  object 'freq' not found
```

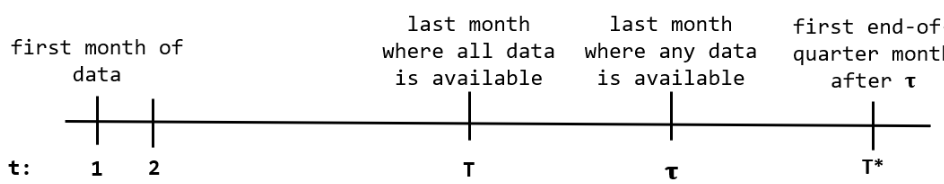
Most datasets have already been deseasonalized if necessary by their original source. We deseasonalize the remaining series by using the U.S. Census Bureau's seasonal adjustment package, X13-ARIMA-SEATS. We interface with it by using the `seasonal` package implementation in R (Sax and Eddelbuettel 2018).

3.2 Time Periods

Now we will segment the data by time periods. The imported monthly data will have ragged edges - i.e., some monthly data will be available for later months than others.

We will let T denote the number of dates for which data is available for all data series. τ will denote the number of dates for which data is available for at least one data series. T^* will denote number of dates up to the end-of-quarter month of the τ date. For example, suppose date τ occurs on February. The end-of-quarter month, T^*, will be March (since Q1 runs through the end of March).

In other words, data will be indexed by $t = 1, 2, \ldots, T, T + 1, \ldots, \tau, \ldots, T^*$, where dates $T + 1$ through τ are the dates for which only some data are available, and dates $\tau + 1$ through T^* are the dates for which no data is available up to the next quarter-ending month.



For our data, we set the dates as follows.

Date	t
2008-01-01	1
2021-12-01	T
2022-03-01	τ
2025-03-01	$T*$

Table 3: Time Periods

3.3 Principal Components Analysis

It is known that a large number of macroeconomic time series are highly correlated; using such covariates as regressors could naturally lead to problems with collinearity and unstable estimates. In addition, it becomes computationally burdensome to analyze data with such a large number of highly correlated variables. Instead, we use principal components analysis (PCA) to shrink our dataset in a way that allows us to retain most of the information in our original data.

Estimation of factors is derived following Stock and Watson (2008). We begin by taking our $T \times N$ data matrix of N monthly leading economic variables, from time 1 through T. The matrix, which we denote X, is normalized to mean 0 and variance 0 across all columns.

The goal is to minimize the error E below.

$$\begin{aligned} X &= F\Lambda' + E, \\ \text{where } X &\text{ is the } T \times N \text{ data matrix,} \\ F &\text{ is the } T \times N \text{ matrix of factors,} \\ \text{and } \Lambda &\text{ is the weighting matrix.} \end{aligned}$$

Estimation of factors is derived following Stock and Watson (2008).

$$\begin{aligned} \widehat{\Lambda} &= \text{eigenvectors of } (X'X) \\ \widehat{F} &= X\widehat{\Lambda} \end{aligned}$$

Once factors are derived, we select the optimal number of factors to use in predictive regressions. To do so, we use the information criteria from Bai and Ng (2002). Let R refer to the number of factors used. We also include alternative specifications of the information criteria from Bai and Ng as a robustness check.

$$IC(R) = MSE + R \times \frac{N+T}{NT} \times log\left(\frac{NT}{N+T}\right)$$

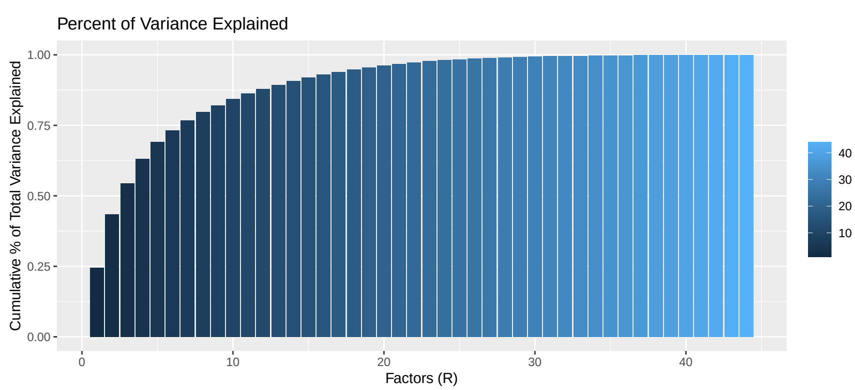
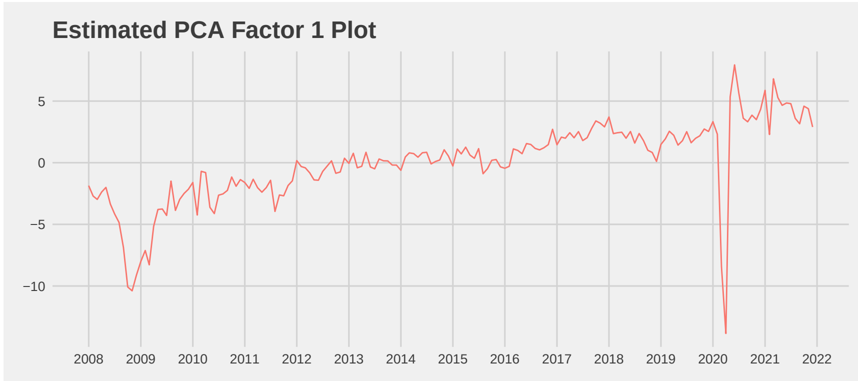


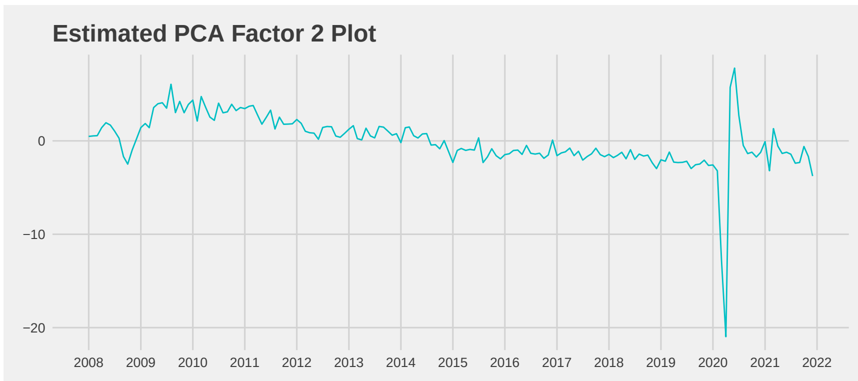
Figure 3: Factor Selection

Factors (R)	Variance Explained	Pct of Total Var Explained	Cumulative Pct	MSE	IC1	IC2	IC3
1	10.78	0.25	0.25	0.75	0.85	0.86	0.84
2	8.33	0.19	0.43	0.56	0.77	0.78	0.73
3	4.89	0.11	0.55	0.45	0.76	0.78	0.71
4	3.76	0.09	0.63	0.37	0.77	0.80	0.71
5	2.67	0.06	0.69	0.31	0.82	0.85	0.74
6	1.77	0.04	0.73	0.27	0.88	0.92	0.78
7	1.57	0.04	0.77	0.23	0.94	0.99	0.83
8	1.31	0.03	0.80	0.20	1.02	1.07	0.89
9	1.05	0.02	0.82	0.18	1.09	1.15	0.95
10	1.02	0.02	0.84	0.15	1.17	1.24	1.01
11	0.81	0.02	0.86	0.14	1.26	1.33	1.08
12	0.72	0.02	0.88	0.12	1.34	1.42	1.15
13	0.65	0.01	0.89	0.11	1.43	1.52	1.22
14	0.58	0.01	0.91	0.09	1.52	1.61	1.30
15	0.55	0.01	0.92	0.08	1.61	1.71	1.37
16	0.47	0.01	0.93	0.07	1.70	1.81	1.45
17	0.42	0.01	0.94	0.06	1.79	1.90	1.52
18	0.36	0.01	0.95	0.05	1.89	2.01	1.60
19	0.34	0.01	0.96	0.04	1.98	2.11	1.68
20	0.30	0.01	0.96	0.04	2.07	2.21	1.76
21	0.25	0.01	0.97	0.03	2.17	2.31	1.84
22	0.24	0.01	0.97	0.03	2.27	2.41	1.92
23	0.18	0.00	0.98	0.02	2.36	2.52	2.00
24	0.16	0.00	0.98	0.02	2.46	2.62	2.08
25	0.13	0.00	0.98	0.02	2.56	2.73	2.17
26	0.11	0.00	0.99	0.01	2.66	2.83	2.25
27	0.10	0.00	0.99	0.01	2.76	2.94	2.33
28	0.09	0.00	0.99	0.01	2.86	3.05	2.42
29	0.07	0.00	0.99	0.01	2.96	3.15	2.50
30	0.06	0.00	0.99	0.01	3.06	3.26	2.59
31	0.05	0.00	1.00	0.00	3.16	3.37	2.67
32	0.04	0.00	1.00	0.00	3.26	3.48	2.76
33	0.03	0.00	1.00	0.00	3.36	3.58	2.84
34	0.03	0.00	1.00	0.00	3.47	3.69	2.93
35	0.02	0.00	1.00	0.00	3.57	3.80	3.01
36	0.02	0.00	1.00	0.00	3.67	3.91	3.10
37	0.01	0.00	1.00	0.00	3.77	4.02	3.18
38	0.01	0.00	1.00	0.00	3.87	4.12	3.27
39	0.01	0.00	1.00	0.00	3.97	4.23	3.35
40	0.01	0.00	1.00	0.00	4.07	4.34	3.44
41	0.01	0.00	1.00	0.00	4.18	4.45	3.53
42	0.00	0.00	1.00	0.00	4.28	4.56	3.61
43	0.00	0.00	1.00	0.00	4.38	4.67	3.70
44	0.00	0.00	1.00	0.00	4.48	4.78	3.78

Table 4: Factor Selection Process

Choosing the IC-minimizing R lets us choose $R = 2$ factors.





Next, we perform a qualitative check of the factors. The first factor usually represents the COVID-19 shock. Typically the second factor should give us something similar to the growth rate of GDP or aggregate production, but on a monthly basis. The third and fourth factors may vary but often represent interest rates or consumption. Note that the sign direction of the factors is irrelevant to the modeling process, and they may be switched negated without consequence.

Finally, we evaluate the components of each factor, i.e. which

f1	f2
baa (-0.25)	ahe (-0.24)
hpermits (0.22)	ipfinal (0.23)
sphfed (0.22)	mpce (0.23)
houst (0.22)	ipi (0.23)
hpi (0.22)	mnocg (0.22)

Table 5: PCA Factor Weights - Top 5 Per Factor

3.4 Factor VAR

The next step is to model the transition of the factors over time. To do so, we utilize a vector-autoregressive (VAR) process, following Stock and Watson (2016). As before, R will refer to the total number of factors we extracted in the previous section, and f_t^i for $i = 1, \ldots, R$ will refer to the value of factor i at time t.

We will use a VAR(1) model of the following form.

$$\underbrace{\begin{bmatrix} f_t^1 \\ f_t^2 \\ \vdots \\ f_t^R \end{bmatrix}}_{z_t} = B \underbrace{\begin{bmatrix} f_{t-1}^1 \\ f_{t-1}^2 \\ \vdots \\ f_{t-1}^R \end{bmatrix}}_{z_{t-1}} + C + \underbrace{\begin{bmatrix} v_t^1 \\ v_t^2 \\ \vdots \\ v_t^R \end{bmatrix}}_{v_t},$$

$$\begin{aligned} \text{where } z_t &\text{ is the } R \times 1 \text{ matrix of time } t \text{ factors,} \\ B &\text{ is the } R \times R \text{ coefficient matrix,} \\ C &\text{ is the } R \times 1 \text{ constant matrix,} \\ \text{and } v_t &\text{ is the } R \times 1 \text{ matrix of errors for time } t. \end{aligned}$$

We wish to estimate the coefficient matrices B and C. This can be done via OLS estimation. We first rewrite the data as the standard linear equation,

$$\underbrace{\begin{bmatrix} f_2^1 & f_2^2 & \dots & f_2^R \\ f_3^1 & f_3^2 & \dots & f_3^R \\ \vdots & \vdots & \vdots & \vdots \\ f_T^1 & f_T^2 & \dots & f_T^R \end{bmatrix}}_{\Gamma} = \underbrace{\begin{bmatrix} 1 & f_1^1 & f_1^2 & \dots & f_1^R \\ 1 & f_2^1 & f_2^2 & \dots & f_2^R \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & f_{T-1}^1 & f_{T-1}^2 & \dots & f_{T-1}^R \end{bmatrix}}_{\Psi} \underbrace{\begin{bmatrix} C' \\ B' \end{bmatrix}}_{\Lambda} + \underbrace{\begin{bmatrix} v_2^1 & v_2^2 & \dots & v_2^R \\ v_3^1 & v_3^2 & \dots & v_3^R \\ \vdots & & & \\ v_T^1 & v_T^2 & \dots & v_T^R \end{bmatrix}}_{V},$$

$$\begin{aligned} \text{where } \Gamma &\text{ is the } T-1 \times R \text{ dependent data matrix,} \\ \Psi &\text{ is the } T-1 \times R+1 \text{ independent data matrix,} \\ \Lambda &\text{ is the } R+1 \times R \text{ matrix of coefficient weightings,} \\ \text{and } V &\text{ is the } T-1 \times R \text{ matrix of residuals.} \end{aligned}$$

The coefficient matrix Λ can be estimated by the standard OLS estimator.

$$\widehat{\Lambda} = (\Psi'\Psi)^{-1}(\Psi'\Gamma)$$

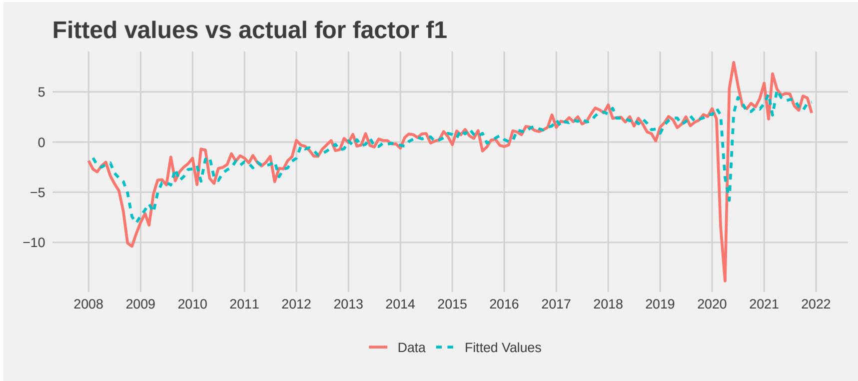
It can then be partitioned to calculate $\widehat{B}'$ and $\widehat{C}'$, which can then be transposed to derive our estimates of the original coefficient matrices B and C, $\widehat{B}$ and $\widehat{C}$.

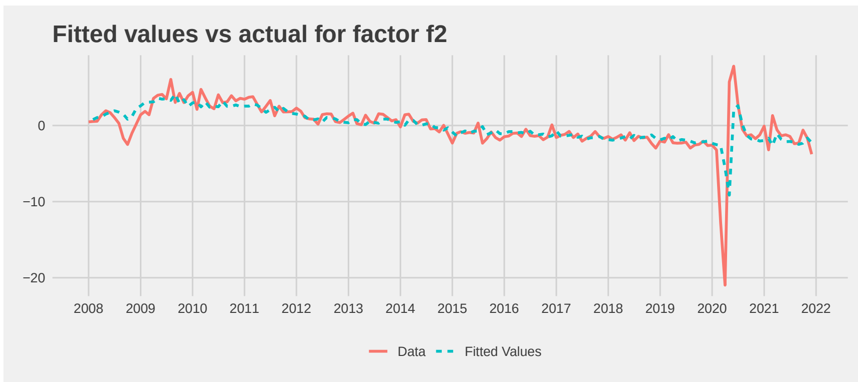
The estimated coefficients in $\widehat{B}$ and $\widehat{C}$ are shown below.

name	constant	f1.l1	f2.l1
f1	0.03	0.80	-0.25
f2	-0.02	-0.27	0.62

Table 6: Factor VAR Coefficients

Finally, we perform a qualitative check of the fitted values and residuals. It is important that factors that are predictable — i.e., factors 2 and 3, since they represent output — have a good fit. Since factor 1 represents the COVID-19 shock, we should expect that the fit is poor; such a shock should not be predictable simply from the time dynamics of the factors; so if the fit is good, our model is likely overfitted.





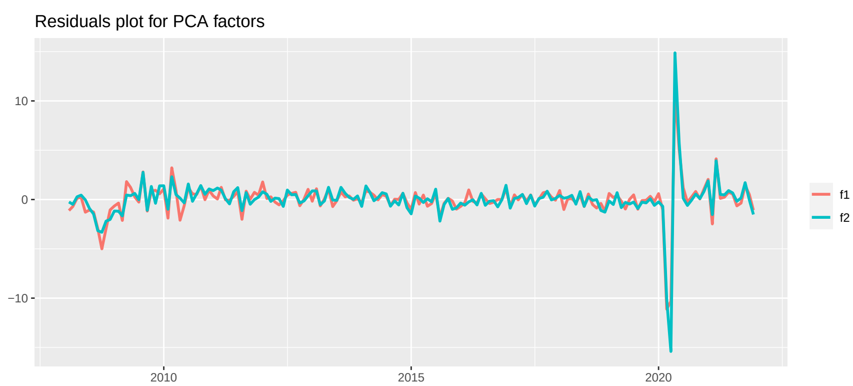


Figure 4: Factor VAR Residuals

Additionally, we expect residuals of the nowcast. Goodness-of-fit statistics are shown below.

varname	MAE	MSE
f1	0.93126717	3.33388276
f2	0.94143730	4.42038307

Table 7: DFM Goodness of Fit

3.5 Dynamic Factor Models

Now let us consider again the monthly leading economic variables which were include in the principal components analysis. We will model these as dynamic factor models (DFMs), i.e. - they are regressed on the factor variables derived from earlier. As before, let x_t^i refer to the time t value for monthly variable x^i, where $i = 1, \ldots, N$.

The factor models take the following form:

$$\underbrace{\begin{bmatrix} x_t^1 \\ x_t^2 \\ \vdots \\ x_t^N \end{bmatrix}}_{y_t} = A \underbrace{\begin{bmatrix} f_t^1 \\ f_t^2 \\ \vdots \\ f_t^R \end{bmatrix}}_{z_t} + D + \underbrace{\begin{bmatrix} w_t^1 \\ w_t^2 \\ \vdots \\ w_t^N \end{bmatrix}}_{w_t},$$

$$\begin{aligned} \text{where } y_t &\text{ is the } N \times 1 \text{ vector of monthly variables at time } t, \\ A &\text{ is the } N \times R \text{ coefficient matrix,} \\ z_t &\text{ is the } R \times 1 \text{ vector of factors at time } t, \\ D &\text{ is the } N \times 1 \text{ constant matrix,} \\ \text{and } w_t &\text{ is the } N \times 1 \text{ vector of errors at time } t. \end{aligned}$$

We wish to estimate the coefficient matrices A and D. As before, we can do this by estimating this as an OLS equation, writing the data matrices as follows

$$\underbrace{\begin{bmatrix} x_2^1 & x_2^2 & \ldots & x_2^N \\ x_3^1 & x_3^2 & \ldots & x_3^N \\ \vdots & \vdots & \vdots & \vdots \\ x_T^1 & x_T^2 & \ldots & x_T^N \end{bmatrix}}_{\Phi} = \underbrace{\begin{bmatrix} 1 & f_2^1 & f_2^2 & \ldots & f_2^R \\ 1 & f_3^1 & f_3^2 & \ldots & f_3^R \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & f_T^1 & f_T^2 & \ldots & f_T^R \end{bmatrix}}_{\Omega} \begin{bmatrix} D' \\ A' \end{bmatrix} + \underbrace{\begin{bmatrix} w_2^1 & w_2^2 & \ldots & w_2^R \\ w_3^1 & w_3^2 & \ldots & w_3^R \\ \vdots & & & \\ w_T^1 & w_T^2 & \ldots & w_T^R \end{bmatrix}}_{W} \tag{1}$$

However, estimation of A and D with the standard OLS estimator is likely to lead to significant overfitting, particularly if the matrix of f factors is particularly high-dimensional. We will instead use an elastic net process to regularize covariate selection (Zou and Hastie 2005). Hyperparameters of the elastic net are chosen through a cross-validated MAE-minimization process.

Estimated coefficients for $\widehat{A}$ and $\widehat{D}$ are below.

name	constant	f1	f2
mpce	0.00	0.00	0.00
psr	7.90	0.29	-0.44
pid	0.00	-0.00	-0.00
bankci	0.00	-0.00	-0.00
bankre	0.00	0.00	-0.00
hpi	0.00	0.00	-0.00
houst	6.88	0.08	-0.07
hsold	6.21	0.06	-0.06
hpermits	6.93	0.08	-0.07
pcepi	1.68	0.08	-0.03
cpi	1.94	0.08	-0.05
ppi	0.00	0.00	0.00
unemp	1.81	-0.05	0.06
jolts	0.01	0.01	0.01
vacancies	0.70	0.79	0.75
ahe	0.00	-0.00	-0.00
kclfi	-0.52	0.19	-0.22
lfpr	63.41	-0.27	0.21
ipi	-0.00	0.00	0.00
ipfinal	-0.00	0.00	0.00
ipcons	-0.00	0.00	0.00
ipfuels	0.00	0.00	0.00
mnodg	0.21	0.54	0.62
mnocg	0.12	0.47	0.63
cons	0.00	0.00	-0.00
isratio	1.34	-0.00	-0.01
cu	75.74	0.32	-0.20
ffr	0.62	0.02	-0.08
mort30y	4.15	-0.16	0.09
aaa	4.03	-0.20	0.15
baa	5.12	-0.30	0.17
vsales	-0.00	0.01	0.01
advsales	0.00	0.00	0.01
csent	99.37	0.28	-0.27
snyfed	18.01	1.30	0.09
stxfed	-0.76	4.60	0.67
sphfed	19.13	2.80	-1.20
spy	0.01	0.01	0.00
nasdaq	0.01	0.01	0.00
vix	20.12	-1.88	-0.06
moo	-0.00	0.00	0.00
metals	0.00	0.01	0.00
oil	70.90	-1.48	3.60
spdw	0.00	0.01	0.00

Table 8: Estimated DFM Coefficients

We perform a qualitative check of the in-sample fit, as well as a quantitative review of the goodness-of-fit. The graphs of the fitted plots are located in the appendix.

varname	MAE	MSE
aaa	0.39344229	0.22828074
advsales	0.00808897	0.00020074
ahe	0.00167021	0.00000602
baa	0.40905583	0.24898455
bankci	0.00873153	0.00015126
bankre	0.00300415	0.00002875
cons	0.00865432	0.00011690
cpi	1.07721344	2.19697726
csent	0.78494594	1.00816983
cu	2.21919553	8.73556325
ffr	0.59735127	0.61255829
houst	0.09413276	0.01625652
hpermits	0.08429309	0.01343295
hpi	0.00338827	0.00001677
hsold	0.13593501	0.02958403
ipcons	0.00680810	0.00008283
ipfinal	0.00549016	0.00005781
ipfuels	0.01350831	0.00052173
ipi	0.00508908	0.00006119
isratio	0.04829056	0.00372746
jolts	0.03759362	0.00228391
kclfi	0.49716247	0.41437379
lfpr	0.45687101	0.40676825
metals	0.03486406	0.00217619
mnocg	1.14488142	2.59104533
mnodg	3.21585987	21.64716232
moo	0.02650468	0.00135727
mort30y	0.43155750	0.30565986
mpce	0.00537784	0.00006352
nasdaq	0.02778795	0.00166279
oil	16.88091114	427.59618333
pcepi	0.83838869	1.32033022
pid	0.00968491	0.00071069
ppi	0.00710778	0.00009908
psr	1.93195222	11.41118130
snyfed	7.61073400	90.95586497
spdw	0.02664204	0.00160124
sphfed	6.18172225	58.10948406
spy	0.02300125	0.00119901
stxfed	11.58513026	218.23903302
unemp	0.17764827	0.05701605
vacancies	3.79725668	23.52365936
vix	5.19217345	46.58433485
vsales	0.03231990	0.00319191

Table 9: DFM Goodness of Fit

3.6 State-Space Setup

Now, combining our equations for the DFM and the VAR, we have the below system.

$$z_t = Bz_{t-1} + Cx + v_t$$
$$y_t = Az_t + w_t$$

This system is now fully specified and in state-space form. The first equation is our state (or transition) equation. The second equation is our measurement equation.

We use our estimated values B, C, A, and D calculated in our previous two sections. To run the Kalman Filter, we will want to create the actual data matrices for z_t and y_t. z_t can be constructed as before, using data for factors from time 1 through T. However, unlike in the previous two sections, we will want to create y_t matrices not for just time periods 1 through T, but now for time periods 1 through τ. Elements in y_t may be set to any value for missing observations; the process of Kalmam filtration will render this choice irrelevant.

Specifically, we construct the matrices below.

$$z_t = \begin{bmatrix} f_t^1 \\ f_t^2 \\ \vdots \\ f_t^R \end{bmatrix}, \forall t \in 1, \ldots, T$$

$$y_t = \begin{bmatrix} x_t^1 \text{ if available, otherwise } 0 \\ x_t^2 \text{ if available, otherwise } 0 \\ \vdots \\ x_t^N \text{ if available, otherwise } 0 \end{bmatrix}, \forall t \in 1, \ldots, \tau$$

For Kalman filtration, we also require an assumed distribution on v_t and w_t. We assume that v_t is distributed normally with mean 0 and constant diagonal covariance matrix denoted Q, with diagonal entries calculated by taking the average squared values of the residuals of the VAR.

We also assume w_t is distributed normally with mean 0. However, we no longer specify the covariance matrix as constant, but as the time-dependent matrices R_t. For $t \in 1, \ldots, T+1$, we let R_t be a diagonal covariance matrix with diagonal entries calcualted by taking the average squared values of the residuals of the DFM. For $t \in T+1, \ldots, \tau$, we let the diagonal elements of R_t be equal to infinity if the corresponding element of y_t is missing for that time period; and equal to the average squared value of the residual if otherwise.

$$\begin{aligned} v_t &\sim \mathcal{N}(0, Q) \\ w_t &\sim \mathcal{N}(0, R_t) \end{aligned}$$

3.7 Kalman Filtration

Now that our state-space model is fully specified, we can begin the Kalman filter recursions.

$$\begin{aligned} z_t &= Bz_{t-1} + Cx + v_t \\ y_t &= Az_t + Dx + w_t \\ v_t &\sim \mathcal{N}(0, Q) \\ w_t &\sim \mathcal{N}(0, R_t) \end{aligned}$$

To solve this programmatically, we will need the previously estimated matrices A, B, C, and D; the matrices z_t from 1 through T; the matrices y_t from 1 through τ; the covariance matrix Q; and finally, the covariance matrices R_t from 1 through τ.

We initialize the Kalman filter with the following standard assumptions.

$$\begin{aligned} \mathbf{z}_{0|0} &= 0 \\ \mathbf{CovZ} &= 0 \end{aligned}$$

Now for $t = 1, \ldots, \tau$, we iterate through the Kalman filter recursions and iteratively calculate the values below.

$$\begin{aligned} \mathbf{z}_{t|t-1} &= B\mathbf{z}_{t-1|t-1} + C \\ \mathbf{CovZ}_{t|t-1} &= B\mathbf{CovZ}_{t-1|t-1} + Q \\ \mathbf{y}_{t|t-1} &= A\mathbf{z}_{t|t-1} + D \\ \mathbf{CovY}_{t|t-1} &= A\mathbf{CovZ}_{t|t-1}A' + R_t \\ P_t &= \mathbf{CovZ}_{t|t-1}A'\mathbf{CovY}_{t|t-1}^{-1} \\ \mathbf{z}_{t|t} &= \mathbf{z}_{t|t-1} + P_t(\mathbf{y}_t - \mathbf{y}_{t|t-1}) \\ \mathbf{CovZ}_{t|t} &= \mathbf{CovZ}_{t|t-1} - P_t(\mathbf{CovY}_{t|t-1})P_t' \end{aligned}$$

Note that the during recursions $T + 1, \ldots \tau$, the infinite values in the R_t matrix will cause infinite values in the $\mathbf{CovY}_{t|t-1}$ matrix. This may prevent standard computational methods from computing the inverse of the matrix needed in the step for calculation of $\mathbf{CovZ}_{t|t}$. Alternative methods, such as a Cholesky decomposition before inversion, are used to subvert this problem.

The Kalman filter allows us to recover all the time t conditional state matrices $z_{t|t}$ that have been adjusted for information from the monthly datasets. However, of more interest to us is the value of the state matrices when conditioned on all data available at time τ, $z_{t|\tau}$. This can be recovered by using the Kalman smoother.

Recursively iterating over $t = \tau - 1, \ldots, 1$, we calculate the following values.

$$\begin{aligned} S_t &= \mathbf{CovZ}_{t|t}B'\mathbf{CovZ}_{t+1|t}^{-1} \\ \mathbf{z}_{t|\tau} &= \mathbf{z}_{t|t} + S_t(\mathbf{z}_{t+1|\tau} - \mathbf{z}_{t+1|t}) \\ \mathbf{CovZ}_{t|\tau} &= \mathbf{CovZ}_{t|t} - S_t(\mathbf{CovZ}_{t+1|t} - \mathbf{CovZ}_{t+1|\tau})S_t' \end{aligned}$$

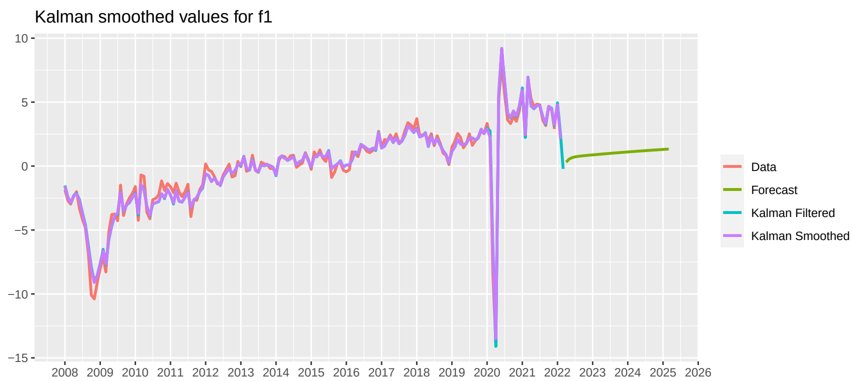
These values $\mathbf{z}_{t|\tau}$ will serve as our estimates of the state variables (i.e., the PCA factors) from time 1 through τ.

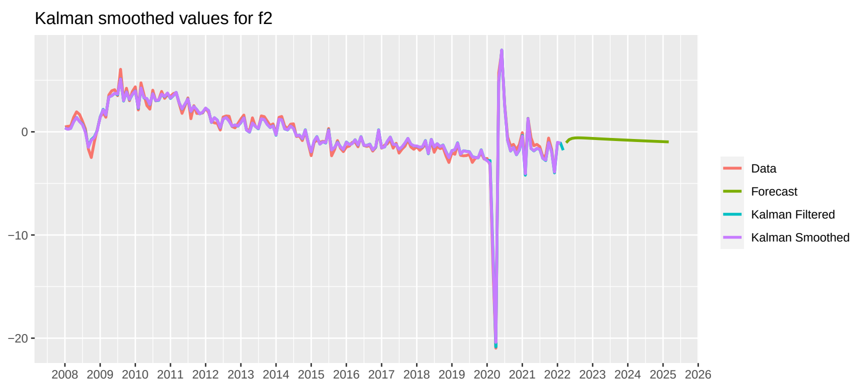
Finally, we want to forecast the the state vector $z_{t|\tau}$ for $t = \tau + 1, \ldots, T^*$. This can be done through the typical Kalman filter forecasting step.

Recursively iterating over $t = \tau + 1, \ldots, T^*$, we calculate the following values.

$$\begin{aligned} \mathbf{z}_{t|\tau} &= B\mathbf{z}_{t-1|\tau} + C \\ \mathbf{CovZ}_{t|\tau} &= B\mathbf{CovZ}_{t-1|\tau}B' + Q \\ \mathbf{y}_{t|\tau} &= A\mathbf{z}_{t|\tau} + D \\ \mathbf{CovY}_{t|\tau} &= A\mathbf{CovZ}_{t|\tau}A' + R_0 \end{aligned}$$

Combining the calculations for $\mathbf{z}_{t|\tau}$ with the ones derived from the Kalman smoother, we will now be able to obtain the full time series for the factors from time 1 through time T^*.





3.8 Nowcast Step

Now that we have our Kalman-smoothed and forecasted factors from time 1 through time T^*, we will be able to use these as covariates to model any monthly time series that we have.

We begin by aggregating these monthly factors into quarterly data by taking a simple monthly average for each factor over each quarter.

In this section, we will use these now-quarterly factors to forecast our quarterly subcomponents of GDP. As discussed in the data import section, these have been transformed for stationarity, typically by taking a log-difference.

Variable
Consumption: Goods: Durable: Motor Vehicles & Parts
Consumption: Goods: Durable: Furnishings & Durable Household Equipment
Consumption: Goods: Durable: Recreational Goods & Services
Consumption: Goods: Durable: Other Durable Goods
Consumption: Goods: Nondurable Goods: Food & Beverages for Off-Premises Use
Consumption: Goods: Nondurable Goods: Clothing
Consumption: Goods: Nondurable Goods: Gasoline & Energy Goods
Consumption: Goods: Nondurable Goods: Other Nondurable Goods
Consumption: Services: Housing & Utilities
Consumption: Services: Health Care
Consumption: Services: Transportation
Consumption: Services: Recreation
Consumption: Services: Food Services & Accomodations
Consumption: Services: Financial Services
Consumption: Services: Other Services
Consumption: Services: Nonprofit Services
Gross Private Domestic Investment: Nonresidential Investment: Structures
Gross Private Domestic Investment: Nonresidential Investment: Equipment
Gross Private Domestic Investment: Nonresidential Investment: Intellectual Property
Gross Private Domestic Investment: Residential Investment
Gross Private Domestic Investment: Change in Private Inventories
Exports: Exported Goods
Exports: Exported Services
Imports: Imported Goods
Imports: Imported Services
Government Spending: Federal
Government Spending: State & Local

Table 10: Quarterly Data Covariates

We will notate each of these gdp subcomponents as y^i and M as the total number of covariates, so that $i = 1, \ldots, M$.

Note that many higher-level components of GDP (including GDP itself) are *not* forecasted directly in this step; these will be forecasted later by aggregating their subcomponents.

Now we will specify that these GDP subcomponents follow a DFM-AR(1) model; i.e., they will be functions of the monthly-aggregated factors as well as the first lag of themselves. For any variable i,

$$y_t^i = \beta \begin{bmatrix} 1 \\ y_{t-1}^i \\ f_t^1 \\ \vdots \\ f_t^R \end{bmatrix} + e_t.$$

The $R + 2 \times 1$ β matrix can be estimated with a GLS process, where the training data is constituted of the y_t^i quarterly covariates, the lagged quarterly covariates y_{t-1}^i, and the quarterly-aggregated factor variables. The data is cut off at the date for which the any data on the quarterly covariates are missing.

In particular, to control for overfitting, we use an elastic net process, a linear combination of the standard LASSO and ridge regression estimators. We use a grid search process to find an optimal combination of the L_1 and $L2$ coefficient hyperparameters. The optimal values for each variable are given below. Plots for the selection process can be found in the appendix.

Variable	l1	l2
Consumption: Goods: Durable: Motor Vehicles & Parts	1.00	0.00
Consumption: Goods: Durable: Furnishings & Durable Household Equipment	0.50	0.00
Consumption: Goods: Durable: Recreational Goods & Services	0.50	0.00
Consumption: Goods: Durable: Other Durable Goods	0.00	0.00
Consumption: Goods: Nondurable Goods: Food & Beverages for Off-Premises Use	0.00	1.93
Consumption: Goods: Nondurable Goods: Clothing	0.50	0.00
Consumption: Goods: Nondurable Goods: Gasoline & Energy Goods	0.50	0.00
Consumption: Goods: Nondurable Goods: Other Nondurable Goods	1.00	0.00
Consumption: Services: Housing & Utilities	1.00	0.00
Consumption: Services: Health Care	1.00	0.00
Consumption: Services: Transportation	0.50	0.00
Consumption: Services: Recreation	0.50	0.01
Consumption: Services: Food Services & Accomodations	0.50	0.00
Consumption: Services: Financial Services	0.00	0.00
Consumption: Services: Other Services	0.50	0.01
Consumption: Services: Nonprofit Services	0.50	0.01
Gross Private Domestic Investment: Nonresidential Investment: Structures	1.00	0.00
Gross Private Domestic Investment: Nonresidential Investment: Equipment	1.00	0.00
Gross Private Domestic Investment: Nonresidential Investment: Intellectual Property	1.00	0.00
Gross Private Domestic Investment: Residential Investment	0.50	0.01
Gross Private Domestic Investment: Change in Private Inventories	0.50	59.83
Exports: Exported Goods	0.50	0.00
Exports: Exported Services	1.00	0.00
Imports: Imported Goods	0.00	0.00
Imports: Imported Services	0.50	0.01
Government Spending: Federal	0.00	0.00
Government Spending: State & Local	1.00	0.00

Table 11: Elastic Net Hyperparameters

After estimation, we then use the same model to iteratively forecast forward the y_t^i quarterly, up through time τ. The Kalman forecast values are used for the f forecasts.

The forecasted results are as follows.

Variable	2022Q1	2022Q2	2022Q3	2022Q4	2023Q1	2023
Consumption: Goods: Durable: Motor Vehicles & Parts	0.03	0.00	0.02	0.02	0.02	0
Consumption: Goods: Durable: Furnishings & Durable Household Equipment	-0.01	-0.01	-0.01	-0.01	-0.01	-0
Consumption: Goods: Durable: Recreational Goods & Services	0.02	0.01	0.02	0.02	0.02	0
Consumption: Goods: Durable: Other Durable Goods	-0.00	-0.01	-0.01	-0.01	-0.01	-0
Consumption: Goods: Nondurable Goods: Food & Beverages for Off-Premises Use	-0.01	-0.01	-0.01	-0.01	-0.01	-0
Consumption: Goods: Nondurable Goods: Clothing	0.00	-0.01	-0.00	-0.00	-0.00	-0
Consumption: Goods: Nondurable Goods: Gasoline & Energy Goods	-0.01	-0.01	-0.01	-0.01	-0.01	-0
Consumption: Goods: Nondurable Goods: Other Nondurable Goods	-0.00	-0.00	-0.00	-0.00	-0.00	-0
Consumption: Services: Housing & Utilities	-0.00	-0.00	-0.00	-0.00	-0.00	-0
Consumption: Services: Health Care	0.01	0.01	0.01	0.01	0.01	0
Consumption: Services: Transportation	0.00	-0.00	-0.00	-0.00	-0.00	-0
Consumption: Services: Recreation	-0.00	-0.01	-0.00	-0.00	-0.00	-0
Consumption: Services: Food Services & Accomodations	0.00	0.00	0.00	0.00	0.00	0
Consumption: Services: Financial Services	0.01	0.00	0.00	0.00	0.00	0
Consumption: Services: Other Services	0.00	0.00	0.00	0.00	0.00	0
Consumption: Services: Nonprofit Services	-0.06	-0.05	-0.05	-0.05	-0.05	-0
Gross Private Domestic Investment: Nonresidential Investment: Structures	-0.00	0.00	0.00	0.01	0.01	0
Gross Private Domestic Investment: Nonresidential Investment: Equipment	0.01	0.00	0.00	0.00	0.01	0
Gross Private Domestic Investment: Nonresidential Investment: Intellectual Property	0.02	0.01	0.02	0.02	0.02	0
Gross Private Domestic Investment: Residential Investment	0.01	0.01	0.01	0.01	0.01	0
Gross Private Domestic Investment: Change in Private Inventories	-52.88	-52.88	-52.88	-52.88	-52.88	-52
Exports: Exported Goods	0.02	0.01	0.01	0.01	0.01	0
Exports: Exported Services	-0.03	0.02	-0.01	0.01	-0.00	0
Imports: Imported Goods	0.02	0.01	0.01	0.01	0.01	0
Imports: Imported Services	0.01	0.01	0.01	0.01	0.01	0
Government Spending: Federal	-0.00	0.00	0.00	0.00	0.00	0
Government Spending: State & Local	-0.00	0.00	-0.00	-0.00	-0.00	-0

Table 12: DFM-AR(1) Forecasted GDP Subcomponents

We can then backtransform the data so that the units are in base values. After backtransformation, we are ready to aggregate these up to higher-level GDP components. In particular, we calculate the variables below.

Variable
Real GDP
Personal Consumption
Consumption: Goods
Consumption: Goods: Durable
Consumption: Goods: Nondurable Goods
Consumption: Services
Gross Private Domestic Investment
Gross Private Domestic Investment: Nonresidential Investment
Net Exports
Exports
Imports
Government Spending

Table 13: Summable Quarterly Data Covariates

These are calculated using the standard GDP aggregation equations, e.g., net exports = exports - imports, and so on. Finally, we convert these into annualized percentage change, as this is the standard format in which GDP subcomponents are reported in. The results are reported in the next section.

We also use a similar DFM-AR(1) specification to forecast out other economic variables of interest, with results reported in the next section.

4 Results

Our final nowcasts of GDP and its subcomponents are below. All units are reported in terms of annualized percentage change (seasonally adjusted) except for change in private inventories and net exports, which are reported in terms of billions of real 2012 dollars.

Variable	2022Q1	2022Q2	2022Q3	2022Q4	2023Q1	2023
Real GDP	3.49	0.08	0.37	0.71	0.59	0
Personal Consumption	7.78	-0.58	0.14	0.13	0.20	0
Goods	6.29	-1.26	0.37	0.24	0.35	0
Durable	19.24	0.17	3.33	2.99	3.20	3
Motor Vehicles & Parts	13.57	1.53	8.33	7.27	7.50	7
Furnishings & Durable Household Equipment	-2.52	-5.05	-3.61	-3.59	-3.59	-3
Recreational Goods & Services	8.52	4.18	6.94	6.49	6.70	6
Other Durable Goods	-1.09	-4.07	-3.65	-3.54	-3.49	-3
Nondurable Goods	-5.50	-2.21	-1.62	-1.62	-1.60	-1
Food & Beverages for Off-Premises Use	-2.70	-2.69	-2.69	-2.69	-2.69	-2
Clothing	1.09	-2.33	-0.01	-0.04	-0.06	-0
Gasoline & Energy Goods	-2.74	-4.53	-2.83	-2.98	-2.98	-3
Other Nondurable Goods	-0.05	-0.98	-0.92	-0.88	-0.84	-0
Services	0.72	-0.12	-0.00	0.06	0.10	0
Housing & Utilities	-0.61	-0.61	-0.61	-0.61	-0.61	-0
Health Care	3.14	3.14	3.14	3.14	3.14	3
Transportation	1.99	-1.46	-1.30	-1.20	-1.16	-1
Recreation	-1.95	-3.01	-1.92	-1.88	-1.93	-1
Food Services & Accomodations	0.84	0.19	0.38	0.41	0.43	0
Financial Services	2.31	1.34	1.21	1.20	1.21	1
Other Services	1.06	0.44	0.53	0.56	0.58	0
Nonprofit Services	-22.34	-18.54	-18.82	-18.81	-18.86	-18
Gross Private Domestic Investment	4.16	2.87	3.43	3.69	3.82	3
Nonresidential Investment	8.79	2.75	3.45	3.78	3.93	4
Structures	-1.37	0.35	1.51	2.03	2.27	2
Equipment	3.51	0.64	1.44	1.90	2.08	2
Intellectual Property	7.76	6.03	6.41	6.48	6.53	6
Residential Investment	4.45	4.41	4.53	4.60	4.65	4
Change in Private Inventories	247.37	247.37	247.37	247.37	247.37	247
Net Exports	-1391.61	-1396.02	-1416.81	-1423.12	-1439.02	-1448
Exports	8.02	5.18	2.17	4.62	3.06	4
Exported Goods	8.36	3.66	4.18	4.37	4.40	4
Exported Services	-11.48	9.78	-3.57	5.37	-0.84	3
Imports	10.74	3.75	3.57	3.59	3.59	3
Imported Goods	8.39	4.01	3.81	3.83	3.83	3
Imported Services	2.02	2.02	2.02	2.02	2.02	2
Government Spending	-0.63	0.12	0.09	0.08	0.09	0
Federal	-1.09	0.16	0.39	0.44	0.45	0
State & Local	-0.20	0.09	-0.11	-0.16	-0.15	-0

apchg = annualized % change

Table 14: Nowcasts for GDP and Subcomponents (Annualized Percent Change)

Backtested forecasts of these variables can be found in the appendix.

We also report nowcasts of other variables of interest, including other quarterly variables, as well as all monthly variables used in this process. Note that because these variables are ragged-edge in nature, there may be "blank" columns. These columns simply indicate that historical data already exists for those columns. Only nowcasts produced by the model are reported.

Variable	2022Q1	2022Q2	2022Q3	2022Q4	2023Q1	2023Q2
Delinquency Rate on Residential Mortgages	2.24	2.15	2.05	1.95	1.85	1.76
Delinquency Rate on Credit Cards	1.64	1.68	1.70	1.73	1.76	1.78
Delinquency Rate on Commercial \& Industrial Loans	1.05	1.01	0.96	0.90	0.85	0.80

Table 15: Nowcasts for Other Quarterly Variables

Variable	2022M2	2022M3	2022M4	2022M5	2
PCE *(apchg)*	0.57	-6.98	-0.32	0.48	
Personal Outlays *(apchg)*	2.16	-5.70	1.29	2.27	
Personal Savings *(apchg)*	130.50	56.48	25.60	16.10	
Personal Savings Rate	0.08	0.08	0.09	0.09	
Personal Income *(apchg)*	7.10	2.73	3.38	2.33	
Disposable Personal Income *(apchg)*	5.79	4.24	2.78	2.59	
Bank Commercial & Industrial Loans *(pchg)*		0.55	0.75	0.71	
Bank Real Estate Loans *(pchg)*		0.30	0.25	0.23	
Case-Shiller National Housing Price Index *(apchg)*	15.32	13.35	11.91	10.79	
Case-Shiller 20-City Composite Home Price Index *(apchg)*	17.46	15.58	14.17	13.05	
Housing Starts	1506.87	1333.88	1236.43	1178.71	1
Houses Sold	774.07	728.10	694.44	668.56	
New Private Housing Permits	1745.42	1558.59	1437.35	1355.42	1
Zillow Home Value Index	330415.98	335156.55	339903.76	344660.22	349
1-Year PCEPI Inflation	6.10	6.01	5.97	5.94	
1-Year CPI Inflation	7.56	7.42	7.35	7.30	
Producer Price Index: All Commodities	246.07	246.42	246.57	246.82	
Unemployment Rate	3.86	3.94	3.90	3.82	
All Employees (Nonfarm)	149919.57	149195.10	148883.76	148786.87	148
Total Nonfarm Job Openings	11422.87	11275.24	11333.57	11374.40	11
Total Unfilled Job Vacancies	11440217.63	11319134.75	11397450.49	11458576.26	11552
Average Hourly Earnings	31.70	31.83	31.93	32.02	
Labor Force Participation Rate	62.19	62.04	61.94	61.86	
Industrial Production index *(apchg)*	3.62	-8.46	-4.48	-2.00	
Industrial Production: Final Products	102.53	101.65	101.25	101.06	
Industrial Production: Consumer Goods	102.68	101.91	101.57	101.41	
Industrial Production: Fuels	90.50	89.59	89.35	89.21	
Manufacturers New Orders: Durable Goods	278415.85	274289.58	274941.51	274874.79	275
Manufacturers New Orders: Consumer Goods	220853.27	218319.64	217162.55	216797.51	216
Total Construction Spending	1691647.50	1698056.73	1703553.37	1709424.45	1715
Inventory to Sales Ratio	1.29	1.31	1.33	1.33	
Capacity Utilization	77.73	77.17	76.92	76.80	
Federal Funds Rate		0.05	0.05	0.06	
30-Year Mortgage Rate		3.75	3.74	3.74	
Corporate Aaa Bond Yield		3.25	3.26	3.27	
Corporate Baa Bond Yield		4.07	4.14	4.20	
Vehicle Sales	14.47	14.37	14.19	14.15	
Retail & Food Sales *(apchg)*	-5.88	-7.41	1.69	2.07	
Consumer Confidence (OECD)	97.70	97.62	97.57	97.53	
S&P 500 *(pchg)*			0.42	0.61	
Nasdaq Composite *(pchg)*			0.46	0.89	
Bitcoin *(pchg)*		3.54	5.42	6.23	
VIX			29.94	27.26	
Agricultural Commodities (DBA) *(pchg)*			0.30	-0.15	
Metal Commodities (DBB) *(pchg)*			1.30	0.42	
Crude Oil Price (West Texas Intermediate)		88.36	86.67	85.52	
U.S. Dollar Index			26.01	25.93	
Ex-US Equities (SPDW) *(pchg)*			-0.65	0.03	

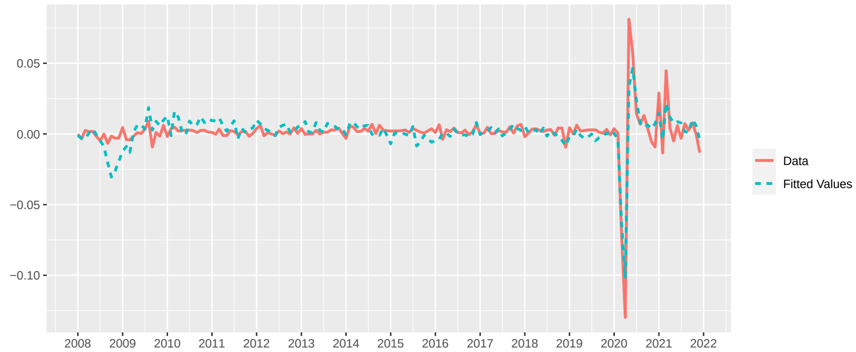
apchg = annualized % change

Table 16: Nowcasts for All Monthly Variables

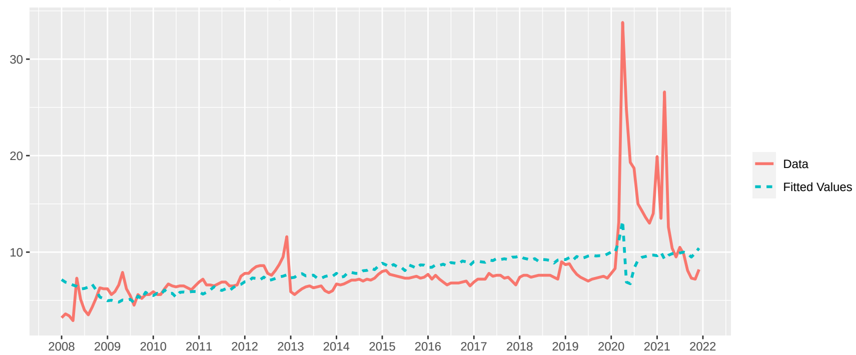
Appendices

A DFM Fitted Plots

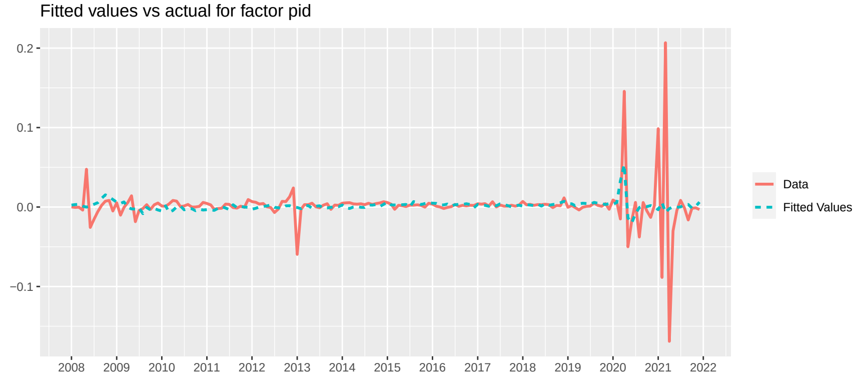
Fitted values vs actual for factor mpce

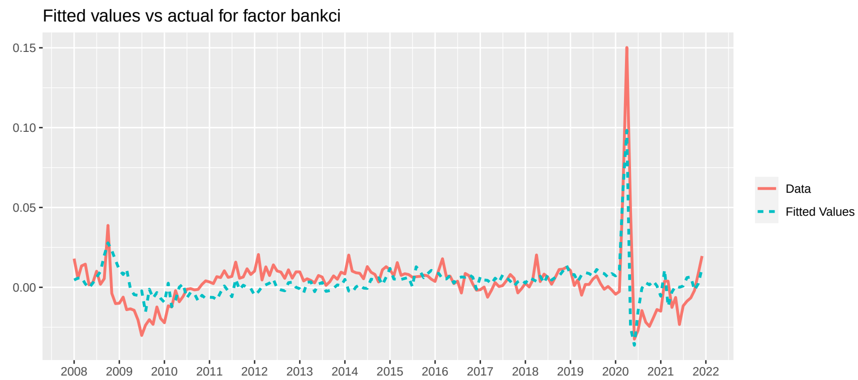


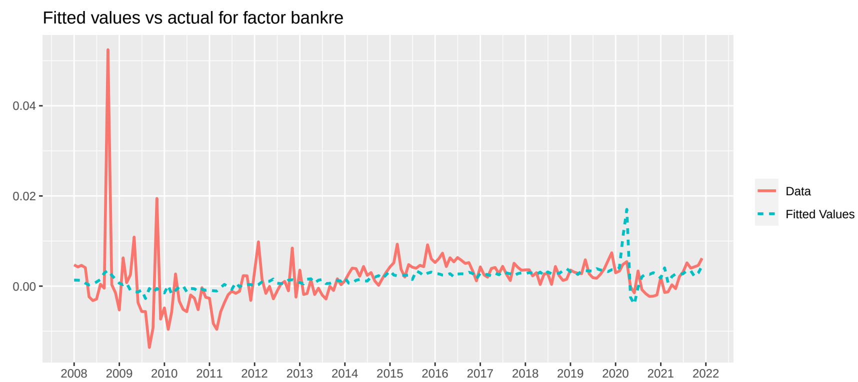
Fitted values vs actual for factor psr

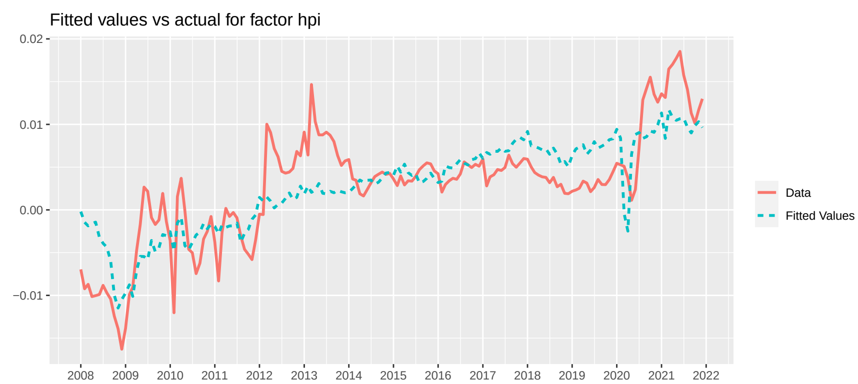


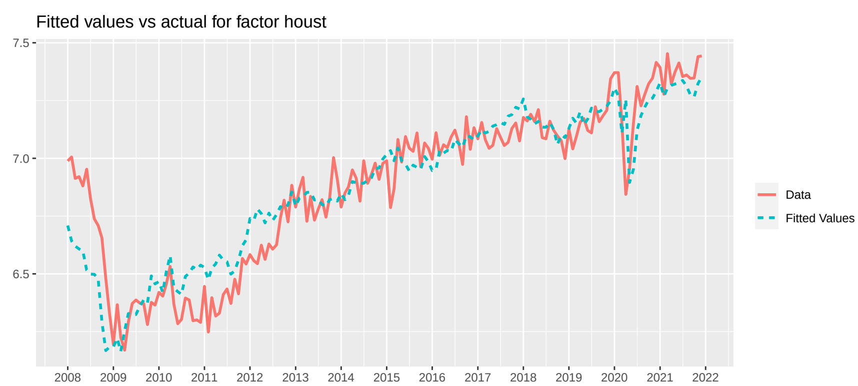
Fitted values vs actual for factor pid



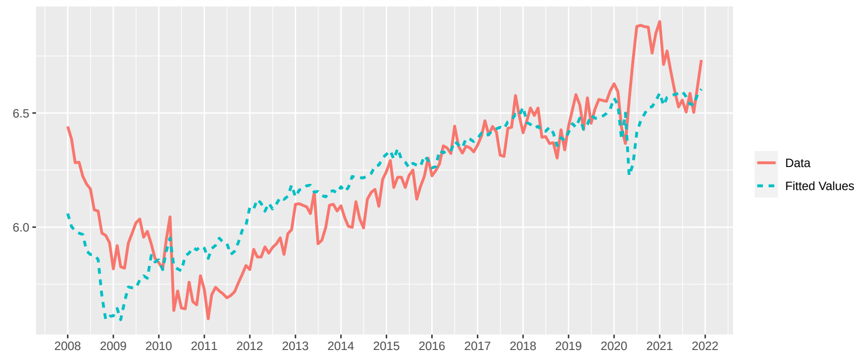

Fitted values vs actual for factor bankci
Data
Fitted Values


Fitted values vs actual for factor bankre
Data
Fitted Values

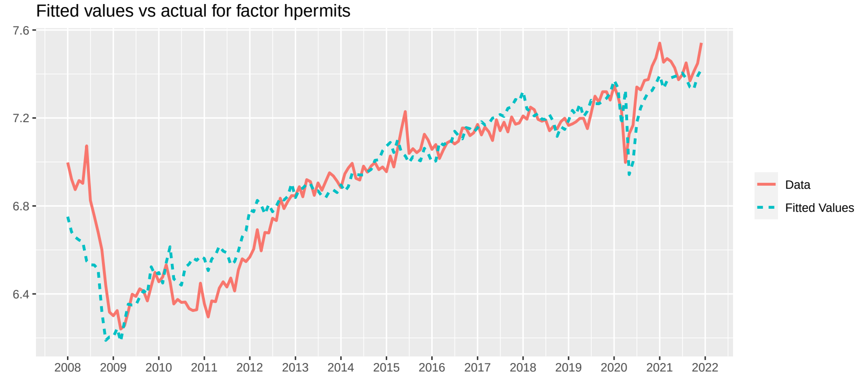

Fitted values vs actual for factor hpi
Data
Fitted Values


Fitted values vs actual for factor houst
Data
Fitted Values

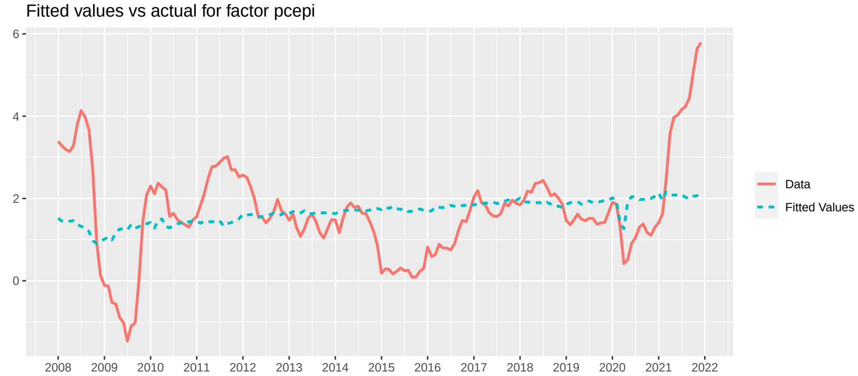
Fitted values vs actual for factor hsold



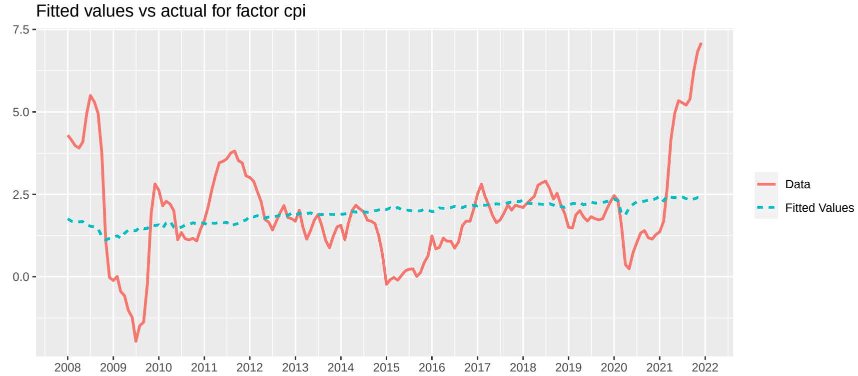
Fitted values vs actual for factor hpermits

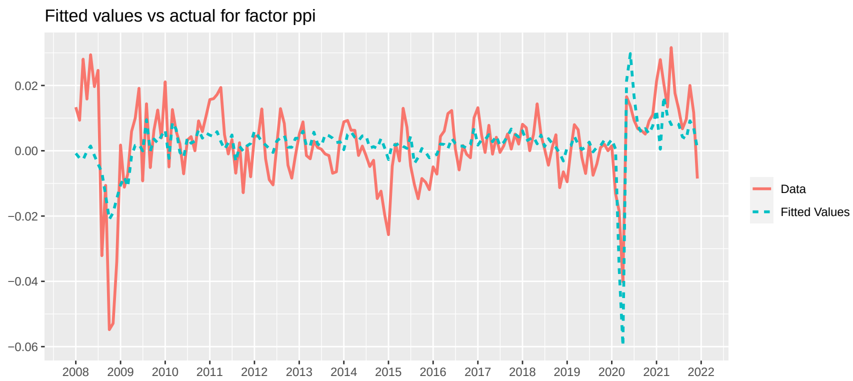


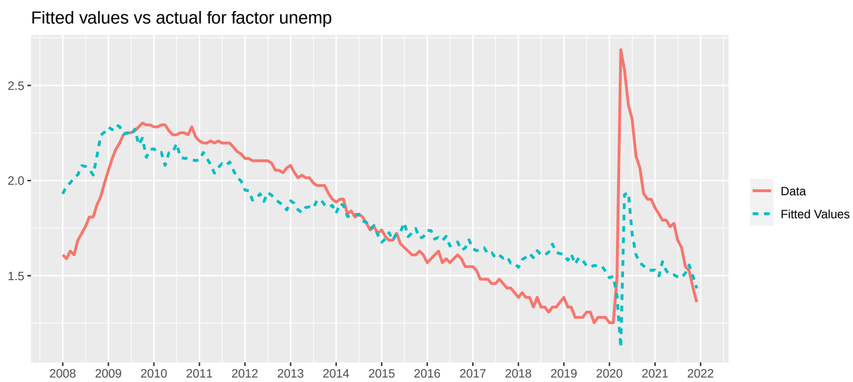
Fitted values vs actual for factor pcepi

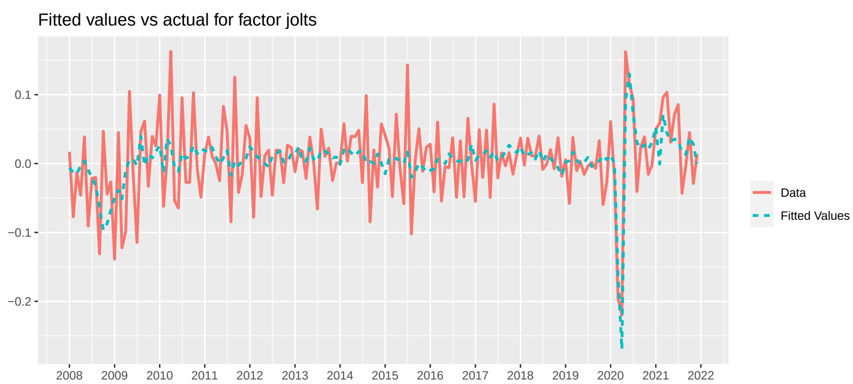


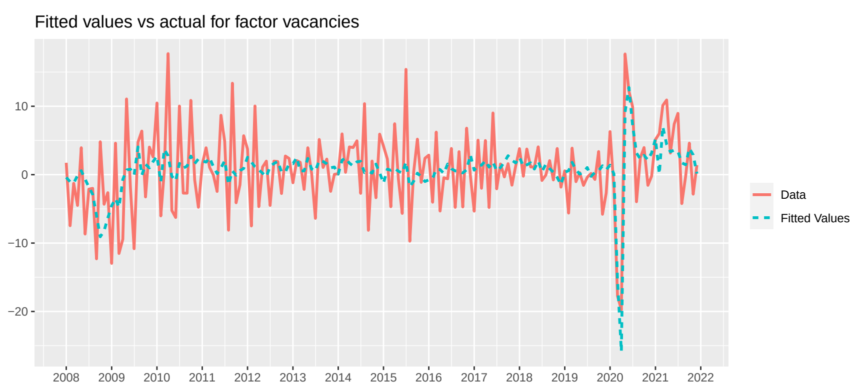
Fitted values vs actual for factor cpi

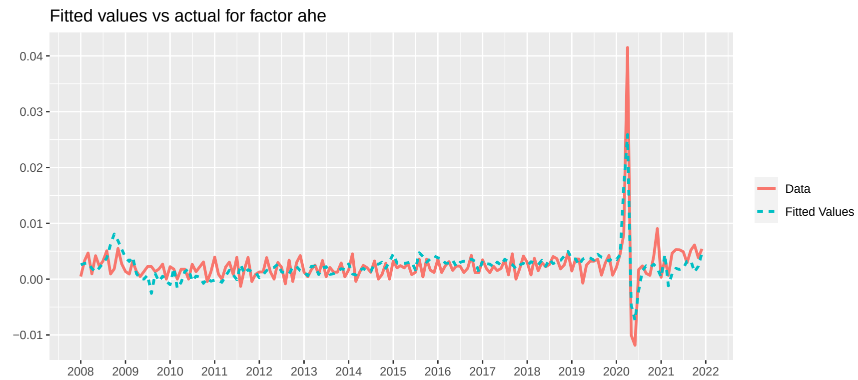


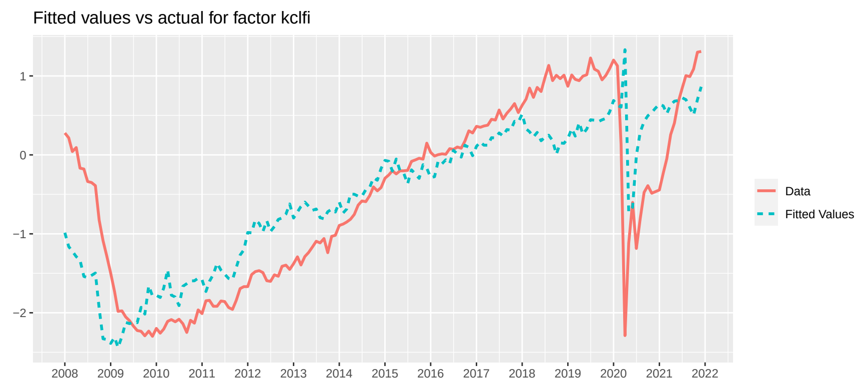

Fitted values vs actual for factor ppi
0.02
0.00
−0.02
−0.04
−0.06
2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022
Data
Fitted Values

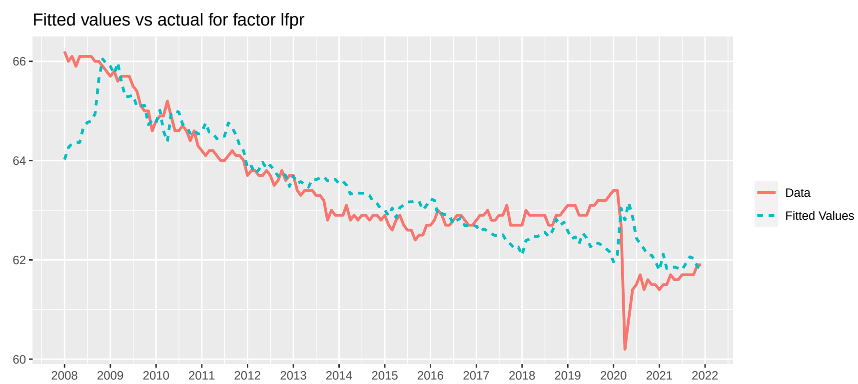

Fitted values vs actual for factor unemp
2.5
2.0
1.5
2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022
Data
Fitted Values

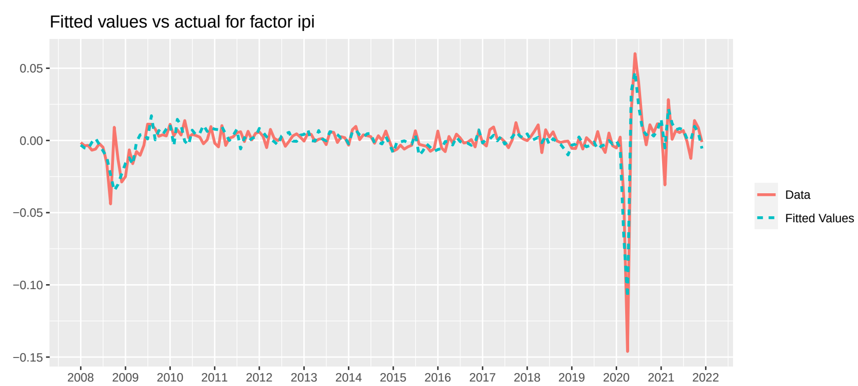

Fitted values vs actual for factor jolts
0.1
0.0
−0.1
−0.2
2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022
Data
Fitted Values


Fitted values vs actual for factor vacancies
10
0
−10
−20
2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022
Data
Fitted Values

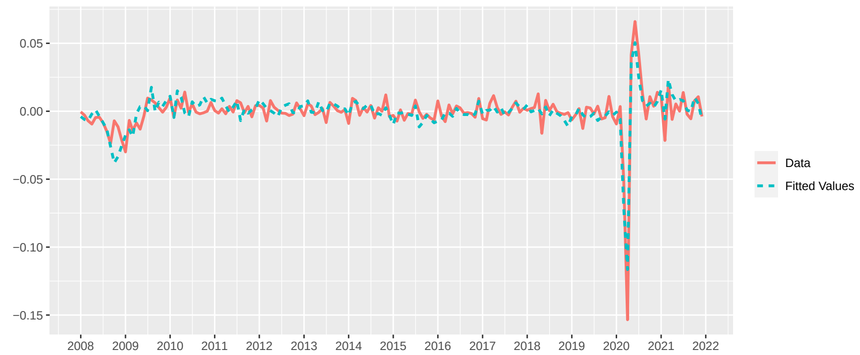

Fitted values vs actual for factor ahe
Data
Fitted Values


Fitted values vs actual for factor kclfi
Data
Fitted Values

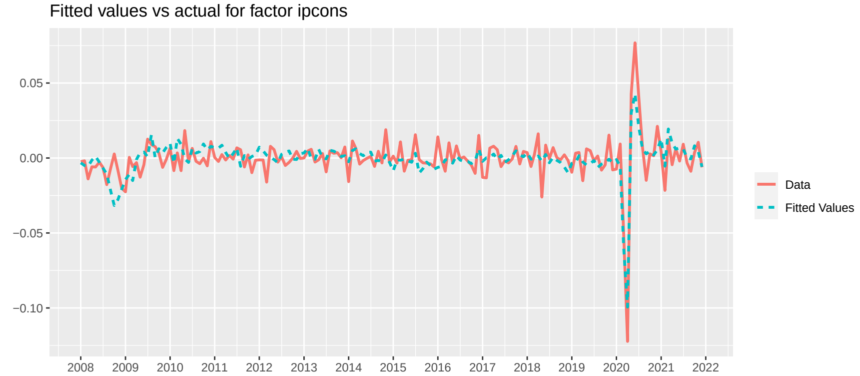

Fitted values vs actual for factor lfpr
Data
Fitted Values


Fitted values vs actual for factor ipi
Data
Fitted Values

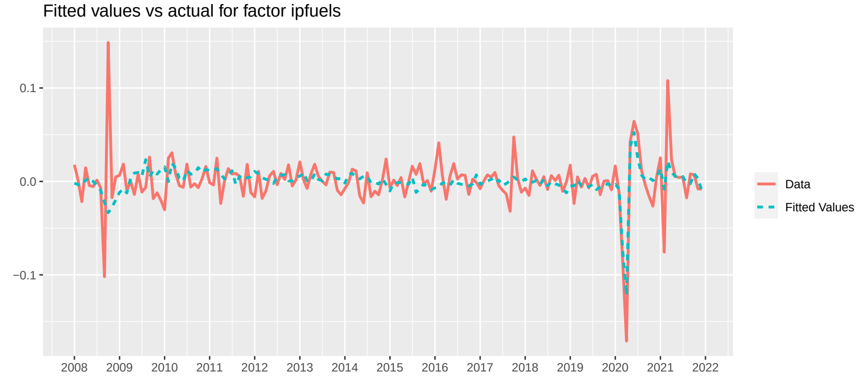
Fitted values vs actual for factor ipfinal



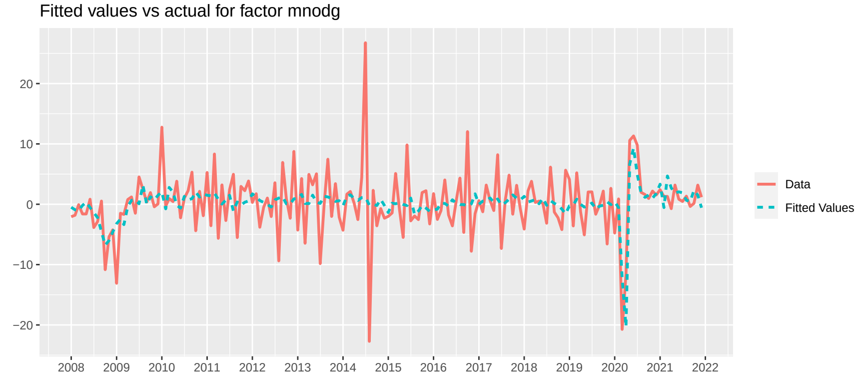
Fitted values vs actual for factor ipcons

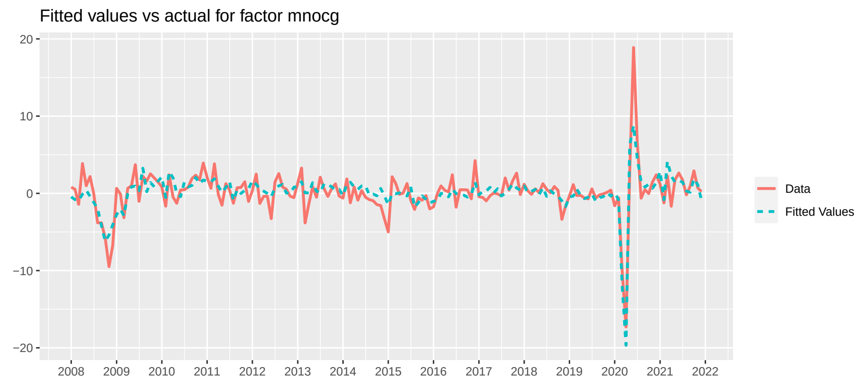


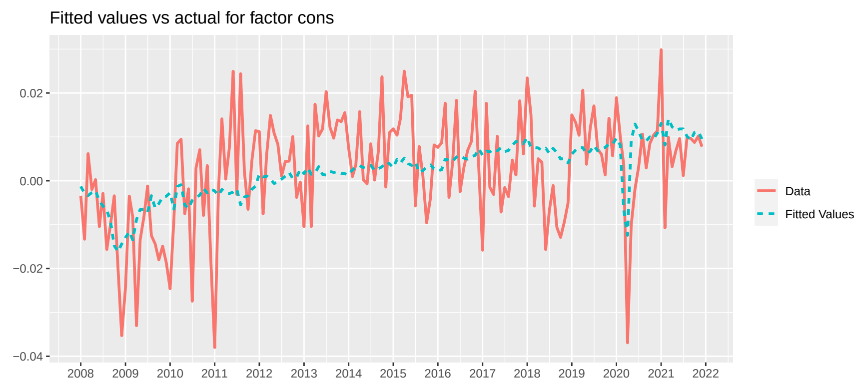
Fitted values vs actual for factor ipfuels

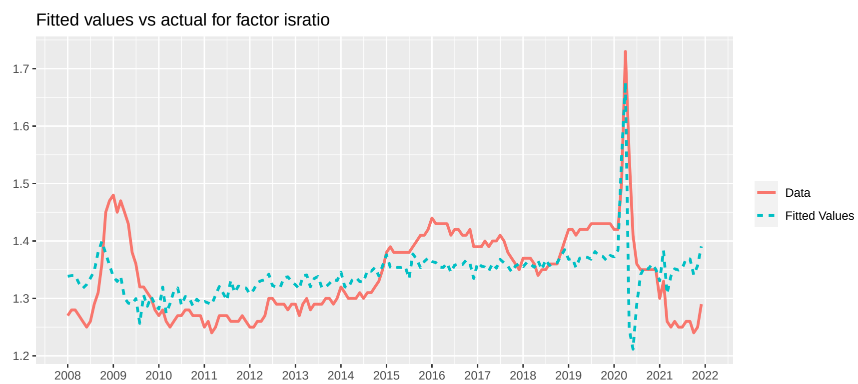


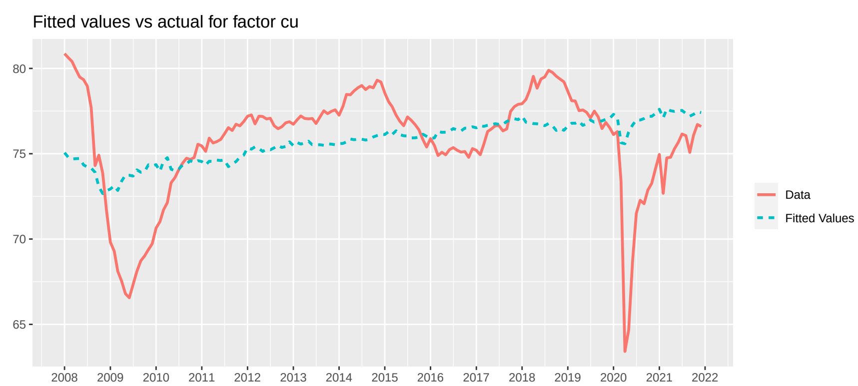
Fitted values vs actual for factor mnodg



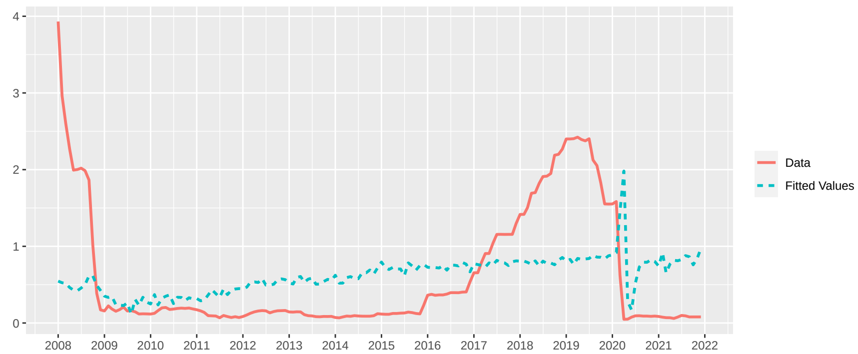

Fitted values vs actual for factor mnocg
20
10
0
−10
−20
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
Data
Fitted Values


Fitted values vs actual for factor cons
0.02
0.00
−0.02
−0.04
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
Data
Fitted Values

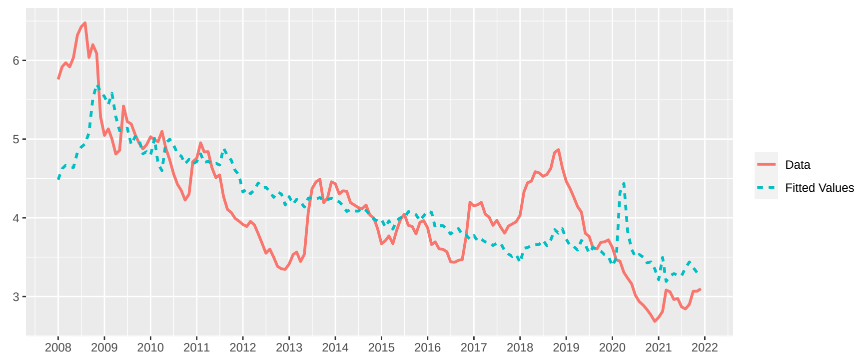

Fitted values vs actual for factor isratio
1.7
1.6
1.5
1.4
1.3
1.2
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
Data
Fitted Values


Fitted values vs actual for factor cu
80
75
70
65
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
Data
Fitted Values

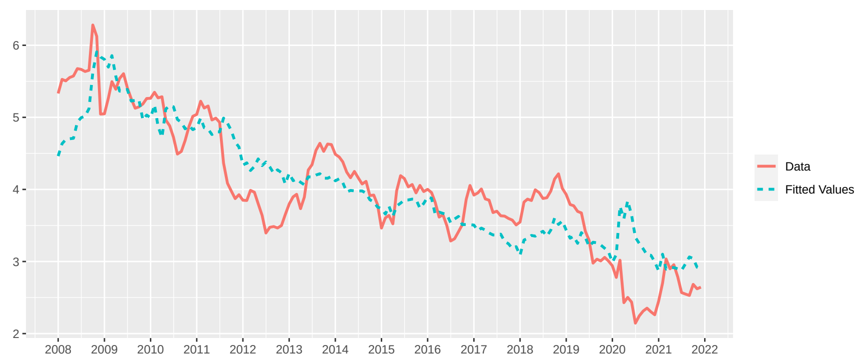
Fitted values vs actual for factor ffr



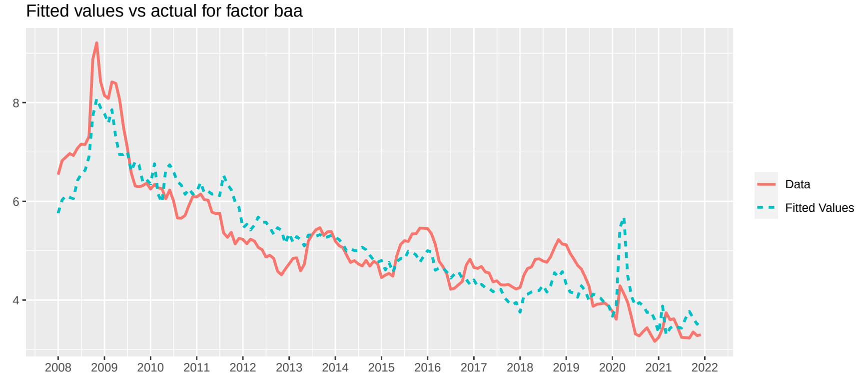
Fitted values vs actual for factor mort30y



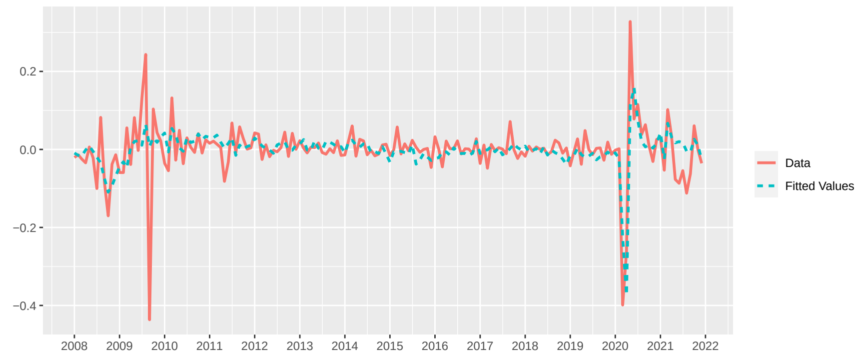
Fitted values vs actual for factor aaa



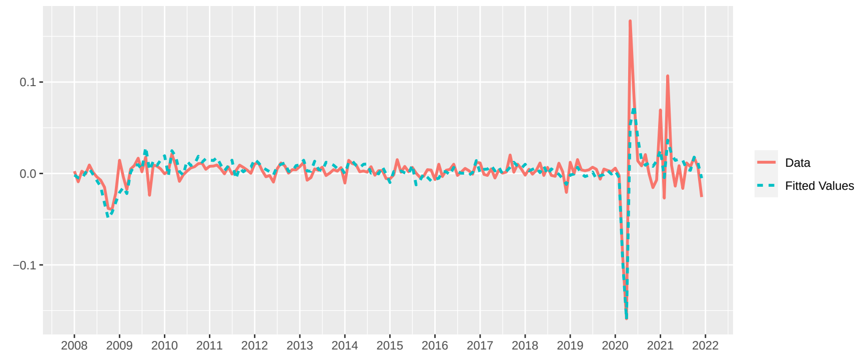
Fitted values vs actual for factor baa



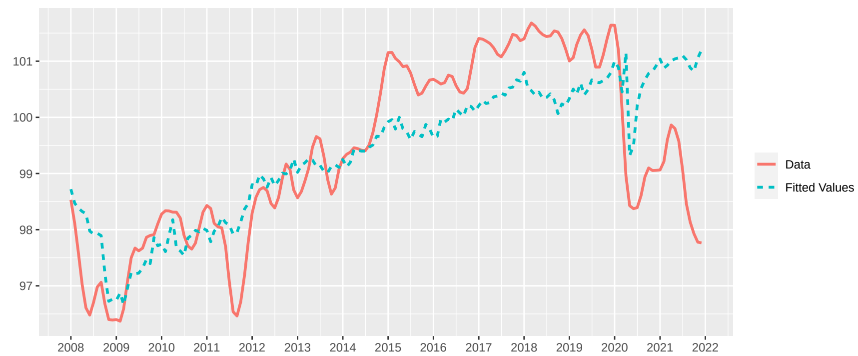
Fitted values vs actual for factor vsales



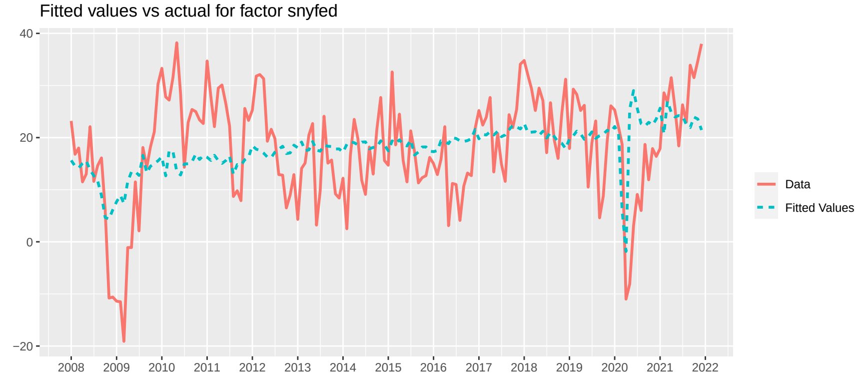
Fitted values vs actual for factor advsales



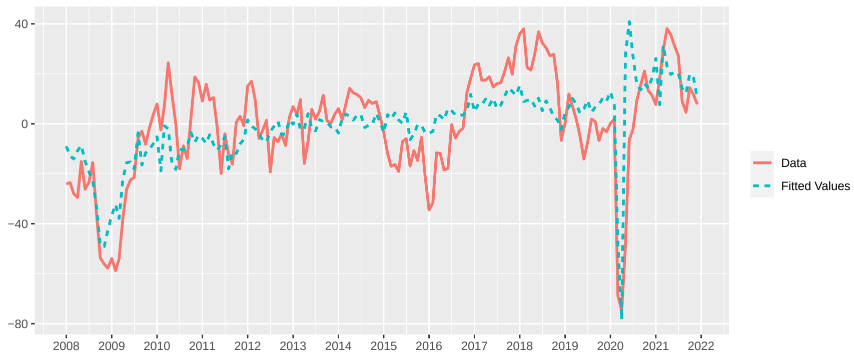
Fitted values vs actual for factor csent



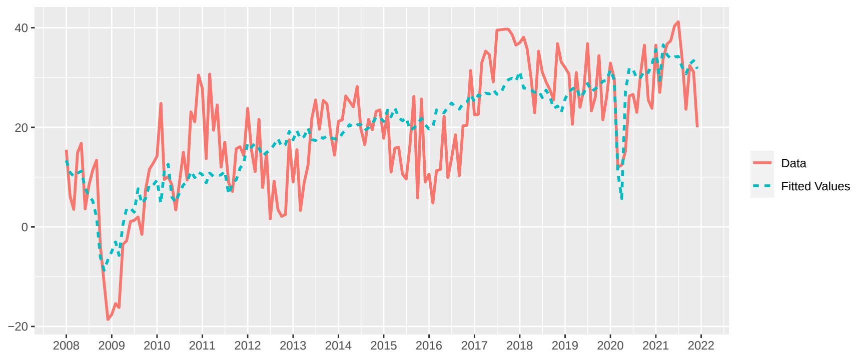
Fitted values vs actual for factor snyfed



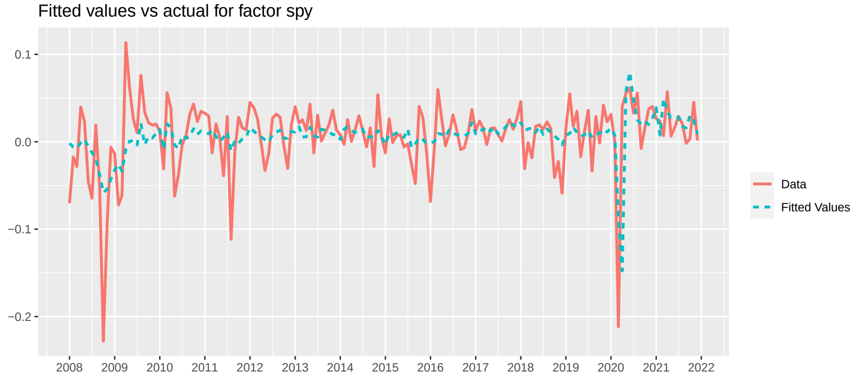
Fitted values vs actual for factor stxfed



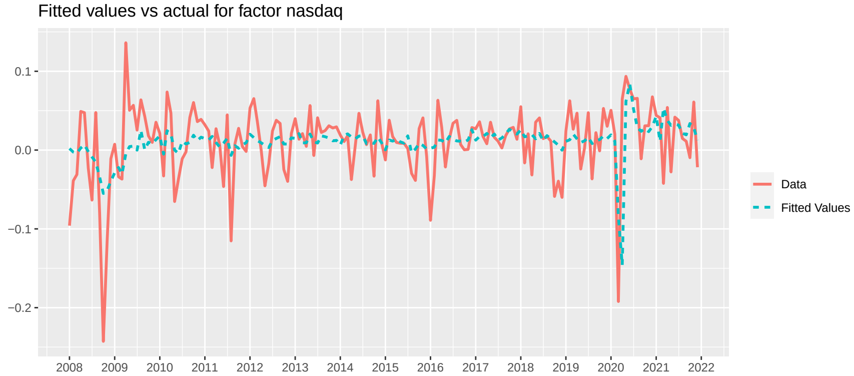
Fitted values vs actual for factor sphfed

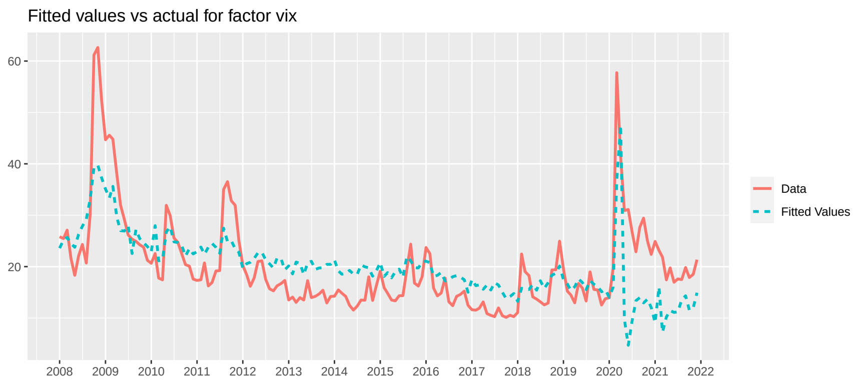


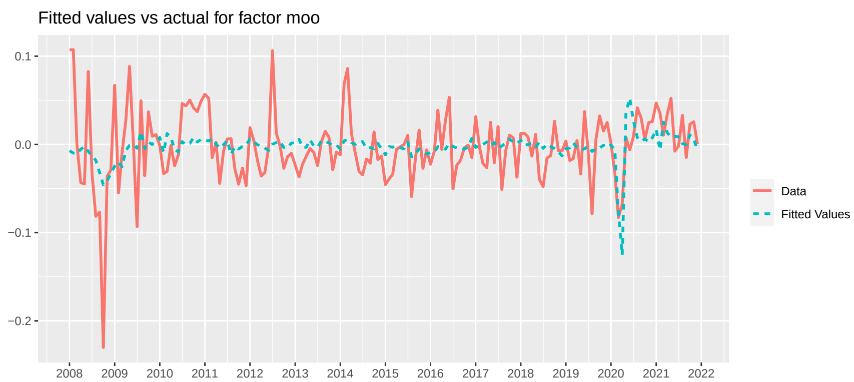
Fitted values vs actual for factor spy

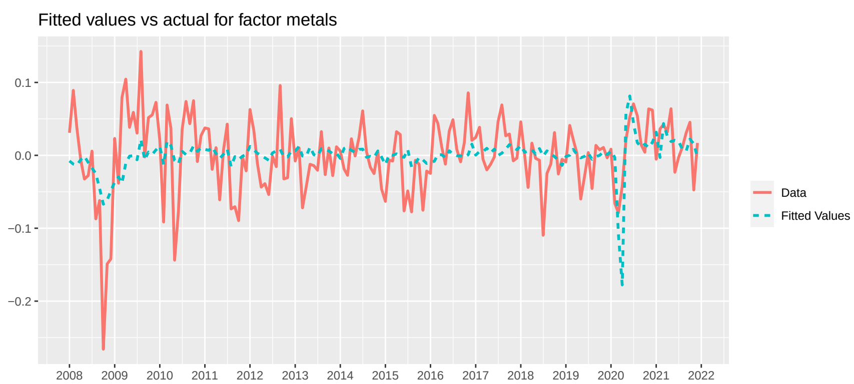


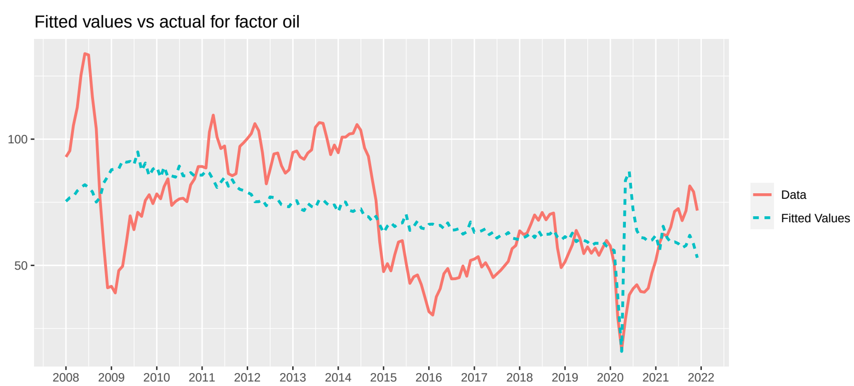
Fitted values vs actual for factor nasdaq

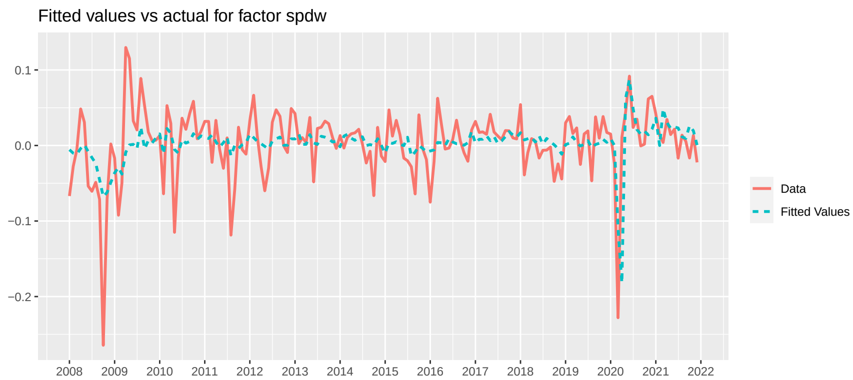



Fitted values vs actual for factor vix
60
40
20
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
Data
Fitted Values

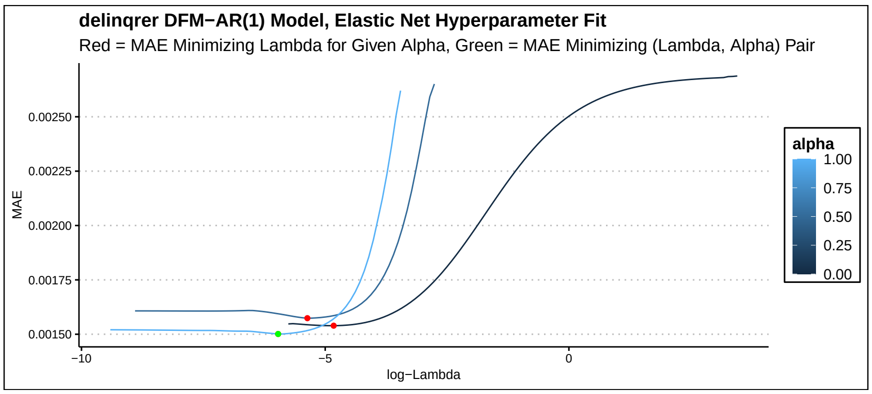

Fitted values vs actual for factor moo
0.1
0.0
−0.1
−0.2
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
Data
Fitted Values

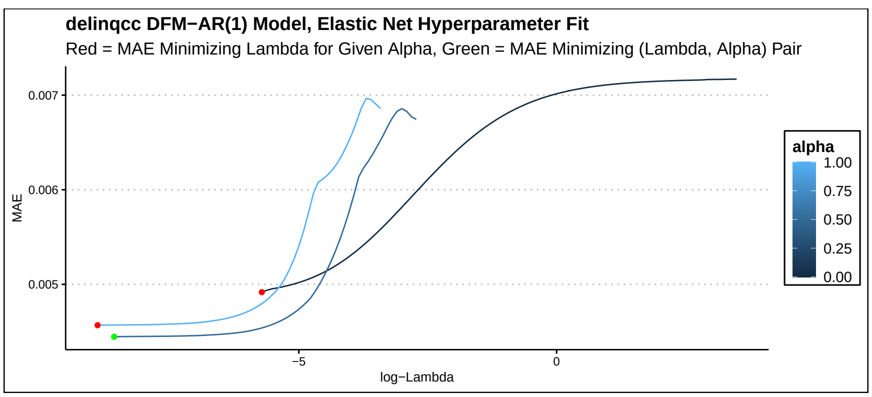

Fitted values vs actual for factor metals
0.1
0.0
−0.1
−0.2
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
Data
Fitted Values

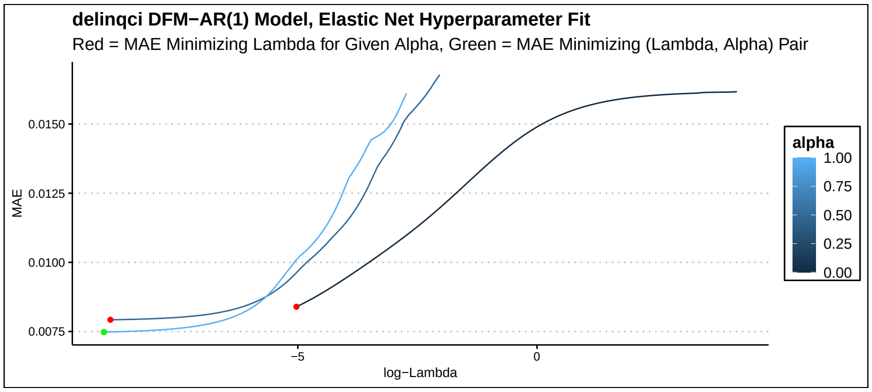

Fitted values vs actual for factor oil
100
50
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
Data
Fitted Values

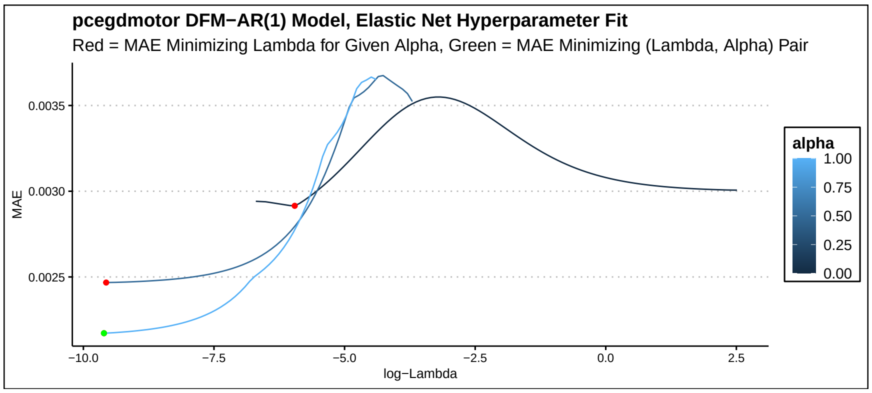


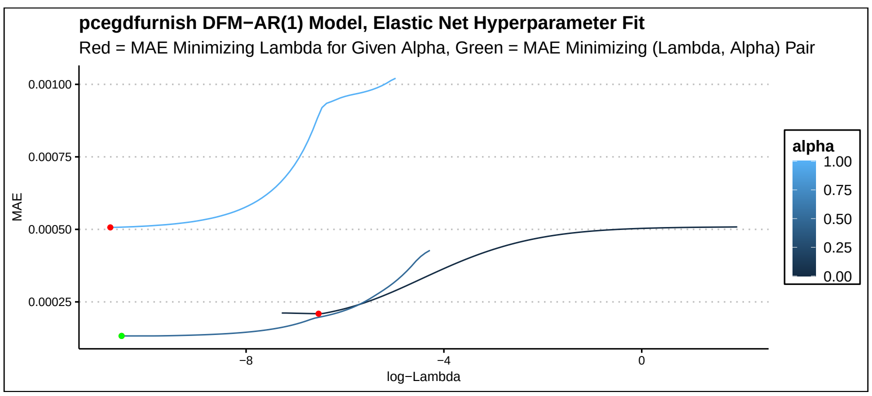
B Elastic Net Hyperparameter Selection

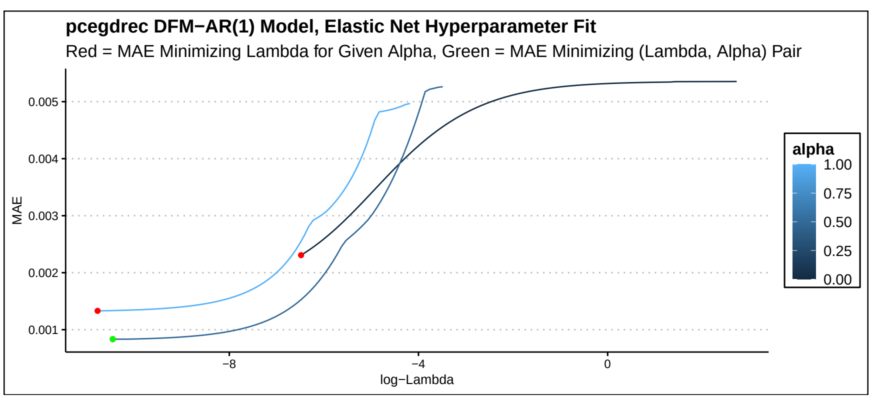


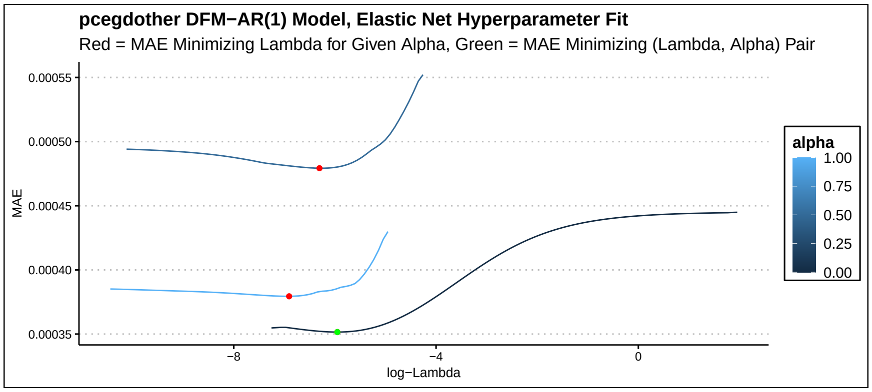


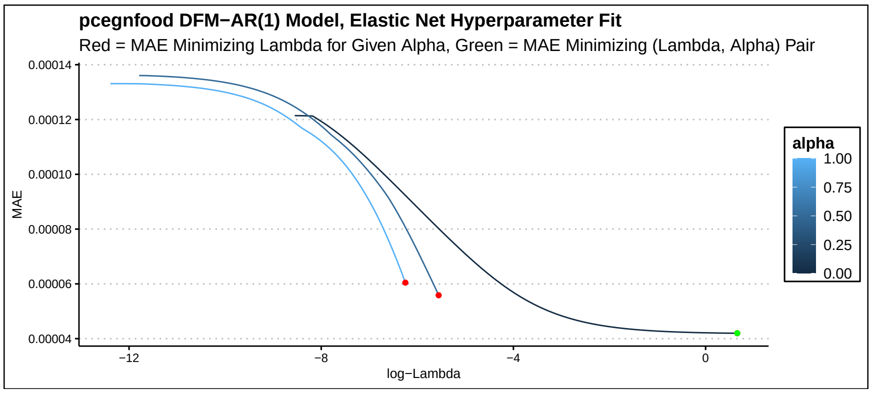

delinqci DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
MAE
0.0150
0.0125
0.0100
0.0075
−5
0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00

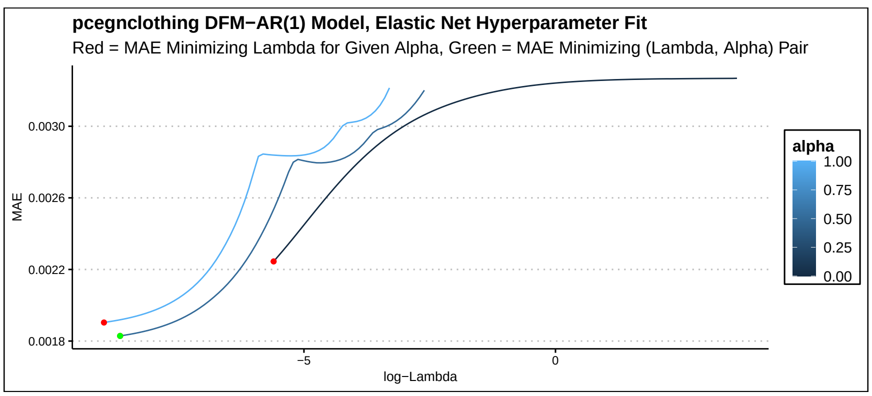

pcegdmotor DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
MAE
0.0035
0.0030
0.0025
−10.0
−7.5
−5.0
−2.5
0.0
2.5
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00

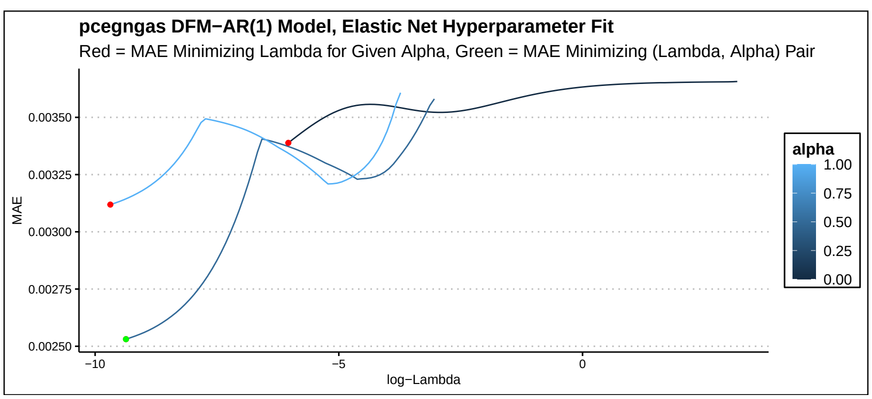

pcegdfurnish DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
MAE
0.00100
0.00075
0.00050
0.00025
−8
−4
0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00

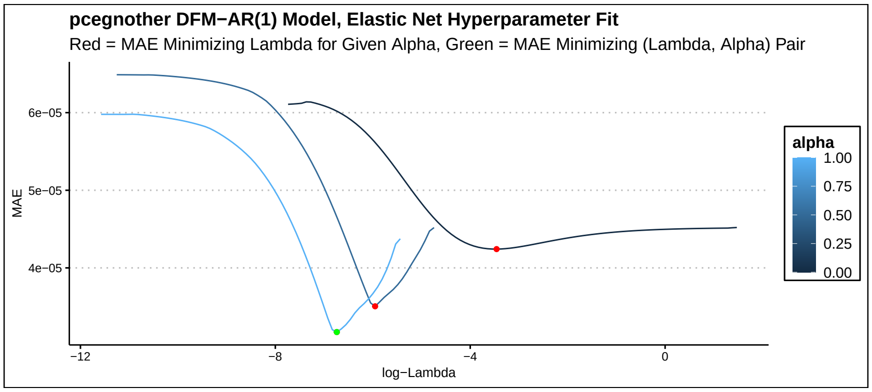

pcegdrec DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
MAE
0.005
0.004
0.003
0.002
0.001
−8
−4
0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00

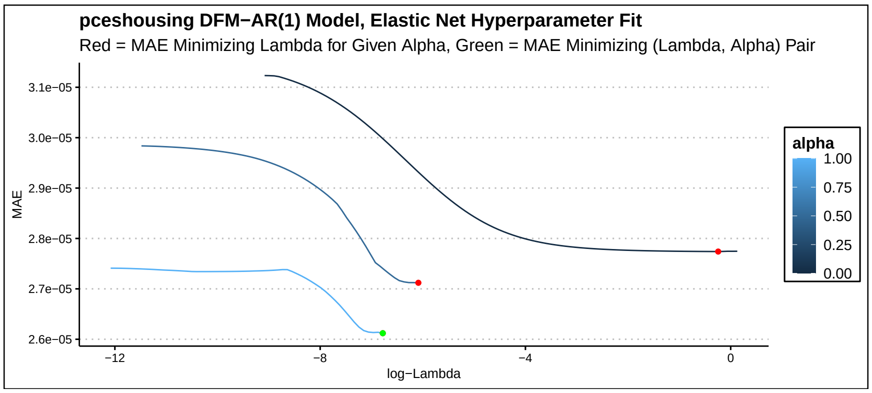

pcegdother DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
MAE
0.00055
0.00050
0.00045
0.00040
0.00035
−8
−4
0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00

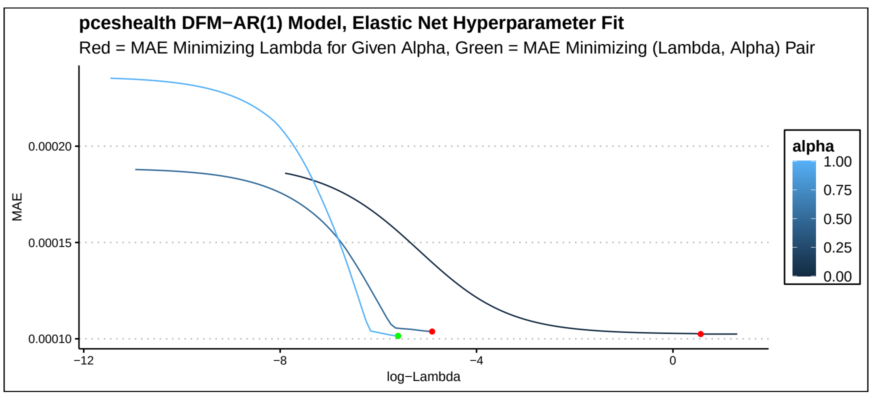

pcegnfood DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
MAE
0.00014
0.00012
0.00010
0.00008
0.00006
0.00004
−12
−8
−4
0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00

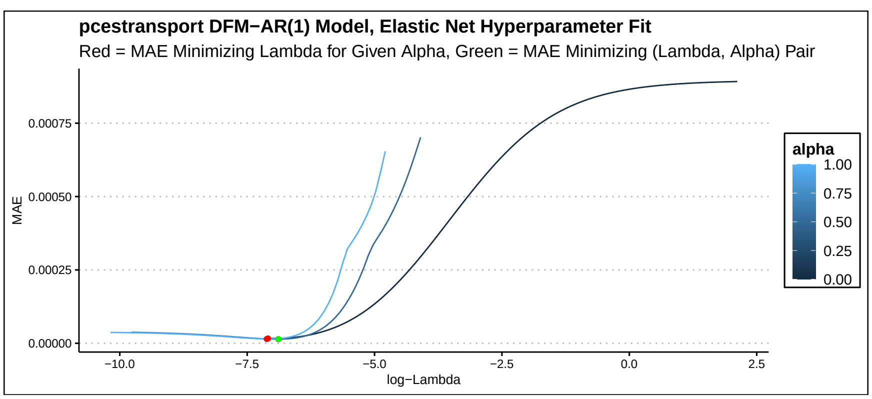

pcegnclothing DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
MAE
0.0030
0.0026
0.0022
0.0018
−5
0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00

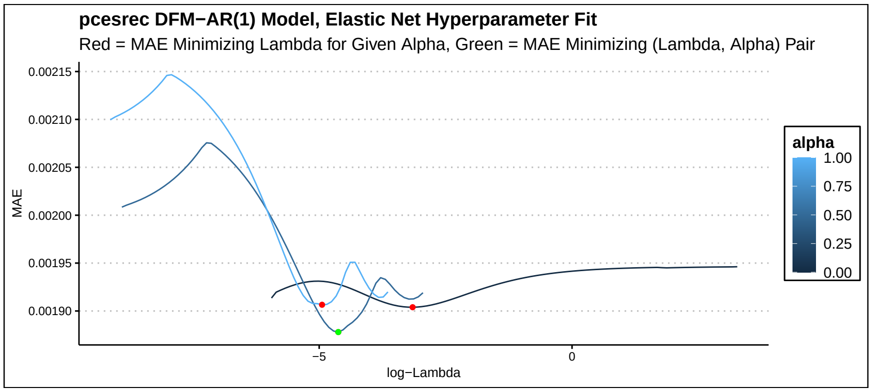

pcegngas DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
MAE
0.00350
0.00325
0.00300
0.00275
0.00250
−10
−5
0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00

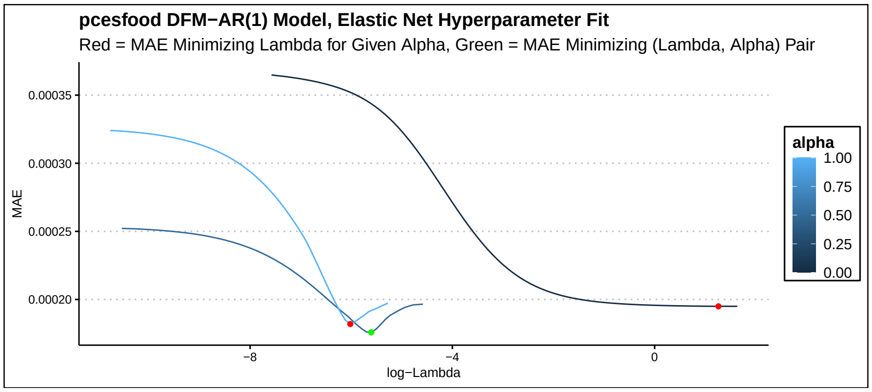

pcegnother DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
MAE
6e−05
5e−05
4e−05
−12
−8
−4
0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00

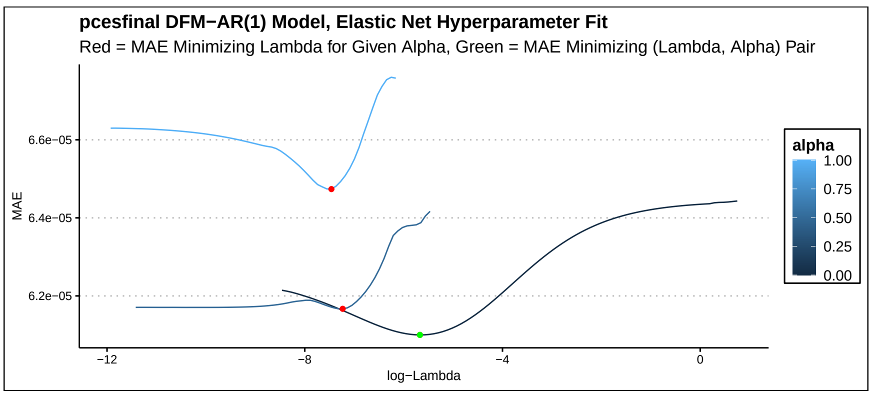

pceshousing DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
MAE
3.1e−05
3.0e−05
2.9e−05
2.8e−05
2.7e−05
2.6e−05
−12
−8
−4
0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00

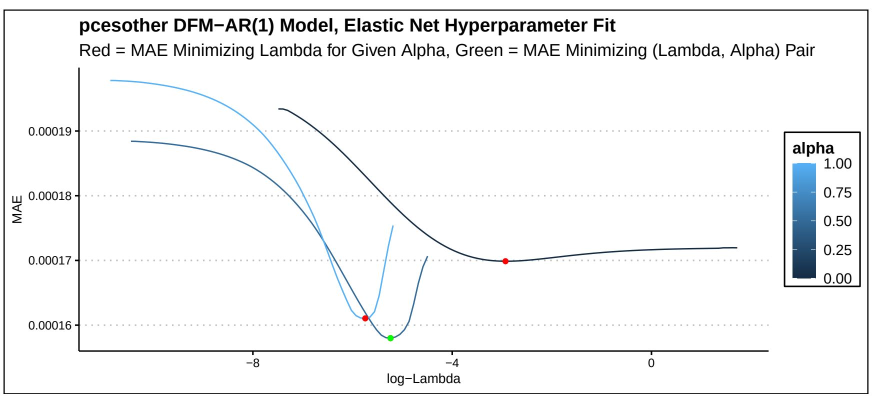

pceshealth DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
MAE
0.00020
0.00015
0.00010
−12
−8
−4
0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00

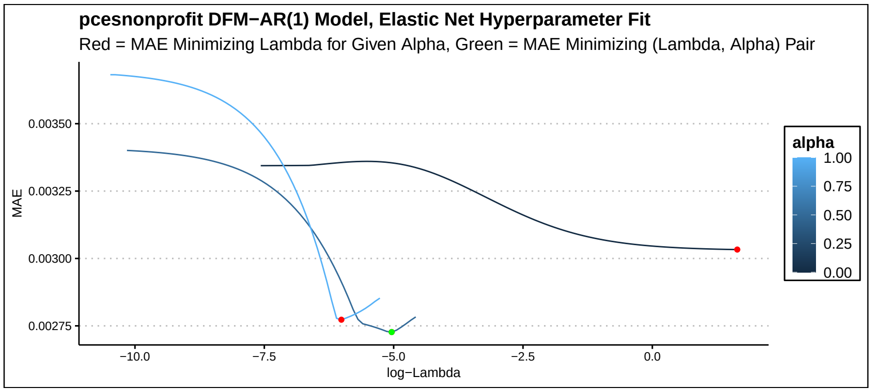

pcestransport DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
MAE
0.00075
0.00050
0.00025
0.00000
−10.0
−7.5
−5.0
−2.5
0.0
2.5
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00

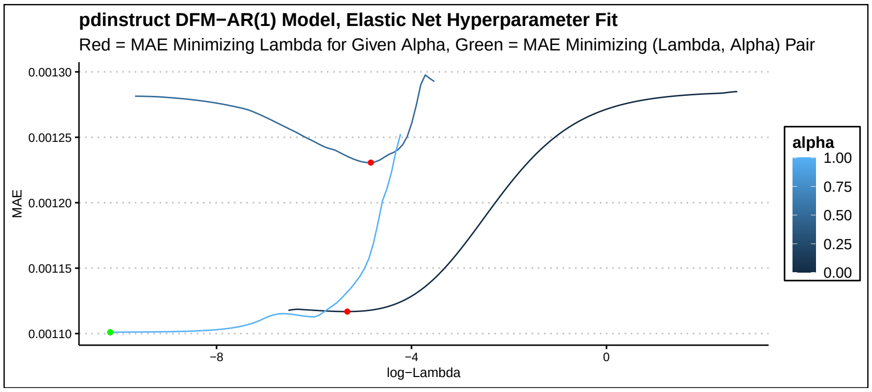

pcesrec DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
MAE
0.00215
0.00210
0.00205
0.00200
0.00195
0.00190
−5
0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00

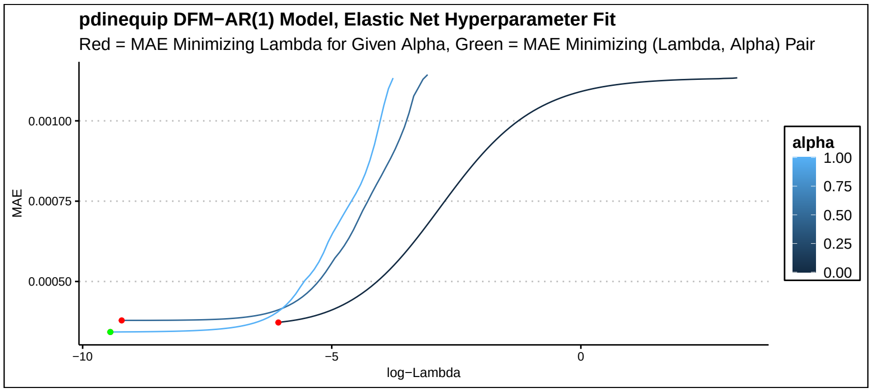

pcesfood DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
MAE
0.00035
0.00030
0.00025
0.00020
−8
−4
0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00

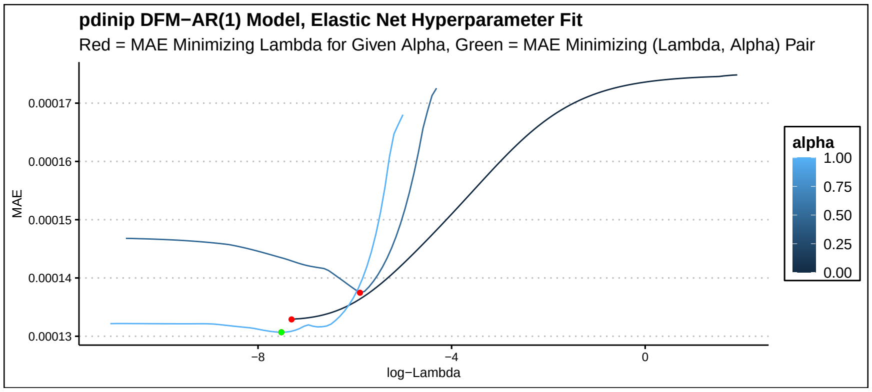

pcesfinal DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
MAE
6.6e−05
6.4e−05
6.2e−05
−12
−8
−4
0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00

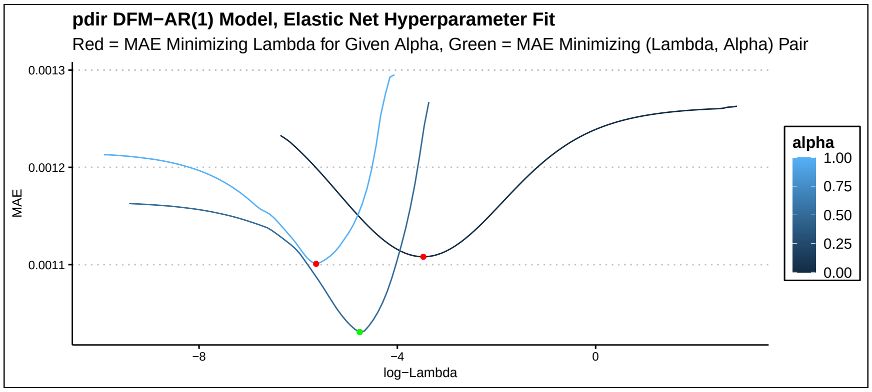

pcesother DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
MAE
0.00019
0.00018
0.00017
0.00016
−8
−4
0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00

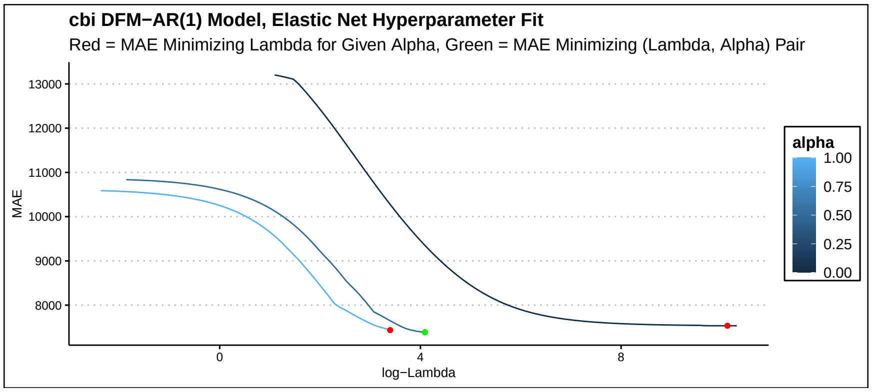

pcesnonprofit DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
MAE
0.00350
0.00325
0.00300
0.00275
−10.0
−7.5
−5.0
−2.5
0.0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00

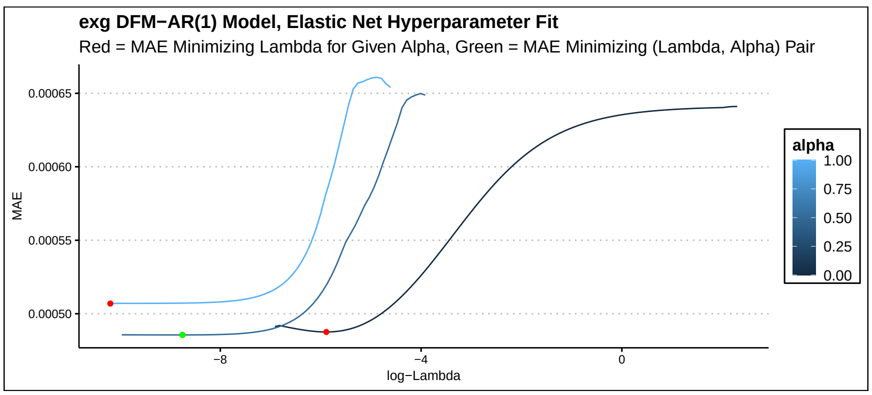

pdinstruct DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
MAE
0.00130
0.00125
0.00120
0.00115
0.00110
−8
−4
0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00

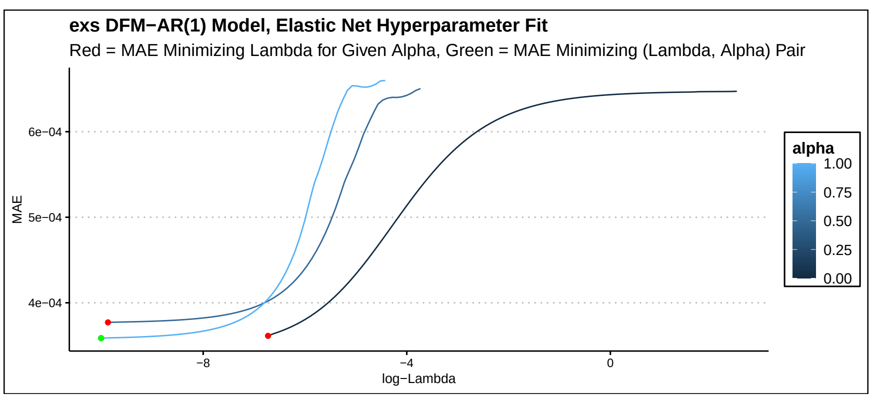

pdinequip DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
MAE
0.00100
0.00075
0.00050
−10
−5
0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00

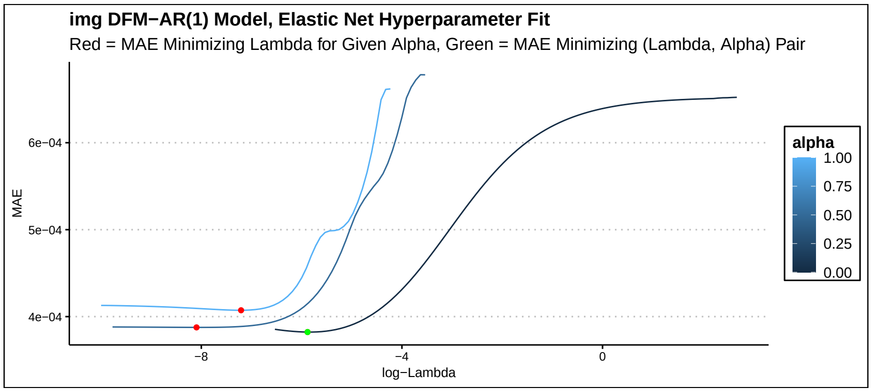

pdinip DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
MAE
0.00017
0.00016
0.00015
0.00014
0.00013
−8
−4
0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00

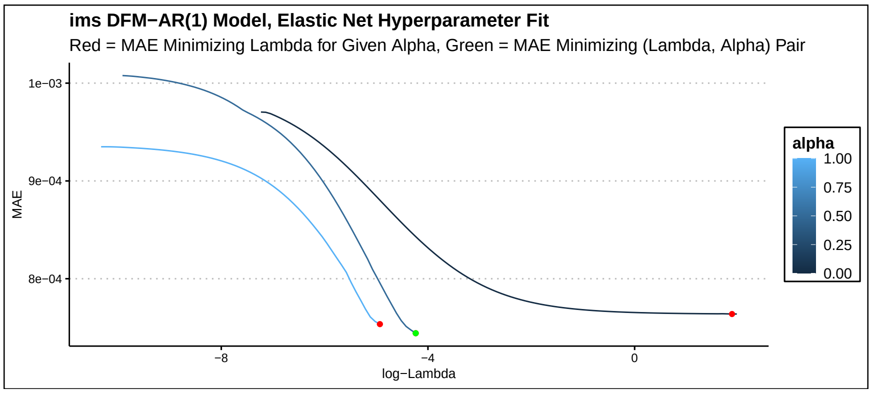

pdir DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
0.0013
0.0012
0.0011
MAE
−8
−4
0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00

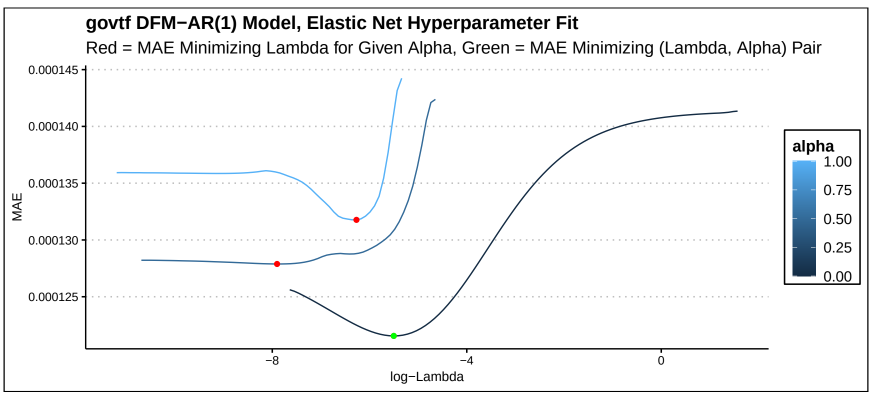

cbi DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
13000
12000
11000
10000
9000
8000
MAE
0
4
8
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00

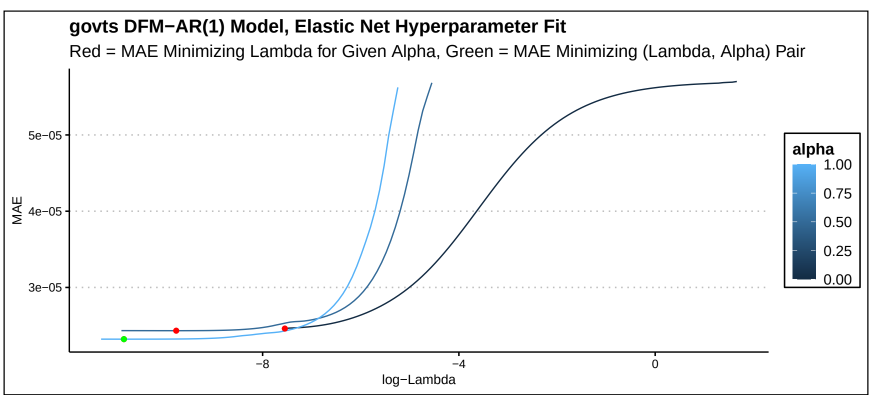

exg DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
0.00065
0.00060
0.00055
0.00050
MAE
−8
−4
0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00


exs DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
6e−04
5e−04
4e−04
MAE
−8
−4
0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00


img DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
MAE
6e−04
5e−04
4e−04
−8
−4
0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00


ims DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
MAE
1e−03
9e−04
8e−04
−8
−4
0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00


govtf DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
MAE
0.000145
0.000140
0.000135
0.000130
0.000125
−8
−4
0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00


govts DFM−AR(1) Model, Elastic Net Hyperparameter Fit
Red = MAE Minimizing Lambda for Given Alpha, Green = MAE Minimizing (Lambda, Alpha) Pair
MAE
5e−05
4e−05
3e−05
−8
−4
0
log−Lambda
alpha
1.00
0.75
0.50
0.25
0.00

C Backtested GDP Forecasts