



---

# A Nowcasting Model for Time Series with Ragged-Edge Data

Model Run Date: March 01, 2021  
Charles Y. — [charles@cmefi.com](mailto:charles@cmefi.com)

## Contents

<b>1</b>	<b>Motivation</b>	<b>2</b>
<b>2</b>	<b>Methodology</b>	<b>3</b>
2.1	Overview . . . . .	3
2.2	Data . . . . .	3
2.3	Time Periods . . . . .	4
2.4	Principal Components Analysis . . . . .	5
2.5	Factor VAR . . . . .	8
2.6	Dynamic Factor Models . . . . .	10
2.7	State-Space Setup . . . . .	12
2.8	Kalman Filtration . . . . .	13
2.9	Predicting Quarterly Variables . . . . .	15
<b>3</b>	<b>Results</b>	<b>18</b>
	<b>Appendices</b>	<b>19</b>
<b>A</b>	<b>DFM Fitted Plots</b>	<b>19</b>

# 1 Motivation

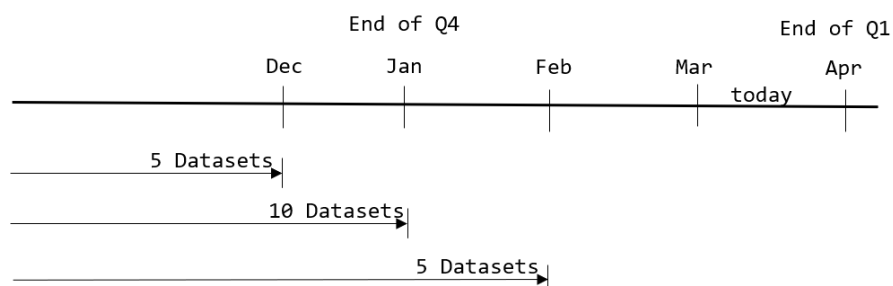
**Nowcasting** is the prediction of the present, the near future, and the near past. Nowcasting is important in economics because many important macroeconomic statistics are released with a lengthy delay. For example, the Bureau of Economic Analysis releases quarterly GDP typically two months after the 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, JP Morgan forecasts in March 13 predicted Q2 GDP growth of -3%; the estimate was revised down to -14% by March 21, -25% by March 25, and -40% by April 10. Yet many of the 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 be 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.

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.



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 are unappealing.

In this paper, we will 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

### 2.1 Overview

This paper utilizes a modified version of the two-step dynamic factor model utilized in Giannone et al (2008) and Doz et al (2011). In this model, we use principal components to extract a few fundamental factors describing the overall structure of the economy. These factors are then projected into the future using a vector autoregression; smoothing over missing input data is conducted by using the Kalman filter and smoother. Finally, we aggregate these up to a quarterly level to forecast our quarterly covariates, then use these to calculate our nowcast of GDP.

The model is run every weekday; this documentation will use actual data and estimates from the most recent model run to illustrate the procedure. The final nowcasted output is located in section 3.

### 2.2 Data

We begin by importing monthly data 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.

Variable	Stationary Form
PCE	dlog
PCE: Goods	dlog
PCE: Goods: Durable Goods	dlog
PCE: Goods: Nondurable Goods	dlog
PCE: Services	dlog
Disposable Personal Income	dlog
Case-Shiller Housing Price Index	dlog
Effective Federal Funds Rate	base
Treasury Yield 10Y	base
Treasury Yield 1Y	base
Vehicle Sales	base
Unemployment rate	dlog
Weekly Economic Index	base
S&P 500	dlog
VIX	base
Outstanding Real Estate Loans	dlog
Outstanding Credit Card Balances	dlog
St. Louis Fed Financial Stress Index	base
U.S. Dollar Index	base
Average Weeks Unemployed	base
Retail & Food Sales	dlog
KC Fed Labor Market Index	base
Consumer Confidence (OECD)	base
Industrial Production index	dlog
New York Fed: Business Conditions Manufacturing Survey	base
Current General Business Activity, Diffusion Index for Texas	base
Chicago Fed National Activity Index	base

Table 1: Imported Monthly Data

We additionally import quarterly data for GDP and its subcomponents.

Variable	Stationary Form
GDP	dlog
Consumption	dlog
Consumption: Goods	dlog
Consumption: Goods: Durable	dlog
Consumption: Goods: Durable: Motor Vehicles & Parts	dlog
Consumption: Goods: Durable: Furnishings & Durable Household Equipment	dlog
Consumption: Goods: Durable: Recreational Goods & Services	dlog
Consumption: Goods: Durable: Other Durable Goods	dlog
Consumption: Goods: Nondurable Goods	dlog
Consumption: Goods: Nondurable Goods: Food & Beverages for Off-Premises Use	dlog
Consumption: Goods: Nondurable Goods: Clothing	dlog
Consumption: Goods: Nondurable Goods: Gasoline & Energy Goods	dlog
Consumption: Goods: Nondurable Goods: Other Nondurable Goods	dlog
Consumption: Services	dlog
Consumption: Services: Housing & Utilities	dlog
Consumption: Services: Health Care	dlog
Consumption: Services: Transportation	dlog
Consumption: Services: Recreation	dlog
Consumption: Services: Food Services & Accomodations	dlog
Consumption: Services: Financial Services	dlog
Consumption: Services: Other Services	dlog
Consumption: Services: Nonprofit Services	dlog
Gross Private Domestic Investment	dlog
Gross Private Domestic Investment: Nonresidential Investment	dlog
Gross Private Domestic Investment: Nonresidential Investment: Structures	dlog
Gross Private Domestic Investment: Nonresidential Investment: Equipment	dlog
Gross Private Domestic Investment: Nonresidential Investment: Intellectual Property	dlog
Gross Private Domestic Investment: Residential Investment	dlog
Gross Private Domestic Investment: Change in Private Inventories	base
Net Exports	base
Exports	dlog
Exports: Exported Goods	dlog
Exports: Exported Services	dlog
Imports	dlog
Imports: Imported Goods	dlog
Imports: Imported Services	dlog
Government Spending	dlog
Government Spending: Federal	dlog
Government Spending: State & Local	dlog
Weekly Economic Index	base
Industrial Production index	dlog
Chicago Fed National Activity Index	base

Table 2: Imported Quarterly Data

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).

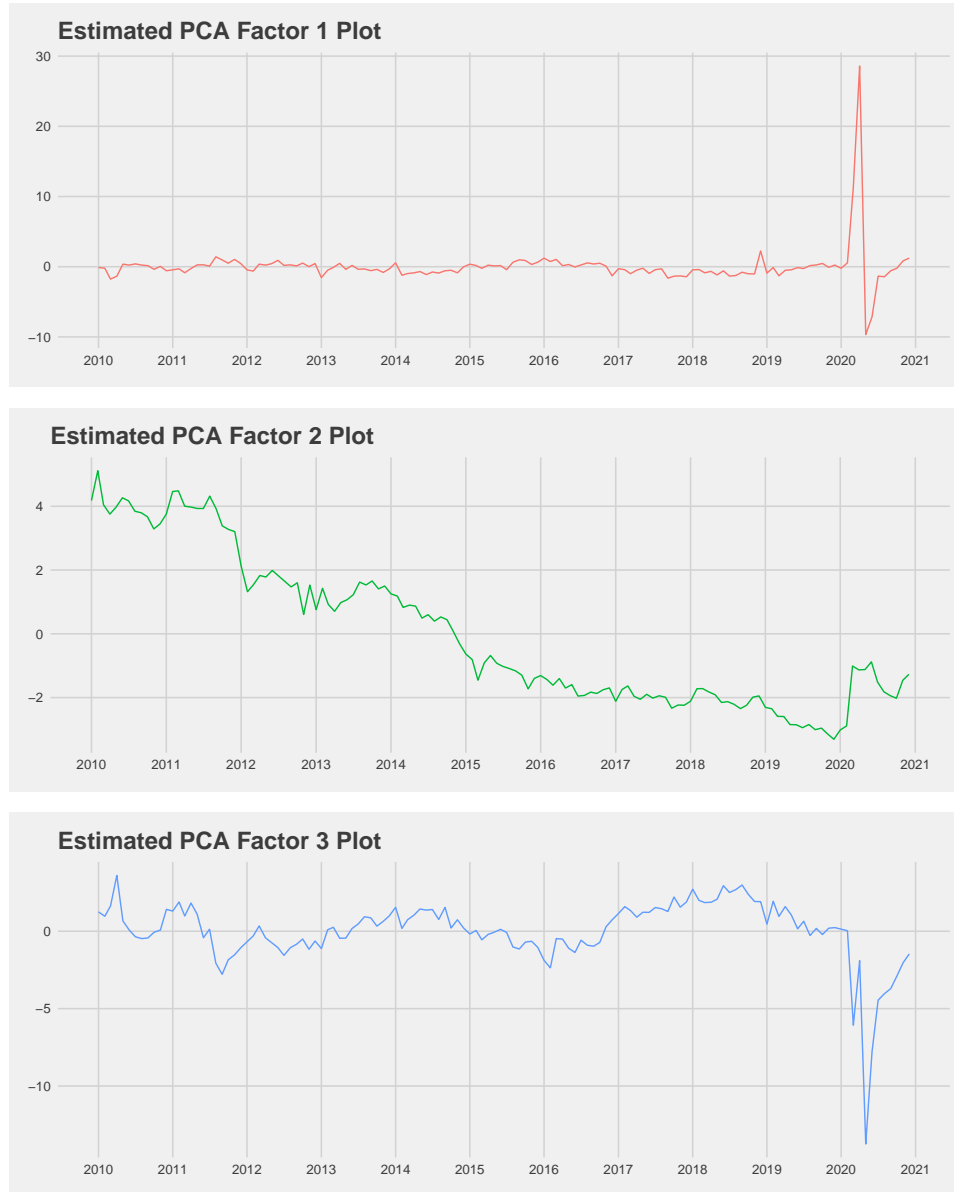
### 2.3 Time Periods

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.  $\tau$  will denote the number of dates for which data is available for at least one data series.  $T^*$  will denote



Once factors are derived, we perform a qualitative check of the factors. Typically the first factor should give us something similar to the growth rate of GDP or aggregate production, but on a monthly basis. The second and third 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.



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 include alternative specifications of the information criteria from Bai and Ng as a robustness check.

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

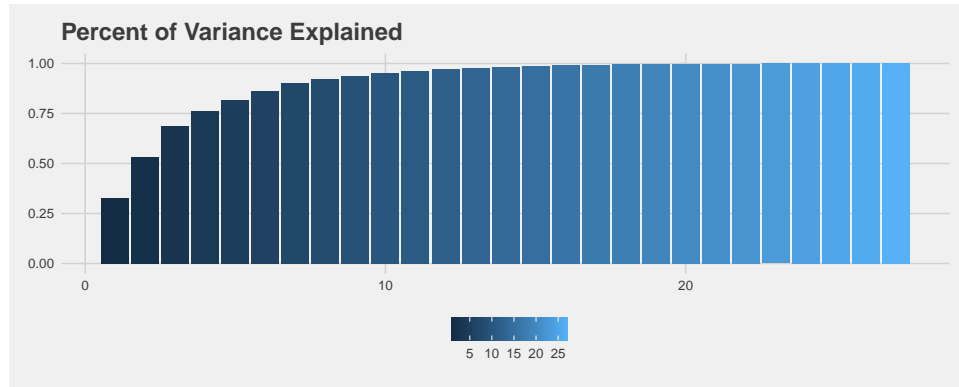


Figure 1: Factor Selection

Factors (R)	Variance Explained	Pct of Total Var Explained	Cumulative Pct	MSE	IC1	IC2	IC3
1	8.87	0.33	0.33	0.67	0.81	0.81	0.79
2	5.46	0.20	0.53	0.47	0.74	0.76	0.71
3	4.24	0.16	0.69	0.31	0.73	0.75	0.68
4	2.05	0.08	0.76	0.23	0.79	0.82	0.72
5	1.50	0.06	0.82	0.18	0.87	0.91	0.79
6	1.18	0.04	0.86	0.14	0.97	1.02	0.87
7	1.01	0.04	0.90	0.10	1.07	1.13	0.95
8	0.56	0.02	0.92	0.08	1.19	1.25	1.05
9	0.42	0.02	0.94	0.06	1.31	1.39	1.16
10	0.36	0.01	0.95	0.05	1.44	1.52	1.27
11	0.31	0.01	0.96	0.04	1.56	1.66	1.38
12	0.23	0.01	0.97	0.03	1.69	1.79	1.49
13	0.18	0.01	0.98	0.02	1.83	1.93	1.61
14	0.13	0.00	0.98	0.02	1.96	2.08	1.73
15	0.12	0.00	0.99	0.01	2.09	2.22	1.84
16	0.10	0.00	0.99	0.01	2.23	2.36	1.96
17	0.08	0.00	0.99	0.01	2.37	2.51	2.08
18	0.06	0.00	0.99	0.00	2.50	2.65	2.20
19	0.04	0.00	1.00	0.00	2.64	2.80	2.32
20	0.03	0.00	1.00	0.00	2.78	2.94	2.44
21	0.02	0.00	1.00	0.00	2.91	3.09	2.56
22	0.02	0.00	1.00	0.00	3.05	3.24	2.69
23	0.01	0.00	1.00	0.00	3.19	3.38	2.81
24	0.01	0.00	1.00	0.00	3.33	3.53	2.93
25	0.00	0.00	1.00	0.00	3.47	3.68	3.05
26	0.00	0.00	1.00	0.00	3.61	3.82	3.17
27	0.00	0.00	1.00	0.00	3.75	3.97	3.30

Table 4: Factor Selection Process

Choosing the IC-minimizing  $R$  lets us choose  $R = 3$  factors.



## 2.5 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).

$$\underbrace{\begin{bmatrix} \mathbf{f1}_t \\ \mathbf{f2}_t \\ \vdots \\ \mathbf{fR}_t \end{bmatrix}}_{z_t} = B \underbrace{\begin{bmatrix} \mathbf{f1}_{t-1} \\ \mathbf{f2}_{t-1} \\ \vdots \\ \mathbf{fR}_{t-1} \end{bmatrix}}_{z_{t-1}} + C + \underbrace{\begin{bmatrix} v1_t \\ v2_t \\ \vdots \\ vR_t \end{bmatrix}}_{v_t},$$

where  $z_t$  is the  $R \times 1$  matrix of time  $t$  factors,

$B$  is the  $R \times R$  coefficient matrix,

$C$  is the  $R \times 1$  constant matrix,

and  $v_t$  is the  $R \times 1$  matrix of errors for time  $t$ .

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,

$$\underbrace{\begin{bmatrix} \mathbf{f1}_2 & \mathbf{f2}_2 & \dots & \mathbf{f6}_2 \\ \mathbf{f1}_3 & \mathbf{f2}_3 & \dots & \mathbf{f6}_3 \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{f1}_T & \mathbf{f2}_T & \dots & \mathbf{f6}_T \end{bmatrix}}_{\Gamma} = \underbrace{\begin{bmatrix} 1 & \mathbf{f1}_1 & \mathbf{f2}_1 & \dots & \mathbf{f6}_1 \\ 1 & \mathbf{f1}_2 & \mathbf{f2}_2 & \dots & \mathbf{f6}_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \mathbf{f1}_{T-1} & \mathbf{f2}_{T-1} & \dots & \mathbf{f6}_{T-1} \end{bmatrix}}_{\Psi} \underbrace{\begin{bmatrix} C' \\ B' \end{bmatrix}}_{\Lambda} + \underbrace{\begin{bmatrix} v1_2 & v2_2 & \dots & vR_2 \\ v1_3 & v2_3 & \dots & vR_3 \\ \vdots & \vdots & \vdots & \vdots \\ v1_T & v2_T & \dots & vR_T \end{bmatrix}}_V,$$

where  $\Gamma$  is the  $T - 1 \times R$  dependent data matrix,

$\Psi$  is the  $T - 1 \times R + 1$  independent data matrix,

$\Lambda$  is the  $R + 1 \times R$  matrix of coefficient weightings,

and  $V$  is the  $T - 1 \times R$  matrix of residuals.

The coefficient matrix  $\Lambda$  can be estimated by the standard OLS estimator.

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

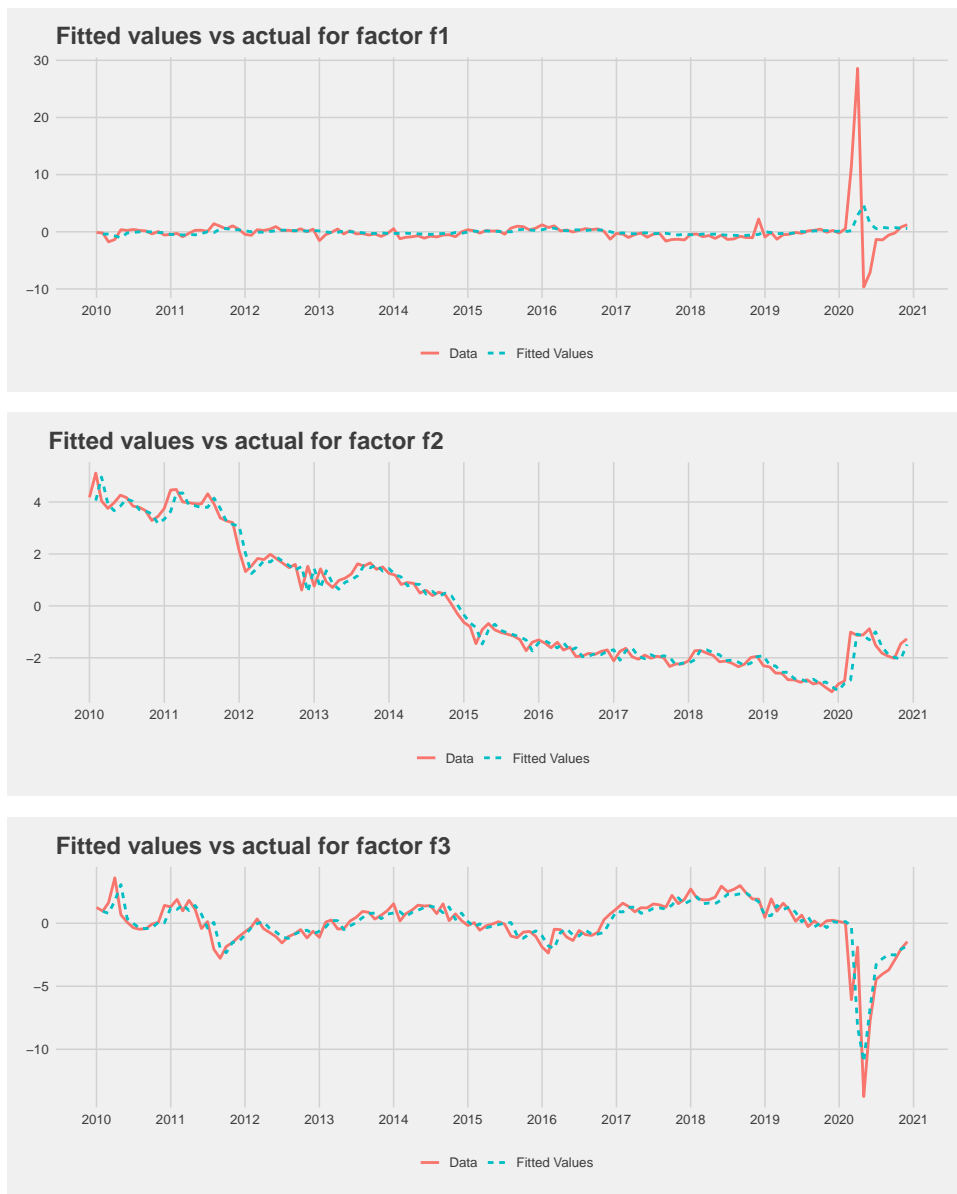
It can then be partitioned to calculate  $\hat{B}'$  and  $\hat{C}'$ , which can then be transposed to derive our estimates of the original coefficient matrices  $B$  and  $C$ ,  $\hat{B}$  and  $\hat{C}$ .

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

name	constant	f1.l1	f2.l1	f3.l1
f1	0.00	0.15	-0.03	-0.20
f2	-0.04	0.00	0.98	0.01
f3	-0.02	-0.34	0.01	0.73

Table 5: Factor VAR Coefficients

Finally, we perform a qualitative check of the fitted values and residuals.



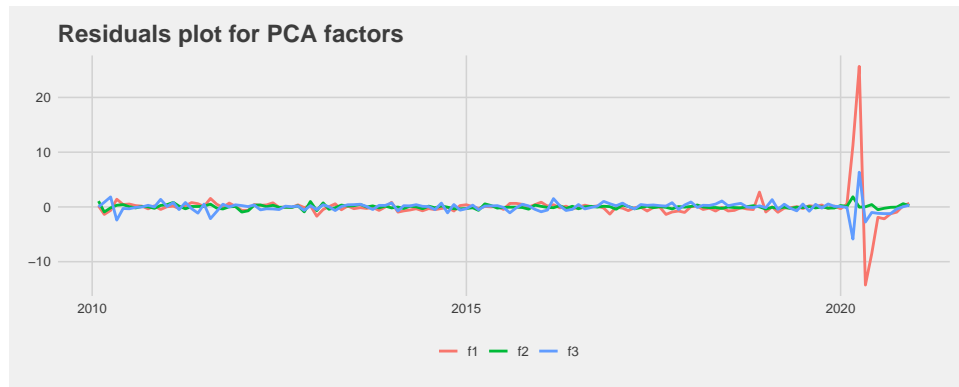


Figure 2: Factor VAR Residuals

Goodness-of-fit statistics are shown below.

varname	MAE	MSE
f1	0.90292071	8.50502097
f2	0.24907981	0.13092583
f3	0.57784082	1.01829474

Table 6: DFM Goodness of Fit

## 2.6 Dynamic Factor Models

Now let us consider again the monthly covariates 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.

The factor models take the following form:

$$\underbrace{\begin{bmatrix} \mathbf{ue}_t \\ \mathbf{pce}_t \\ \vdots \\ \mathbf{claims}_t \end{bmatrix}}_{y_t} = A \underbrace{\begin{bmatrix} \mathbf{f1}_t \\ \mathbf{f2}_t \\ \vdots \\ \mathbf{fR}_t \end{bmatrix}}_{z_t} + D + \underbrace{\begin{bmatrix} w1_t \\ w2_t \\ \vdots \\ wN_t \end{bmatrix}}_{w_t},$$

where  $y_t$  is the  $N \times 1$  vector of monthly variables at time  $t$ ,

$A$  is the  $N \times R$  coefficient matrix,

$z_t$  is the  $R \times 1$  vector of factors at time  $t$ ,

$D$  is the  $N \times 1$  constant matrix,

and  $w_t$  is the  $N \times 1$  vector of errors at time  $t$ .

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

$$\underbrace{\begin{bmatrix} \mathbf{ue}_2 & \mathbf{pce}_2 & \dots & \mathbf{claims}_2 \\ \mathbf{ue}_3 & \mathbf{pce}_3 & \dots & \mathbf{claims}_3 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{ue}_T & \mathbf{pce}_T & \dots & \mathbf{claims}_T \end{bmatrix}}_{\Phi} = \underbrace{\begin{bmatrix} 1 & \mathbf{f1}_2 & \mathbf{f2}_2 & \dots & \mathbf{f6}_2 \\ 1 & \mathbf{f1}_3 & \mathbf{f2}_3 & \dots & \mathbf{f6}_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \mathbf{f1}_T & \mathbf{f2}_T & \dots & \mathbf{f6}_T \end{bmatrix}}_{\Omega} \begin{bmatrix} D' \\ A' \end{bmatrix} + \underbrace{\begin{bmatrix} w1_2 & w2_2 & \dots & wR_2 \\ w1_3 & w2_3 & \dots & wR_3 \\ \vdots & \vdots & \ddots & \vdots \\ w1_T & w2_T & \dots & wR_T \end{bmatrix}}_W \quad (1)$$

As before we can estimate  $A$  and  $D$  with the standard OLS estimator.

$$\begin{bmatrix} \hat{A} \\ \hat{D} \end{bmatrix} = (\Omega' \Omega)^{-1} (\Omega' \Phi)$$

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

name	constant	f1	f2	f3
mpce	0.00	-0.01	-0.00	-0.00
mpceg	0.00	-0.01	-0.00	-0.00
mpcegd	0.01	-0.01	-0.00	-0.01
mpcegn	0.00	-0.00	-0.00	-0.00
mpces	0.00	-0.01	0.00	-0.00
pid	0.00	0.00	0.00	0.00
hpi	0.00	0.00	-0.00	-0.00
ffr	0.59	-0.02	-0.20	0.14
t10y	2.27	-0.05	0.13	0.24
t01y	2.27	-0.05	0.13	0.24
vsales	15.97	-0.19	-0.77	0.27
ue	-0.00	0.03	-0.00	0.00
wei	1.62	-0.19	0.29	0.86
spy	0.01	-0.00	-0.00	0.00
vix	17.97	0.99	0.85	-2.02
loansre	0.00	0.00	-0.00	0.00
loanscc	0.01	-0.00	0.00	0.01
stlfsi	-0.26	0.11	0.03	-0.14
usd	103.78	0.36	-4.22	-0.93
uemean	29.67	-0.36	2.45	0.60
advsales	0.00	-0.01	0.00	-0.00
kclfi	-0.48	-0.04	-0.42	0.12
csent	99.78	-0.02	-0.54	0.24
ipi	0.00	-0.00	0.00	-0.00
sny	6.18	-2.33	0.32	4.09
stx	2.41	-2.81	-0.61	5.39
cfnai	-0.07	-0.57	0.00	-0.11

Table 7: 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
advsales	0.00430707	0.00005898
cfnai	0.28580822	0.21848328
csent	0.34812380	0.21615133
ffr	0.41858659	0.25395389
hpi	0.00278684	0.00001420
ipi	0.00454008	0.00004977
kclfi	0.21296417	0.08280090
loanscc	0.01593077	0.00222734
loansre	0.00194009	0.00000643
mpce	0.00172203	0.00000874
mpceg	0.00395408	0.00004021
mpcegd	0.00833818	0.00024282
mpcegn	0.00440096	0.00007556
mpces	0.00284379	0.00002649
pid	0.00454171	0.00007823
sny	6.26255117	64.58654599
spy	0.02227264	0.00106228
stlfsi	0.24408582	0.11816037
stx	8.24467958	110.67871487
t01y	0.26668411	0.11444823
t10y	0.26668411	0.11444823
ue	0.03144553	0.00227886
uemean	3.32119293	17.27072698
usd	3.42148149	19.54427327
vix	3.20970152	17.53893033
vsales	0.84270626	0.99521910
wei	0.96199804	1.70988155

Table 8: DFM Goodness of Fit

## 2.7 State-Space Setup

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

$$\begin{aligned}
z_t &= Bz_{t-1} + Cx + v_t \\
y_t &= Az_t + w_t
\end{aligned}$$

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  $B$ ,  $C$ , and  $A$  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  $\tau$ . Elements in  $y_t$  may be set to any value for missing observations; the process of Kalman filtration will render this choice irrelevant.

Specifically, we construct the matrices below.

$$z_t = \begin{bmatrix} \mathbf{f1}_t \\ \mathbf{f2}_t \\ \vdots \\ \mathbf{fR}_t \end{bmatrix}, \forall t \in 1, \dots, T$$

$$y_t = \begin{bmatrix} \mathbf{ue}_t \text{ if available, otherwise } 0 \\ \mathbf{pce}_t \text{ if available, otherwise } 0 \\ \vdots \\ \mathbf{claims}_t \text{ if available, otherwise } 0 \end{bmatrix}, \forall t \in 1, \dots, \tau$$

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  $t \in 1, \dots, T+1$ , we let  $R_t$  be a diagonal covariance matrix with diagonal entries calculated by taking the average squared values of the residuals of the DFM. For  $t \in T+1, \dots, \tau$ , 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.

$$v_t \sim \mathcal{N}(0, Q)$$

$$w_t \sim \mathcal{N}(0, R_t)$$

We begin with the unconditional mean of  $\hat{Z}_{0|-1} = 0$  and unconditional variance of VAR  $\Sigma_{0|-1} = 0$ .

## 2.8 Kalman Filtration

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

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

$$y_t = Az_t + Dx + w_t$$

$$v_t \sim \mathcal{N}(0, Q)$$

$$w_t \sim \mathcal{N}(0, R_t)$$

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  $\tau$ ; the covariance matrix  $Q$ ; and finally, the covariance matrices  $R_t$  from 1 through  $\tau$ .

We initialize the Kalman filter with the following standard assumptions.

$$\mathbf{z}_{0|0} = 0$$

$$\mathbf{CovZ} = 0$$

Now for  $t = 1, \dots, \tau$ , 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  $T + 1, \dots, \tau$ , 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, can be 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  $\tau$ ,  $z_{t|\tau}$ . This can be recovered by using the Kalman smoother.

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

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

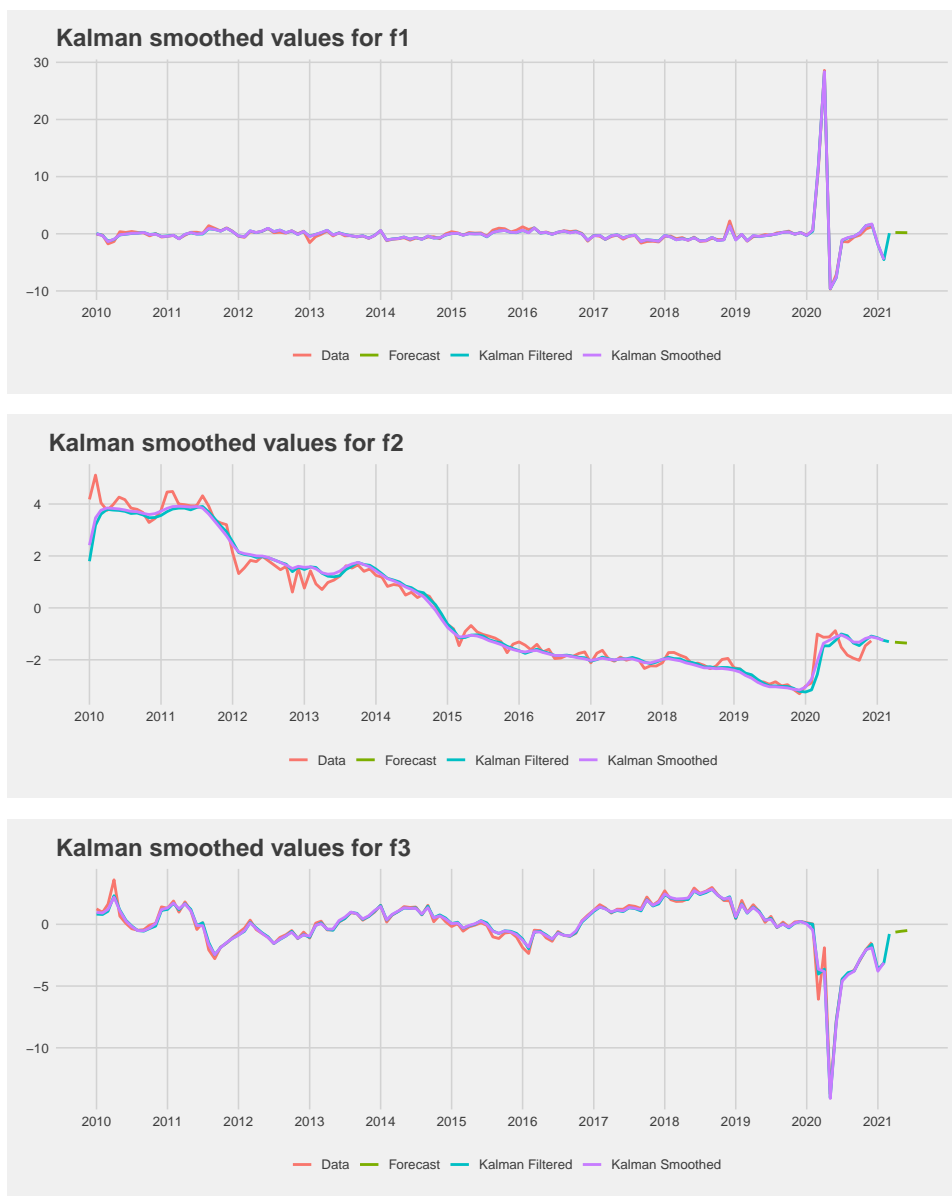
These values  $\mathbf{z}_{t|\tau}$  will serve as our estimates of the state variables (i.e., the PCA factors) from time 1 through  $\tau$ .

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

Recursively iterating over  $t = \tau + 1, \dots, T^*$ , 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  $T^*$ .



WORK IN PROGRESS

## 2.9 Predicting Quarterly Variables

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 9: Quarterly Data Covariates

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

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.

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

The  $\beta$  coefficients can be estimated with a typical OLS 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.

After estimation, we then use the same model to forecast forward the  $y_t^i$  quarterly, up through time  $\tau$ . The forecasted results are as follows.

Variable	2021Q1	2021Q2
Consumption: Goods: Durable: Motor Vehicles & Parts	0.13	0.02
Consumption: Goods: Durable: Furnishings & Durable Household Equipment	0.09	0.00
Consumption: Goods: Durable: Recreational Goods & Services	0.09	0.02
Consumption: Goods: Durable: Other Durable Goods	0.15	-0.02
Consumption: Goods: Nondurable Goods: Food & Beverages for Off-Premises Use	-0.02	0.02
Consumption: Goods: Nondurable Goods: Clothing	0.16	-0.01
Consumption: Goods: Nondurable Goods: Gasoline & Energy Goods	0.04	-0.02
Consumption: Goods: Nondurable Goods: Other Nondurable Goods	0.03	0.01
Consumption: Services: Housing & Utilities	0.01	0.00
Consumption: Services: Health Care	0.07	-0.01
Consumption: Services: Transportation	0.02	-0.02
Consumption: Services: Recreation	-0.01	-0.03
Consumption: Services: Food Services & Accomodations	0.06	-0.03
Consumption: Services: Financial Services	0.01	0.00
Consumption: Services: Other Services	-0.00	-0.00
Consumption: Services: Nonprofit Services	-0.18	-0.01
Gross Private Domestic Investment: Nonresidential Investment: Structures	-0.05	-0.02
Gross Private Domestic Investment: Nonresidential Investment: Equipment	0.08	0.01
Gross Private Domestic Investment: Nonresidential Investment: Intellectual Property	0.01	0.01
Gross Private Domestic Investment: Residential Investment	0.04	0.01
Gross Private Domestic Investment: Change in Private Inventories	55.21	38.85
Exports: Exported Goods	-0.02	0.00
Exports: Exported Services	-0.02	-0.01
Imports: Imported Goods	0.08	-0.00
Imports: Imported Services	0.02	-0.01
Government Spending: Federal	-0.00	0.00
Government Spending: State & Local	-0.00	-0.00

Table 10: 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
GDP
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 11: 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.

### 3 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	2021Q1	2021Q2
<b>GDP</b>	8.51	0.32
<b>Consumption</b>	21.43	-0.20
Goods	31.84	3.71
Durable	67.03	5.45
Motor Vehicles & Parts	69.56	9.62
Furnishings & Durable Household Equipment	41.37	0.99
Recreational Goods & Services	42.67	10.49
Other Durable Goods	81.26	-7.34
Nondurable Goods	7.88	2.48
Food & Beverages for Off-Premises Use	-7.67	9.89
Clothing	88.24	-3.01
Gasoline & Energy Goods	16.92	-7.89
Other Nondurable Goods	10.93	2.26
Services	7.17	-2.79
Housing & Utilities	2.58	1.18
Health Care	31.50	-3.70
Transportation	6.79	-7.94
Recreation	-2.01	-10.64
Food Services & Accommodations	26.47	-12.21
Financial Services	4.27	1.31
Other Services	-0.90	-1.49
Nonprofit Services	-50.65	-2.96
<b>Gross Private Domestic Investment</b>	13.27	1.27
Nonresidential Investment	17.79	3.00
Structures	-17.22	-6.43
Equipment	36.46	4.86
Intellectual Property	3.79	4.72
Residential Investment	17.43	3.73
Change in Private Inventories	55.21	38.85
<b>Net Exports</b>	-1424.31	-1416.37
<b>Exports</b>	-1.95	-0.49
Exported Goods	-7.35	1.02
Exported Services	-8.11	-4.84
<b>Imports</b>	38.77	-1.16
Imported Goods	36.29	-0.67
Imported Services	7.89	-4.88
<b>Government Spending</b>	-1.47	0.34
Federal	-1.69	1.03
State & Local	-1.53	-0.11

Table 12: Nowcasts for GDP and Subcomponents (Annualized Percent Change)

# Appendices

## A DFM Fitted Plots



