A New Identification Procedure for Unknown Integrated Common Factors with a Convergent Panel

Jihun Kwak & Donggyu Sul
University of Texas at Dallas
Two-way fixed effects (or panel augmented) regressions eliminate important information (co-movement)

\[ y_{it} = a_i + \theta_t + \beta x_{it} + u_{it} \]

Rather than eliminating, suggest to identify unknown (nonstationary or trending) common factors

\[ y_{it} = a_i + \delta_i G_t + \beta x_{it} + u_{it} \]

Utilizing weak \(\sigma\)-convergence test proposed by Kong, Phillips and Sul (2019)

\[ \hat{z}_{it} = y_{it} - \delta_i G_t, \text{ If } V(\hat{z}_{it}) \downarrow \text{ over time, } G_t \text{ becomes } F_t \]

Requirement: \(F_t\) should have a (non)stochastic trend or nonstationary & \(y_{it}\) should converge over time.
Issues: Two-way FE Regression

• Two-way FE Panel Regression - See Sul (2019 Chapter 6)

\[ y_{it} = a_i + \theta_t + \beta x_{it} + u_{it} \]

Example: Log Infant Mortality Rates
• What do you see?
• Want to explain co-movements?
• Possibly determinants?
• Do you think you can do with the two-way FE regression?
Issues: Two-way FE Regression

- Two-way FE Panel Regression - See Sul (2019 Chapter 6)

\[ y_{it} = a_i + \theta_t + \beta x_{it} + u_{it} \]
Identification Procedure

• Assume $y_{it} = c_i + \lambda_i F_t + y_{it}^0$. Further assume that $F_t$ has a (non)stochastic trend. Then the weak sigma convergence test is usually done with $z_{it} = c_i + y_{it}^0$

• Suppose that $F_t = \varphi' G_t + u_t$ where $u_t = u_{o,t} T^{\gamma - \frac{1}{2}}$ with $0 \leq \gamma < 1$ and $u_{o,t}$ has a finite variance. Then $G_t$ can be a long run determinant of $F_t$ since as $T \to \infty$, $\varphi' G_t \to F_t$.

• The weak sigma convergence test can be used for identifying $F_t$.

• Run the following regression:

  \[
  y_{it} = c_i + \delta_i' G_t + e_{it} \text{ for each } i
  \]

  \[
  \hat{z}_{it} = \hat{c}_i + \hat{e}_{it}
  \]

  \[
  K_{nt} = a + \varphi t + u_t
  \]

If $\hat{\varphi} < 0$ significantly, then $G_t$ are determinants of $F_t$
Features of the Proposed Method

• Works with non(stochastic) panel data: Don’t need to identify the number of common factors or estimate the common factors. Of course, estimating them gives extra & useful info.

• Allow flexible deviation from true factors

\[ F_t = \varphi' G_t + u_t \]

I. Assume \( F_t = \varphi' G_t \) if \( t \in \tau \), but \( F_t = \varphi' G_t + u_t^* \) if \( t \in \tau^c \).

II. Denote \( T_m \) as the number of times of \( t \in \tau^c \), and let \( T_m = T^\gamma \).

III. Then \( u_t = u_{o,t} T^2 \gamma^{\frac{1}{2}} \) with \( 0 \leq \gamma < 1 \) and \( u_{o,t} \) has a zero mean with a finite variance.

• Parker and Sul (2016) assume \( \gamma = 0 \).

• If \( \gamma = 1 \), \( F_t \) is cointegrated with \( G_t \), but \( F_t \neq \varphi' G_t \) for all \( t \).
Features of the Proposed Method

• Method by Parker and Sul (2016):

1. Based on the estimated number of common factors
2. Works only with stationary (more strictly speaking near white noise) variables
3. Hard to identify unknown common factors even with a single factor

\[
y_{it} = a_i + \lambda_i F_t + y_{it}^o, \quad F_t = F_{t-1} + v_t, \quad y_{it}^o \sim iid(0,1)
\]
\[
\Delta y_{it} = \lambda_i \Delta F_t + y_{it}^o - y_{it-1}
\]

Signal is reduced, but noise increases.

• Newly proposed method works with nonstationary & trended data
Models

\[ y_{it} = c_i + \lambda'_i F_t + y_{it}^o \text{ with } y_{it}^o = \varepsilon_{it} t^{-\alpha} \]

- Consider only a case where \( \alpha > 0 \): \( y_{it} \) is weakly \( \sigma \)-converging to its cross-sectional average of common components.

- Consider the regression

\[ y_{it} = c_i + \delta'_i G_t + u_{it} \]

- Define \( \hat{z}_{it} = y_{it} - \delta'_i G_t \).

- Construct the cross-sectional variance of \( \hat{z}_{it} \), \( K_{nt} \). Running the following trend regression

\[ K_{nt} = a + \phi t + w_t \]

- Use Newey-West long run variance for constructing the t-ratio.

\[
\hat{t}_\phi = \frac{\hat{\phi}_{nt}}{\sqrt{\hat{\Omega}_w^2 / \sum_{t=1}^{T} \hat{\tau}_t^2}} \quad \Omega_w^2 = \frac{1}{T} \sum_{t=1}^{T} \hat{W}_t^2 + \frac{2}{T} \sum_{\ell=1}^{L} \sum_{t=1}^{T-\ell} \left(1 - \frac{\ell}{L+1}\right) \hat{W}_t \hat{W}_{t+\ell}
\]
Theorem 1

Under Assumption A, as $n, T \to \infty$ regardless of the $n/T$ ratio, the t-ratio has the following asymptotic properties.

(i) If $\beta > 0$, $\alpha \leq 1/2$, and all variables in $G_t$ are true determinants of $F_t$, then

$$t_{\hat{\phi}} \to \begin{cases} 
-\infty & \text{if } 0 < \alpha < 1/2, \\
-\sqrt{6/\kappa^2} & \text{if } \alpha = 1/2 
\end{cases}$$

(ii) If $\beta > 0$ but some variables in $G_t$ are not true determinants of $F_t$, then

$$t_{\hat{\phi}} \to +\infty.$$
Theorem 1 requires $\alpha \leq 1/2$

If $\alpha > 1/2$, then the convergence occurs too fast -> Not realistic

The trend regression does not capture the convergence behavior.

If $\alpha > 1/2$, the asymptotic t-ratio becomes random with a negative mean.
Remark 3

• Factor loadings are identical for most of individuals: Local heterogeneity

• Then \( y_{it} \) is weakly \( \sigma \)–converging, and \( z_{it} (\equiv y_{it} - \lambda_t^t F_t) \) is weakly \( \sigma \)-converging as well.

• In this case, DO NOT use any stationary variable as a potential factor.
Monte Carlo Results

- Very Good. Skip
Empirical Example

46 disaggregate PCE price indexes from 1978 to 2016 (from KPS’ example)
Find one factor by using BN’s IC2

- After taking growth rate, another first difference, and standardization, See Sul (2019, Chap 4)
- Before eliminating the common factor, the t-ratio becomes -3.67 (factor loadings are similar)
- After eliminating the common factor, the t-ratio becomes -4.76

Table 1: (Non)Stochastic Trends in Inflation and Potential Determinants

<table>
<thead>
<tr>
<th>Potential Determinant</th>
<th>Trend</th>
<th>ADF (Trend)</th>
<th>ADF (No Trend)</th>
<th>Relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_t$</td>
<td>Yes</td>
<td>-3.989</td>
<td>n.a.</td>
<td></td>
</tr>
<tr>
<td>FFR</td>
<td>Yes</td>
<td>-3.629</td>
<td>n.a.</td>
<td>Yes</td>
</tr>
<tr>
<td>ln(GDP)</td>
<td>Yes</td>
<td>-1.549</td>
<td>n.a.</td>
<td>Yes</td>
</tr>
<tr>
<td>ln(IM)</td>
<td>Yes</td>
<td>-1.757</td>
<td>n.a.</td>
<td>Yes</td>
</tr>
<tr>
<td>$\Delta$ lnGDP</td>
<td>No</td>
<td>n.a.</td>
<td>-3.560</td>
<td>No</td>
</tr>
<tr>
<td>$\Delta$ ln(IM)</td>
<td>No</td>
<td>n.a.</td>
<td>-3.626</td>
<td>No</td>
</tr>
</tbody>
</table>
Empirical Example

- Use Federal Fund Rate as a potential factor:
- \( t \)-ratio for \( \varphi = -4.28 \)
- FFR becomes the common factor to all 46 PCE inflations
- Parker and Sul (2016) method: Can't identify FFR as unknown common factor to PCE inflations
When a wrong variable is included:

Total import & Real GDP as examples.

Both explain the trend of the headline inflation well.

However... weak sigma convergence does not hold with both variables.
Empirical Example

When an irrelevant variable is included:

Real GDP growth rate: Stationary

A stationary variable cannot explain a nonstationary trend

Weak sigma convergence holds but meaningless!
Conclusion

• If a (non) stochastic panel data of interest is converging over time, then the unknown common factors can be identified by using the weak $\sigma$-convergence test.

• If not, use Parker and Sul (2016) method with first differenced + prewhitened + standardizied data to identify the unknown common factors.

• Practically useful: Good forecasting performance + (more importantly) new theory can be developed from new empirical findings.

• Several interesting extensions are promising:
  i. A few outliers -> panel is not converging. How to eliminate them?
  ii. Sample cross-sectional variance is not robust. How to get a better (robust) measure for cross-sectional dispersion?