# Imputing Weekly Values of Monthly Economic Indicators by Dynamic Matrix Completion

Serena Ng, Columbia University and NBER
Susannah Scanlan, Columbia University
July 2022

# Big Picture

- Goal: simple way to construct historical weekly indicator $\mathcal{I}^w$ from $X^m$ and $X^w$ for use in regressions.

- Large literature constructing monthly indicator $\mathcal{I}^m$ from $X^q$ and $X^m$.

  a. Latent variable approach: takes $\widetilde{F}_1^m$ as $\mathcal{I}^m$, where $X_{it} = \Lambda_i' F_t + e_{it}$.
     - $X^m$ only: (eg.dynamic SW-89, static WEI).
     - $X^m$ and $X^q$: state space, stock-flow constraints, (eg ADS, ECI).

  b. Interpolation/temporal disaggregation: eg. Chow-Lin, MIDAS.

- This paper: $\mathcal{I}^w \neq \widetilde{F}_1^w$ but is predicted by $\widetilde{F}_1^w, \ldots, \widetilde{F}_r^w$ and uses $X^m$.
  - Take CFNAI as $\mathcal{I}_t^w$, impute missing $(\mathcal{I}_t^w, \mathcal{I}_{t+1/w_t}^w, \mathcal{I}_{t+2/w_t}^w, \ldots, \mathcal{I}_{t+1}^w)$.
  - Chow-Lin type interpolation, but high frequency bridge equation.
  - Complete data estimation, no missing mechanism.
  - We first estimate $F$, then impute. (matrix completion vs. EM)
  - PC+ADL. Model dynamics of latent series without Kalman filter.

Outline:

1. Static TP imputation of the missing $\mathcal{I}^m$ from (sa) data $X^m$.
   i Efficient estimation of $F$ and $\Lambda$.
   ii Imputation when $e_{it}$ is serially correlated.
2. Dynamic TP imputation of the missing $\mathcal{I}^w$ from (nsa) data $X^w$.

Summary of Findings:

a. Estimation: improvements over static TP come from controlling for non-sphericalness of $e_{it}$, less so from dynamic estimation of $F$
b. Imputation: lots of information in lags of $e_{it}$.
c. Possible to work with seasonally unadjusted.

Improving model more important than improving estimator.

# Factor Estimation, $X$ incompletely observed

$$X_{it} = \Lambda_i' F_t + e_{it}.$$

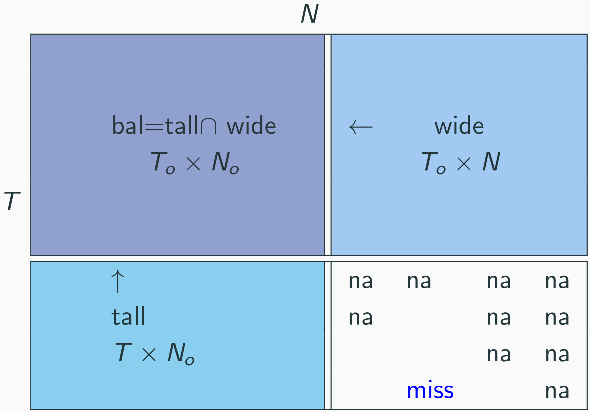
- Common assumption: MCAR/MAR, missing mechanism is ignorable
- Classical factor models and covariance structure modeling
  - Maximizing correctly assumed Gaussian likelihood gives asymptotically normal estimates of loadings under MAR.
  - For non-normal data, result holds only under MCAR.
  - Estimates can be biased if MCAR is assumed, but data are MAR.
- Large factor models: jointly estimate $F$ and missing values.
  - Static factors: iterative PC.
  - Dynamic factors: parametric state space modeling, QMLE.
  - EM+ implicit assumption of MAR.

# Recent Distribution Results for (static) FBI

EM algorithms work well, but theoretical results on imputed values only recently obtained for static PC estimates.

- Jin-Miao-Su (2021): MAR, re-weigh data
- Xiong-Pelger (2022): general missingness, also re-weigh data.
- Bai-Ng (2021), Cahan-Bai-Ng (2022): reorganize data.
  - Missing data assumption can be difficult to verify or justify.
  - Complete data estimation, distribution theory worked out.

# Tall-Wide View of $X = F\Lambda' + e$




Many use BAL estimates as initial values and iterate.

Need $F$ and $\Lambda$ for imputation. Can get them from TALL and WIDE .

Matrix completion view: systems analysis without joint estimation.

# Algorithm TW

1 Pre-process data: demean, etc.

2 From TALL, PC gives $(\widetilde{F}_{\text{tall}}, \widetilde{\Lambda}_{\text{tall}})$. Compute $\widetilde{H}_{\text{tall}}$.

3 From WIDE, PC gives $(\widetilde{F}_{\text{wide}}, \widetilde{\Lambda}_{\text{wide}})$. Compute $\widetilde{H}_{\text{wide}}$.

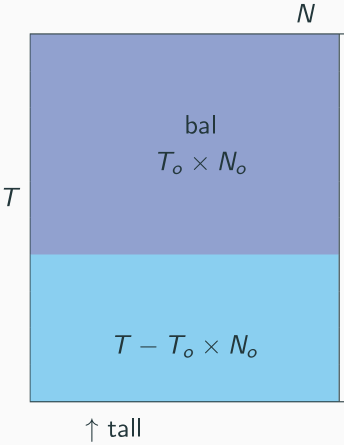
$$\begin{aligned} \widetilde{C}_{\text{tall},it} &= \widetilde{F}'_{\text{tall},t}\widetilde{\Lambda}_{\text{tall},i} = C_{it} + o_p(1) \\ \widetilde{C}_{\text{wide},it} &= \widetilde{F}'_{\text{wide},t}\widetilde{\Lambda}_{\text{wide},i} = C_{it} + o_p(1). \end{aligned}$$

4 Re-rotate: Get $\widetilde{H}_{\text{miss}} = \widetilde{H_{\text{tall}}^{-1} H_{\text{wide}}}$ by regressing the $\widetilde{\Lambda}_{\text{tall}}$ on the $N_o \times r$ submatrix of $\widetilde{\Lambda}_{\text{wide}}$. Show $\widetilde{H}_{\text{wide}}^{-1'}\widetilde{H}_{\text{miss}}\widetilde{H}_{\text{tall}} \xrightarrow{p} I_r$.

5 Output $\widetilde{X}_{it} = \begin{cases} X_{it} & (i,t) \quad \text{observed} \\ \widetilde{C}_{it} = \widetilde{F}_{\text{tall}}\widetilde{H}_{\text{miss}}\widetilde{\Lambda}'_{\text{wide}} & (i,t) \quad \text{missing} \end{cases}$

Note: Steps 2 and 3 are based on complete data. Bai-Ng (2021, JASA)

# Algorithm TP




| | | | |
|---|---|---|---|
| y | y | y | y |
| y | y | y | y |
| y | y | y | y |
| y | y | y | y |
| y | na | y | y |
| y | y | y | y |
| na | na | na | na |
| na | y | na | na |
| y | y | na | na |
| y | y | y | na |

↑ incomplete

i. Estimate $F$ from TALL.

ii. For each $i$, regress the $T_{o_i} \times 1$ observed values of $X_i$ on the corresponding submatrix of $\widetilde{F}$ to obtain $\widetilde{\Lambda}_i$. Cahan-Bai-Ng (JOE)

# Assumptions

A: Moment conditions.

B: $N, T, N_o, T_o$ large, $\frac{\sqrt{N}}{\min\{N_o, T_o\}} \to 0$, $\frac{\sqrt{T}}{\min\{N_o, T_o\}} \to 0$.

C: strong factors in the sub-blocks.

D: blocks have same moments (block stationarity).

Under A-D: $\widetilde{C}_{it} - C^0_{it} = u_{it} + v_{it} + r_{it}$.

- $u_{it}/v_{it}$ are errors from estimating $F$ and $\Lambda_i$.
- $r_{it} = O_p(\min(T_{oi}, N_o))$ uniformly in $i$ and $t$ (higher order est. errors)
- Impute (when $e_{it}$ is uncorrelated)

$$\widetilde{X}_{it} = \begin{cases} X_{it} & \text{if} \quad X_{it} \text{ observed} \\ \widetilde{C}^+_{it} & \text{if} \quad X_{it} \quad \text{missing.} \end{cases}$$

# Asymptotic results

- Proposition 1: $\min(\sqrt{N_o}, \sqrt{T_{o_i}})\left(\frac{\widetilde{C}_{it}-C_{it}^0}{\sqrt{\widetilde{\mathbb{V}}_{it}(N_o, T_{o_i})}}\right)\xrightarrow{d}N(0,1).$

  Assumptions on factor model, not missing data mechanism

  Consistent and asymptotically normal estimates, no need to iterate.

- Proposition 2: Let $\widetilde{C}_{it}^{+} = \widetilde{\Lambda}_i^{+\prime}\widetilde{F}_t^{+}$ be obtained by applying PC to $\widetilde{X}$.

$$\min(\sqrt{N_{o_t}}, \sqrt{T_{o_i}})\left(\frac{\widetilde{C}_{it}^{+} - C_{it}^0}{\sqrt{\widetilde{\mathbb{V}}_{it}^{+}(N_{o_t}, T_{o_i})}}\right)\xrightarrow{d}N(0,1).$$

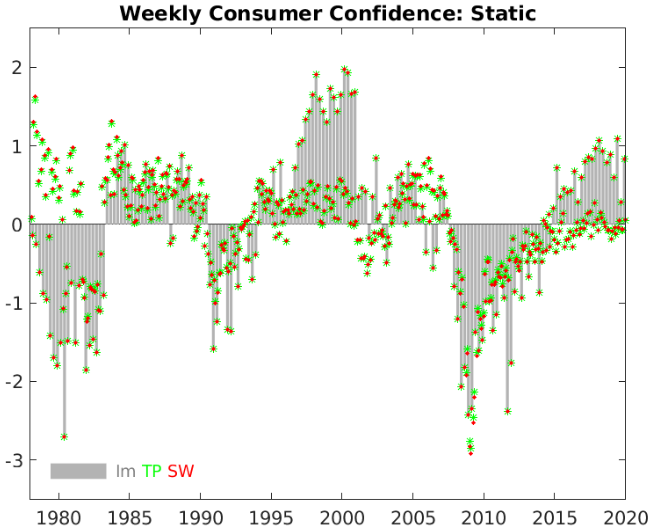
  Re-estimation improves rate since $N_{ot} > \min_s N_{os} = N_o$.

# Example: balanced panel of FRED-MD + CS

CS series reported quarterly before 1978

| date | x1 | x2 | x3 | x4 | CS |
|---|---|---|---|---|---|
| 1960-01 | 0.319 | 0.463 | 2.660 | 2.410 | na |
| 1960-02 | 0.114 | 0.091 | 0.369 | -0.569 | 100.0 |
| 1960-03 | 0.190 | 0.091 | -0.110 | -0.342 | na |
| 1960-04 | 0.341 | 0.361 | 2.590 | 0.228 | na |
| 1960-05 | 0.240 | 0.244 | -1.504 | 0.569 | 93.3 |
| 1960-06 | 0.078 | -0.024 | -0.136 | -1.141 | na |
| 1960-07 | 0.182 | 0.190 | -1.009 | -0.576 | na |
| 1960-08 | -0.162 | -0.256 | 0.341 | -0.231 | 97.2 |
| ⋮ | ⋮ | | | ⋮ | |
| 1977-10 | 0.861 | 1.135 | 1.969 | -0.065 | na |
| 1977-11 | 0.784 | 0.623 | 0.789 | 0.146 | 84.4 |
| 1977-12 | 0.443 | 0.535 | 0.457 | 1.004 | na |
| 1978-01 | -0.354 | -0.383 | -1.994 | -1.887 | 83.7 |
| 1978-02 | 0.448 | 0.526 | 2.334 | 1.208 | 84.3 |
| 1978-03 | 0.566 | 0.620 | 2.436 | 2.352 | 78.8 |

# Hypothetical exercise: CS available quarterly even after 1978




Problem: (mean,$\sigma$): TP(-.0042,0.742), SW=(-0.0013, 0.7513).

# Limitations of TP

TP does not use dynamic information in estimation or prediction.

- Estimation:
  i $\widetilde{F}$ based on cross-section information only.
  ii $\widetilde{\Lambda}$ inefficient if $e_{it}$ is serially correlated and/or heteroskedastic.

- Imputation:
  - $\widetilde{\mathcal{I}}_s^m = \widetilde{\Lambda}'_{N_m+1}\widetilde{F}_s$ as if $e_{it}$ was unpredictable.
  - Chow-Lin/MIDAS: $\widetilde{\mathcal{I}}_s^m = \widehat{B}_1^m(L^{1/m}; \widehat{\gamma})' Z_s^m$ (distributed lag)
    i Monthly data: $\mathcal{I}_s^m = B_1^m(L^{1/m}; \gamma)' Z_s^m + e_s^m$, $e_s^m$ correlated.
    ii GLS estimation of bridge equation $\mathcal{I}_t^q = \beta_1^q(L)' z_t^q + e_t^q$. Efficient estimation but not prediction.

# Model: $X_{it} = \Lambda_i' F_t + e_{it}$

| Estimator | goal | dynamics | $\Sigma_e$ |
|---|---|---|---|
| pc-ols | $(F, \Lambda)$ | - | unrestricted |
| pc-gls (Breitung) | $(F, \Lambda)$ | $e$ | $\Phi = \text{diag}(\Sigma_e)$. |
| pc-ks (Doz et al) | $(F, \Lambda)$ | $F$ | $\Phi$ |
| mle-gls (Bai-Li) | $(F, \Phi)$ | e | $\Phi$ |

All estimators give $\sqrt{N}$ consistent estimates of $F$ if $\sqrt{N}/T \to 0$.

- pc-ols is inefficient if $e$ is heteroskedastic/serially correlated.
- pc-gls assumes $\rho_i(L)e_{it} = \epsilon_{it}$, and efficiently estimates $\Lambda_i$ from:

$$[\widehat{\rho}_i(L)X_{it}] = [\widehat{\rho}_i(L)\widetilde{F}_t'] \Lambda_i + \text{error}$$

But $\rho$ irrelevant for estimating $F_t$ (Breitung-Tenhoven 2011) :

$$\widetilde{F}_t^{gls} = (\widetilde{\Lambda}'\widehat{\Phi}^{-1}\widetilde{\Lambda})^{-1}\widetilde{\Lambda}'\widehat{\Phi}^{-1}X_t.$$

# GLS vs PC-KS

- Projections estimator: $F_t^p = \overbrace{(\Sigma_F^{-1} + \Lambda'\Phi^{-1}\Lambda)^{-1}}^{\mathcal{G}} \Lambda'\Phi^{-1}X_t$.

- By Woodbury inversion in Doz et al (2011): Details

$$F_t^p = \underbrace{(\Lambda'\Phi^{-1}\Lambda)^{-1}\Lambda'\Phi^{-1}X_t}_{F_t^{gls}} - \underbrace{\mathcal{G}\Sigma_F^{-1}(\Lambda'\Phi^{-1}\Lambda)^{-1}\Phi^{-1}X_t}_{O_p(N^{-1/2})}.$$

$F_t^p$ is shrinkage estimator. But with strong factors, $\Lambda'\Phi\Lambda$ dominates. $F_t^{gls}$ and $F_t^p$ are asymptotically equivalent. Differ when $N$ is small.

- If $F$ is VAR(p): $F_t^{ks} = F_t^p$. If $T/N^3 \to 0$, $\sqrt{N}(\widetilde{F}_t^{ks} - \widetilde{F}_t^{gls}) = o_p(1)$, Bai and Li (2016):

# Results from Calibrated monte-carlo

- From 2021:12 vintage of FRED-MD, get $\widetilde{F}$, $\widetilde{\Lambda}$ by PCA.
- calibration: VAR(2) for $\widetilde{F}$, and AR(1) for $\widetilde{e}_{it}$
- Set $\Lambda$ to $\widetilde{\Lambda}$, randomly generate $F^*$, $e^*$ in each iteration.

$$\frac{\text{trace}(F^{*\prime}\widehat{F}(\widehat{F}'\widehat{F})^{-1}\widehat{F}'F^*)}{\text{trace}(F^{*\prime}F^*)}$$

| | | | $N$ | | | | |
|---|---|---|---|---|---|---|---|
| Nsim | 10 | 20 | 30 | 40 | 50 | 75 | 122 |
| pc | 0.669 | 0.703 | 0.868 | 0.901 | 0.875 | 0.923 | 0.957 |
| mle-h | 0.736 | 0.755 | 0.890 | 0.920 | 0.900 | 0.935 | 0.964 |
| pc-ks | 0.759 | 0.789 | 0.895 | 0.922 | 0.909 | 0.936 | 0.964 |
| pcgls-h | 0.774 | 0.785 | 0.890 | 0.921 | 0.901 | 0.935 | 0.964 |
| pcgls-har | 0.761 | 0.773 | 0.883 | 0.919 | 0.897 | 0.934 | 0.964 |

pc-ks: Barrigozi-Luciani (2019)

mle-h: Bai-Li (2016)

pcgls: Breitung-Tenhofen (2011)

Regression: $F_j^* = a + b_1\widetilde{F}_1 + b_2\widetilde{F}_2 + b_3\widetilde{F}_3 +$ error

| Nsim | 10 | 20 | 30 | 40 | 50 | 75 | 122 |
|---|---|---|---|---|---|---|---|
| | | | $R^2$: $j=1$ | | | | |
| pc | 0.736 | 0.841 | 0.897 | 0.936 | 0.922 | 0.956 | 0.971 |
| mle-h | 0.793 | 0.872 | 0.919 | 0.947 | 0.935 | 0.962 | 0.976 |
| pc-ks | 0.774 | 0.877 | 0.926 | 0.949 | 0.941 | 0.962 | 0.976 |
| pcgls-h | 0.817 | 0.874 | 0.920 | 0.947 | 0.935 | 0.962 | 0.976 |
| pcgls-har | 0.790 | 0.856 | 0.916 | 0.946 | 0.933 | 0.961 | 0.976 |
| | | | $R^2$: $j=2$ | | | | |
| pc | 0.628 | 0.614 | 0.868 | 0.877 | 0.861 | 0.907 | 0.954 |
| mle-h | 0.713 | 0.686 | 0.886 | 0.903 | 0.896 | 0.924 | 0.962 |
| pc-ks | 0.770 | 0.743 | 0.888 | 0.905 | 0.902 | 0.924 | 0.962 |
| pcgls-h | 0.769 | 0.741 | 0.885 | 0.905 | 0.897 | 0.924 | 0.962 |
| pcgls-har | 0.769 | 0.740 | 0.879 | 0.901 | 0.893 | 0.923 | 0.962 |
| | | | $R^2$: $j=3$ | | | | |
| pc | 0.642 | 0.655 | 0.837 | 0.891 | 0.842 | 0.905 | 0.946 |
| mle-h | 0.700 | 0.706 | 0.864 | 0.910 | 0.870 | 0.920 | 0.954 |
| pc-ks | 0.729 | 0.745 | 0.871 | 0.912 | 0.883 | 0.921 | 0.954 |
| pcgls-h | 0.736 | 0.738 | 0.864 | 0.911 | 0.870 | 0.920 | 0.954 |
| pcgls-har | 0.722 | 0.722 | 0.853 | 0.909 | 0.864 | 0.919 | 0.954 |

Result 1. Estimator of $F$ matters when $N$ is very small

# $(1-\rho_{\mathcal{I}} L)e_{\mathcal{I},t} = \epsilon_{\mathcal{I},t}$ and Prediction

- $e_{\mathcal{I}t}$ serially correlated implies $\mathcal{I}_s = \Lambda'_{\mathcal{I}} F_s + \underbrace{\rho_{\mathcal{I}}(\mathcal{I}_{s-1} - \Lambda'_{\mathcal{I}} F_{s-1})}_{\rho_{\mathcal{I}} e_{\mathcal{I}s|s-1}} + \epsilon_{\mathcal{I}s}$.
  - Static TP imputation is biased because of omitted lags.
  - Chow-Lin use lags of $F$ but not lags of $\mathcal{I}$ as covariates.
  - Ideal regression is infeasible because $\mathcal{I}_s$ has missing values.

- We replace $\mathcal{I}_{s-1}$ by $\mathcal{I}^{(k)}_{s-1}$ and iteratively estimate Durbin equation:

$$\mathcal{I}^{(k)}_s = \Lambda'_{\mathcal{I}} \widetilde{F}_s + \rho \mathcal{I}^{(k)}_{s-1} + \gamma'_{\mathcal{I}} \widetilde{F}_{s-1} + \epsilon_{\mathcal{I}s}.$$

  using static $\widetilde{\mathcal{I}}_s$ as initial values and update $\mathcal{I}^{(k+1)}_s$ by the fit.
  - ADL bridge regression in high frequency data. Sargan-Drettakis (74).
  - Estimate dynamic model with missing values without Kalman filter.
  - Can be used with any estimator of $F$.
  - Precise estimates of $(\Lambda_{\mathcal{I}}, \rho_{\mathcal{I}}, \gamma_{\mathcal{I}})$. Works!

# KSM: Kalman Smoothing with Missing Values

$$
\begin{aligned}
X_t^+ &= \Lambda^{+'} F_t + e_t^+, \quad X_t^+ = \begin{pmatrix} X_t^{w'} & \mathcal{I}_t^w \end{pmatrix}', \\
e_{it} &= \rho_i e_{it-1} + \epsilon_{it} \qquad \epsilon_{it} \sim N(0, \sigma_{\epsilon,i}), i = 1, \ldots, N_w \\
e_{\mathcal{I},t} &= \rho e_{\mathcal{I}} \mathcal{I}, t-1 + \epsilon_{\mathcal{I},t} \\
F_t &= A F_{t-1} + u_t \quad \eta_t \sim N(0, \Sigma_u).
\end{aligned}
$$

- Jungbacker et al (2011): rewrite state space model to selectively include $e_t$ in state vector. We define $X_t^+ = Z\alpha_t + c_t$, $c_t = \rho^+ X_{t-1}^+$, $Z = \begin{pmatrix} \Lambda^+ & -\rho^+ \Lambda^+ \end{pmatrix}$, $\alpha_t = (F_t' \; e_{\mathcal{I},t})'$, $\rho^+ = \text{diag}(\rho_1, \ldots, \rho_{N_w}, 0)$.
- Missing values omitted from estimation as in Banbura-Modugno.

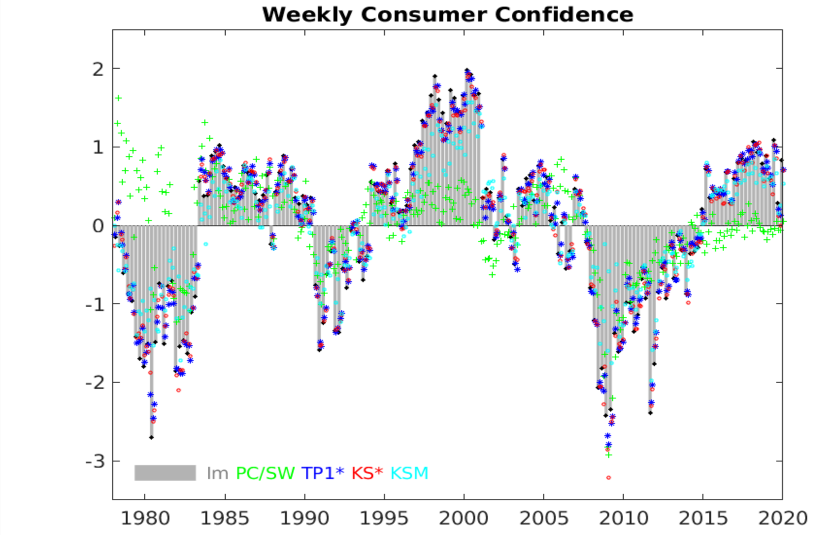
Note: results not very robust.

# Results from a hypothetical exercise

- CS hypothetically released quarterly even after 1978.
- $T = 504, N = 127$. Use realized values as true values in evaluation.

$$\|\widehat{\mathcal{I}}^m - \mathcal{I}^m\|_F$$

| | | static | | | dynamic | | | | |
|---|---|---|---|---|---|---|---|---|---|
| start | end | tp | sw | mle-h | tp* | mle* | pc-ks* | pc-gls* | ksm |
| Jan-78 | Dec-19 | 15.62 | 15.62 | 15.18 | 6.771 | 7.062 | 7.053 | 7.109 | 8.078 |
| Jan-78 | Dec-83 | 10.31 | 10.44 | 10.14 | 2.766 | 3.003 | 2.988 | 2.939 | 3.868 |
| Jan-84 | Jun-94 | 4.483 | 4.431 | 4.267 | 3.111 | 3.209 | 3.137 | 3.204 | 3.655 |
| Jul-94 | Dec-00 | 7.174 | 7.083 | 6.990 | 1.841 | 1.906 | 1.949 | 1.987 | 2.728 |
| Jan-01 | Dec-06 | 3.849 | 3.858 | 3.630 | 3.215 | 3.439 | 3.296 | 3.342 | 3.463 |
| Jan-07 | Dec-10 | 3.433 | 3.424 | 3.290 | 2.300 | 2.482 | 2.572 | 2.586 | 2.667 |
| Jan-11 | Dec-19 | 6.279 | 6.231 | 6.056 | 3.081 | 2.990 | 3.114 | 3.134 | 3.223 |




Result 2. serial correlation in $e_{it}$ matters!

Outline:

1. TP estimation of $F$ and $\Lambda$ from incomplete data. ✓
2. Static TP imputation of regularly spaced data. ✓
   - i Dynamic imputation and modeling $e_{it}$. ✓
   - ii Imputing weekly CFNAI.
3. Weekly Data: 1990-2019 $T_w = 1566$. $N_w = 20$ (or 80).

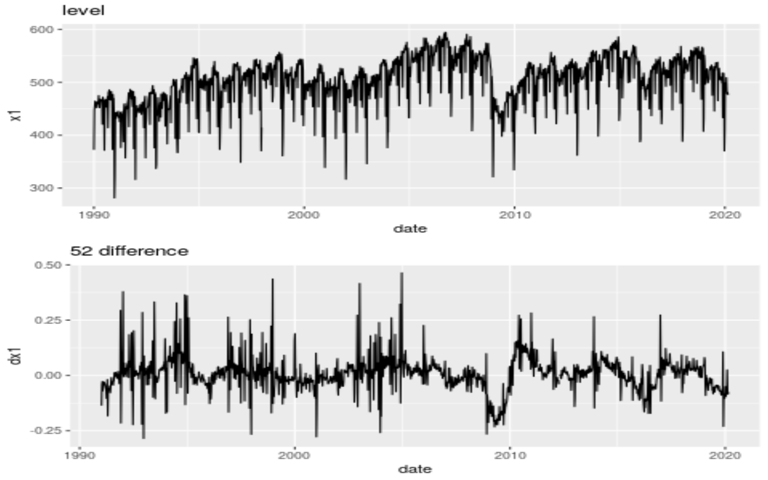
| Monthly data | Weekly data |
|---|---|
| regularly spaced | irregularly spaced |
| seasonally adjusted | seasonally unadjusted |
| $N > 100$ | much smaller $N$ |

# Weekly Panel

| date | wiy | wim | eleout | uiinit | wtioil | cfnai |
|---|---|---|---|---|---|---|
| 1990-12-24 | 52 | 4 | 134 | 454 | 27.60 | na |
| 1990-12-31 | 53 | 5 | 139 | 415 | 24.90 | -0.92 |
| 1991-01-07 | 1 | 1 | 122 | 437 | 27.30 | na |
| ⋮ | ⋮ | ⋮ | ⋮ | | | |
| 2020-02-24 | 8 | 4 | 178 | 217 | 44.83 | 0.08 |
| 2020-03-02 | 9 | 1 | 172 | 211 | 41.14 | na |
| 2020-03-30 | 13 | 5 | 164 | 6615 | 28.36 | -4.26 |
| 2020-04-20 | 16 | 3 | 170 | 3867 | 15.99 | na |
| 2020-04-27 | 17 | 4 | 167 | 3176 | 19.72 | -17.89 |
| | | | ⋮ | | | |
| 2020-05-18 | 20 | 3 | 166 | 2123 | 33.49 | na |
| 2020-05-25 | 21 | 4 | 169 | 1897 | 35.57 | 4.64 |
| | | | ⋮ | ⋮ | | |
| 2020-12-21 | 51 | 3 | 177 | 790 | 48.18 | na |
| 2020-12-28 | 52 | 4 | 160 | 787 | 48.35 | 0.36 |

- weekly $\mathcal{I}^w$ matches $\mathcal{I}^m$ the last week of every month.

# Strong but not strictly periodic seasonal variations




- Traditional seasonal filters are for monthly/quarterly data.
- 52-week differencing creates spikes.

# Imperfect or No Adjustment?

Factor representation of raw data:

$$X^w_{i,t+j/w_t} = \Lambda_i^{w\prime} F^w_{t+j/w_t} + e^w_{i,t+j/w_t} + S^w_{i,t+j/w_t}$$

Seasonal adjustment: $X^w_{i,t+j/w_t} - S^w_{i,t+j/w_t}$.

- Estimate $F$ from $X^w$ since $S^w$ (idio.) not needed for imputation.
- pc-gls and ksm: possible Haywood problem.
- expand data matrix to include three lags of $X^w$ (hence $N_w = 80$).

# Data= $X^w, X^w_{-1}, X^w_{-2}, X^w_{-3}$

Let principle components do the smoothing.

| $\widetilde{F}_1$, 0.35 | | $\widetilde{F}_2$, 0.15 | | $\widetilde{F}_3$, 0.12 | |
|---|---|---|---|---|---|
| uicont | 0.885 | ccinrv | 0.461 | wtioil | 0.431 |
| uicont | 0.885 | ccinrv | 0.460 | wtioil | 0.431 |
| uicont | 0.885 | ccinrv | 0.459 | wtioil | 0.430 |
| uicont | 0.884 | ccinrv | 0.456 | wtioil | 0.427 |
| uiinit | 0.819 | crbcom | 0.351 | altot1 | 0.394 |
| uiinit | 0.816 | crbcom | 0.348 | altot1 | 0.393 |
| uiinit | 0.810 | crbcom | 0.344 | altot1 | 0.392 |
| uiinit | 0.803 | crbcom | 0.338 | altot1 | 0.388 |
| fspcom | 0.776 | altot1 | 0.259 | rigcou | 0.321 |
| fspcom | 0.776 | altot1 | 0.257 | rigcou | 0.314 |

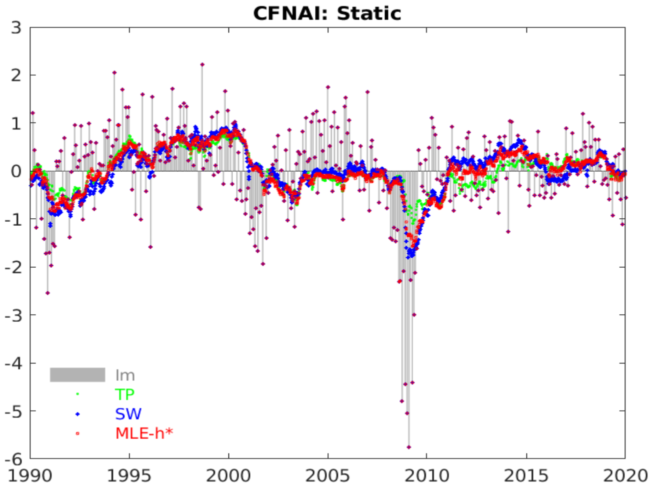
crbcom=commodity price

altot1= avg. total mortgage loan

rigcou=rotary rig count

ccinrv=consumer credit.

# Pre-covid



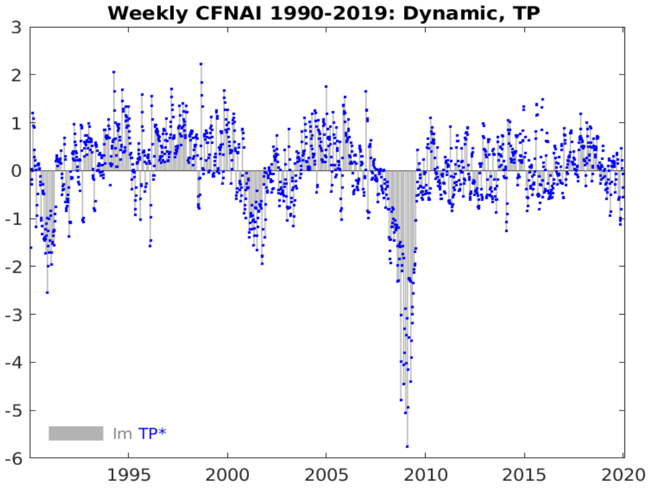
CFNAI: Static
3
2
1
0
-1
-2
-3
-4
-5
-6
1990
1995
2000
2005
2010
2015
2020
Im
TP
SW
MLE-h*

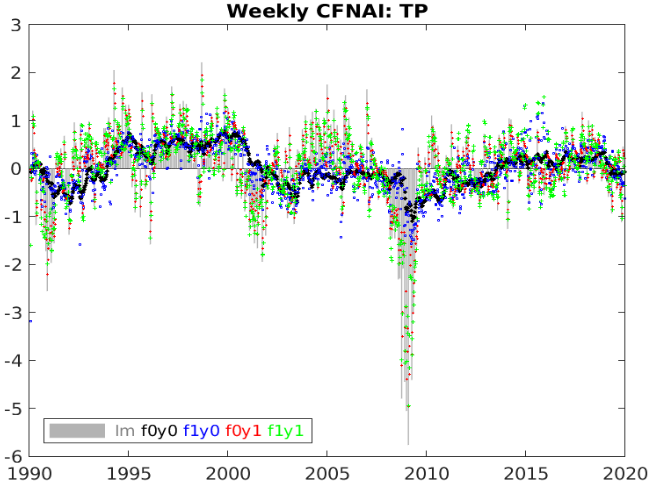
# TP Estimation of $\mathcal{I}^w$ given $\mathcal{I}^m$

1. Let $X^w$ be $T \times N_w$ panel data with missing values. Standardize.
2. Estimate $F^w$ and $\Lambda^w$ by TP. Impute $\widetilde{\mathcal{I}}^{w,(0)}_{k,t/w_t} = \widetilde{\lambda}^{w\prime}_k \widetilde{F}^w_{t/w_t}$ if missing.
3. Until convergence, let $\widetilde{\mathcal{I}}^{w,(k+1)}_s$ be the fit of

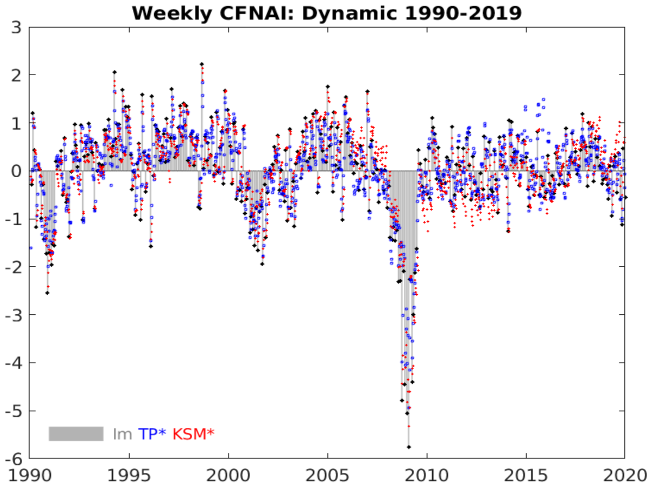
$$\mathcal{I}^{w,(k)}_s = \Lambda_i' \widetilde{F}^w_s + \beta_i' \widetilde{F}^{w,(k)}_{s-1} + \rho^w \mathcal{I}^{w,(k)}_{s-1} + \epsilon^w_s$$

   Dynamic bridge regression in high frequency data.


Weekly CFNAI 1990-2019: Dynamic, TP
3
2
1
0
-1
-2
-3
-4
-5
-6
1995
2000
2005
2010
2015
2020
Im TP*

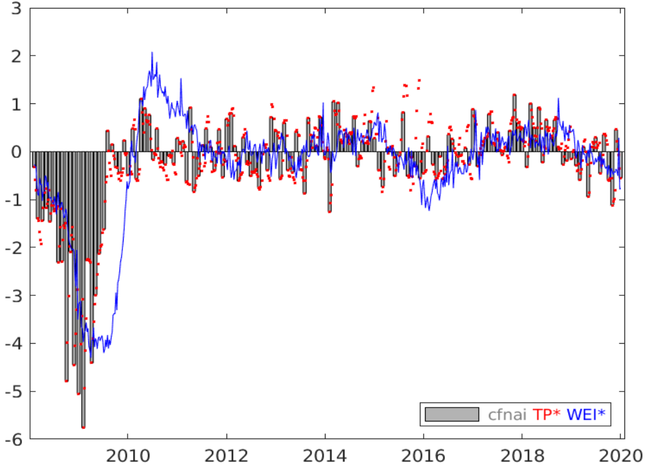


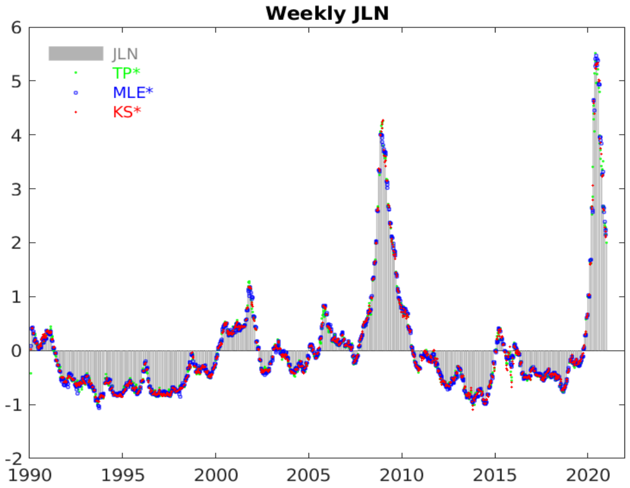

| $\text{std}(\mathcal{I}^w)(p_F, p_e)$ | | | |
|---|---|---|---|
| (0,0): 0.582 | (1,0): 0.639 | (0,1): 0.814 | (1,1): 0.845 |



Weekly CFNAI: Dynamic 1990-2019
3
2
1
0
-1
-2
-3
-4
-5
-6
1990
1995
2000
2005
2010
2015
2020
lm TP* KSM*

# Comparison with WEI




Weekly JLN
JLN
TP*
MLE*
KS*
6
5
4
3
2
1
0
-1
-2
1990
1995
2000
2005
2010
2015
2020

# COVID headaches

From Ng (2021)

- For factor estimation, an identification problem
  - pre-covid: $X_{it} - \mu_i = \Lambda_i' F_t + e_{it}^X$
  - post-covid: $X_{it} - \mu_i = \Lambda_i' F_t + \Gamma_i V_t + e_{it}^X$

  PCA of economic variables no longer estimate economic factors

- For forecasting and estimation: omitted variables problem:

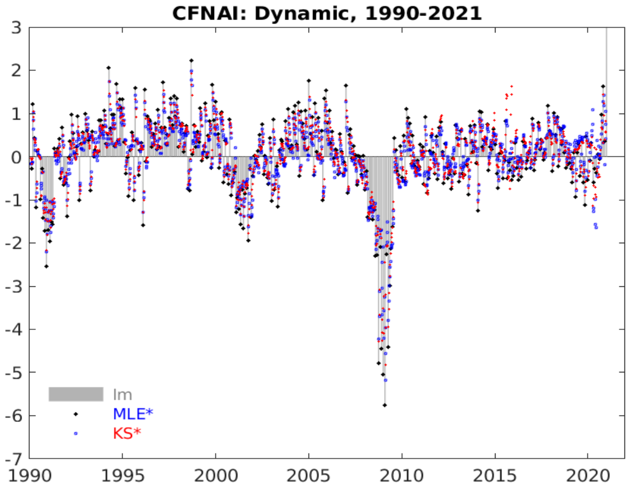
- For VAR: $n$ variables, now $n + 1$ shocks. Identification problem.

- De-covid Regressions: : $\text{CFNAI}_t$ =a + b(L) $\text{COVID}_t$ + error
  - Regressors: change in positivity rate, hospitalization, deaths.
  - $P_t, P_{t-1}, P_{t-2}, H_{t-1}, H_{t-2}, D_{t-1}, D_{t-2}$ has $R^2 = 0.8$
  - (mean,var): precovid= (-0.076, 0.493). after: (-0.009, 0.438).

# COVID


Systematic outlier adjustment: $\widehat{F}_1$ is still $13\sigma$ in 2020:03.

# Post-covid



CFNAI: Dynamic, 1990-2021
3
2
1
0
-1
-2
-3
-4
-5
-6
-7
1990
1995
2000
2005
2010
2015
2020
Im
MLE*
KS*

# Summary

- A weekly indicator pegged to a monthly diffusion index.
- Matrix completion view:
  - system analysis without joint modeling
  - separates estimation of $F$ from imputation.
  - imputes missing values in time series without Kalman filter.
- An ADL high frequency bridge equation that uses latent lagged dependent variable in prediction.
- Model specification more important than choice of estimator for $F$.