An Enhanced Estimation Routine for SEFMs

Version 1; February 23, 2021

1 SEFM Estimation

Let there be N countries with T time series observations each (yearly frequency here). The set $\mathcal{N}$ shall contain all countries, hence $|\mathcal{N}| = N$. For each country $i,\ i = 1, \ldots, N$, we have information about a certain variable of interest, y_{it}, called the *outcome*, for $t = 1, \ldots, T$. The overall goal is to estimate a Single Equation Forecasting Model (SEFM) for each country.

1.1 Step 1

For a given country $i \in \mathcal{N}$, run $N - 1$ OLS time series regressions of the type

$$y_{it} = \delta_{0,i} + \delta_{1,i} y_{jt} + \omega_{jt}, \quad t = 1, \ldots, T, \tag{1}$$

for all $j = 1, \ldots, N$, with $j \neq i$. Compute for the estimated residuals $\hat{\omega}_{it}^{(j)}$ the (estimated) first order autocorrelation $\hat{\rho}_i^{(j)} = \mathrm{Cov}\big(\hat{\omega}_{it}^{(j)}, \hat{\omega}_{i,t-1}^{(j)}\big)/\mathrm{Var}\big(\hat{\omega}_{it}^{(j)}\big),\ \forall j \in \mathcal{N}\backslash i$, and use this quantity to calculate the Cointegration Regression Durbin-Watson (CRDW) test statistic as

$$\mathrm{CRDW}_i^{(j)} = 2(1 - \hat{\rho}_i^{(j)}), \quad j = 1, \ldots, N,\ j \neq i. \tag{2}$$

Then, fix a τ. If $\mathrm{CRDW}_i^{(j)} > \tau$, we retain the regressor y_{jt} in (1). Let a set $\mathcal{R}_i \subseteq \mathcal{N}$ hold all retained variables for country i, including country i itself.

1.2 Step 2

For each country contained in $\mathcal{R}_i$, run the Johansen cointegration method for a given country i. Let $\mathcal{J}_i \subseteq \mathcal{R}_i$ contain all countries for which the Johansen cointegration method returned a cointegration relation (excluding i itself). Let the variable that emerges from the exact cointegration relation be denoted $c_{it},\ t = 1, \ldots, T$, which is a linear function of the outcome variables contained in $\mathcal{J}_i$ and the outcome of i. Define furthermore a vector $\mathbf{j}_{it} \equiv \big\{y_{kt}\big\}_{k \in \mathcal{J}_i}$ that contains the outcomes of all countries for which a cointegration relation with country i was found.

1.3 Step 3

For a given country i, compute all pairwise differentiated correlations over the time dimension, that is,

$$\varrho_{ij} \equiv \text{Corr}(\Delta y_{it}, \Delta y_{j,t-1}) \tag{3}$$

for $j = 1, \ldots, N, j \neq i$. Then, (pre-) define by $\tau_{pc} \geq 0$ and $\tau_{nc} \leq 0$ thresholds for a minimum positive or negative correlation, respectively. Next, let the sets $\mathcal{C}_i^+ = \{k : \varrho_{ik} > \tau_{pc} > 0\ , k = 1, \ldots, N,\ k \neq i\}$ and $\mathcal{C}_i^- = \{k : \varrho_{ik} < \tau_{nc} \leq 0, k = 1, \ldots, N,\ k \neq i\}$ contain all countries with a sufficiently – in absolute value – large positive and negative differentiated autocorrelation, respectively, as in (3).

1.4 Step 4

Fix an integer K_{pc} (K_{nc}), which denotes how many of the – in absolute value – largest positively (negatively) correlated countries[1] shall be considered. Then, for a given country i, collect the K_{pc} countries that are strongest positively correlated with i in a K_{pc}-vector $\mathbf{p}_{it}$. Analogously, the K_{nc}-vector $\mathbf{n}_{it}$ contains the K_{nc} strongest negatively correlated countries with i. Formally,

$$\mathbf{p}_{it} \equiv \left\{y_{(k),t}\right\}_{k\in\mathcal{C}_i^+}^{K_{pc}} \quad \text{and} \quad \mathbf{n}_{it} \equiv \left\{y_{(k),t}\right\}_{k\in\mathcal{C}_i^-}^{K_{nc}}\ , \tag{4}$$

for $t = 1, \ldots, T$, where $x_{(w)}$ denotes the w-th largest value (in absolute terms) in a vector $\mathbf{x}$. Hence, p_{it1} (the first value in $\mathbf{p}_{it}$) is equal to the outcome of the country that has the strongest positive correlation with i; and n_{it1} is the outcome of the country that is strongest negatively correlated with i.

1.5 Step 5

Finally, we can obtain estimates for the SEFM where we only have to fix the values of five parameters: $\tau, \tau_{pc}, \tau_{nc}, K_{pc}$ and K_{nc}. We can do so by means of classical OLS and NLS. Before we proceed, let us recall what we have obtained in the previous steps for a given country $i \in \mathcal{N}$:

- the variable c_{it}, which is the cointegration variable for i,
- the vector $\mathbf{j}_{it}$, which contains the outcomes of the countries for which a cointegration relationship with i was found,
- the vector $\mathbf{p}_{it}$, which contains the outcomes of the K_{pc} countries that are most strongly positively correlated with i (ordered by strength of correlation),
- the vector $\mathbf{n}_{it}$, which contains the outcomes of the K_{nc} countries that are most strongly negatively correlated with i (ordered by strength of correlation).

[1] Henceforth, the term "correlation" refers to the differentiated autocorrelation as defined in (3).

1.5.1 OLS Estimation

Using the obtained variables, run the following OLS regression for a given country $i \in \mathcal{N}$:

$$\Delta y_{it} = \mu_i + \gamma_i c_{i,t-1} + \alpha_i(\Delta y_{i,t-1}) + \boldsymbol{\beta}_i^\top(\Delta \mathbf{p}_{i,t-1}) + \boldsymbol{\lambda}_i^\top(\Delta \mathbf{n}_{i,t-1}) + \varepsilon_{it}, \quad (5)$$

for $t = 3, \ldots, T$. Note that $\boldsymbol{\beta}_i$ is a K_{pc}-vector of coefficients and $\boldsymbol{\lambda}_i$ is a K_{nc}-vector of coefficients.

1.5.2 NLS Estiamation

For a given country $i \in \mathcal{N}$, fit the following model by means of NLS estimation:

$$\begin{aligned}\Delta y_{it} = \mu_i + & \\ & + \gamma_i\Big(y_{i,t-1} - \boldsymbol{\zeta}_i^\top \mathbf{j}_{i,t-1}\Big) + \\ & + \alpha_i(\Delta y_{i,t-1}) + \\ & + \phi_i \sum_{j=1}^{K^{pc}} \beta_i^{j-1}(\Delta p_{i,t-1,j}) + \\ & + \psi_i \sum_{j=1}^{K^{nc}} \lambda_i^{j-1}(\Delta n_{i,t-1,j}) + \\ & + \varepsilon_{it},\end{aligned} \quad (6)$$

for $t = 3, \ldots, T$. The parameter $\boldsymbol{\zeta}_i$ in (6) is a $|\mathcal{J}_i|$-vector of coefficients. Further (scalar) parameters are $\mu_i, \gamma_i, \alpha_i, \phi_i, \beta_i, \psi_i$ and λ_i. Hence, the NLS needs to numerically estimate $7+|\mathcal{J}_i|$ parameters. Note that the problem associated with (6) is non-convex,[2] hence we cannot guarantee that the returned local optimum is also the global optimum, let alone that the chosen optimization algorithm even converges (especially given the number of parameters to estimate here).

We observed that the NLS optimization algorithm applied to (6) tends to shrink certain parameter estimates to zero in order to set one parameter estimate's value to be very high (in absolute terms). Since this is an undesired behavior, we impose the following restrictions in (6):

- $|\alpha_i| \leq 1$,
- $\gamma_i \in [-1,\ 0]$,
- $|\beta_i| \leq 1$,
- $|\lambda_i| \leq 1$,

while we leave $\boldsymbol{\zeta}_i$ and μ_i unrestricted.

[2]The objective function to optimize the parameters by means of NLS is the usual sum of squared residuals, i.e. $\sum_{t=3}^{T} \varepsilon_{it}^2$ in (6).