A Nowcasting Model for Time Series with Ragged-Edge Data

Model Run Date: June 25, 2021

charles@cmefi.com

Contents

List of Tables

List of Figures

1 Central Structural Model

1.1 Equations

Consider the following structural equations:

$$
\begin{aligned}
dlog(\mathbf{gdp}_t) &= \frac{\mathbf{pce}_{ss}}{\mathbf{gdp}_{ss}} dlog(\mathbf{pce}_t) + \frac{\mathbf{pdi}_{ss}}{\mathbf{gdp}_{ss}} dlog(\mathbf{pdi}_t) + \frac{\mathbf{im}_{ss}}{\mathbf{gdp}_{ss}} dlog(\mathbf{im}_t) - \\
&\quad \frac{\mathbf{ex}_{ss}}{\mathbf{gdp}_{ss}} dlog(\mathbf{ex}_t) + \frac{\mathbf{govt}_{ss}}{\mathbf{gdp}_{ss}} dlog(\mathbf{govt}_t) \\
dlog(\mathbf{pce_t}) &= \widehat{\beta}_0 + \widehat{\beta}_1 \frac{1}{4} \sum_{j=0}^{1} dlog(\mathbf{dpi}_{t-j}) + e_t \\
dlog(\mathbf{pdi_t}) &= \widehat{\beta}_0 + \widehat{\beta}_1 \mathbf{pdi}_{t-1} + e_t \\
dlog(\mathbf{govt_t}) &= \widehat{\beta}_0 + \widehat{\beta}_1 \mathbf{govt}_{t-1} + e_t \\
dlog(\mathbf{ex_t}) &= \widehat{\beta}_0 + \widehat{\beta}_1 \mathbf{ex}_{t-1} + e_t \\
dlog(\mathbf{im_t}) &= \widehat{\beta}_0 + \widehat{\beta}_1 \mathbf{im}_{t-1} + e_t
\end{aligned}
$$

We can move this into matrix form:

$$
\begin{bmatrix} 0 \\ \widehat{\beta_0^2} \\ \widehat{\beta_0^3} \\ \widehat{\beta_0^4} \\ \widehat{\beta_0^5} \\ \widehat{\beta_0^6} \end{bmatrix}
=
\begin{bmatrix}
1 & -\frac{pce}{gdp} & -\frac{pdi}{gdp} & -\frac{im}{gdp} & \frac{ex}{gdp} & -\frac{govt}{gdp} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\widehat{\beta_1^2} & -\widehat{\beta_1^2} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix} gdp_t \\ pce_t \\ pdi_t \\ im_t \\ ex_t \\ govt_t \\ pce_{t-1} \\ pdi_{t-1} \\ im_{t-1} \\ ex_{t-1} \\ govt_{t-1} \\ dpi_t \\ dpi_{t-1} \end{bmatrix}
$$

$$
dlog(ue_t) =
$$

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.

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

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,

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

where Γ is the $T-1 \times R$ dependent data matrix,
Ψ is the $T-1 \times R+1$ independent data matrix,
Λ is the $R+1 \times R$ matrix of coefficient weightings,
and V is the $T-1 \times R$ matrix of residuals.

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 $\widehat{C}$.

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.