Simulation Design

Monte Carlo Design

Simulate a latent factor model with stochastic volatility for excess return, r_{t+1}, for $t = 1, \ldots, T$:

$$
\begin{aligned}
r_{i,t+1} &= g\left(z_{i,t}\right) + \beta_{i,t+1} v_{t+1} + e_{i,t+1}, \quad z_{i,t} = (1, x_t)' \otimes c_{i,t}, \quad \beta_{i,t} = (c_{i1,t}, c_{i2,t}, c_{i3,t}) \\
e_{i,t+1} &= \exp(\sigma_{i,t+1}/2) \varepsilon_{i,t+1}, \\
\sigma_{i,t+1}^2 &= \omega + \alpha_i e_{i,t+1}^2 + \gamma_i \sigma_{t,i}^2 + w_{i,t+1}.
\end{aligned}
$$

Let v_{t+1} be a 3×1 vector of errors, and $w_{i,t+1}, \varepsilon_{i,t+1}$ scalar error terms. The matrix C_t is an $N \times P_c$ vector of latent factors, where the first three columns correspond to $\beta_{i,t}$, across the $1 \leq i \leq N$ dimensions, while the remaining $P_c - 3$ factors do not enter the return equation. The $P_x \times 1$ vector x_t is a multivariate time series, and ε_{t+1} is a $N \times 1$ vector of idiosyncratic errors.

One of my key concerns with the Gu et al. (2019) design is that the factors are uncorrelated across i, and, in particular, that the factors which do not matter in the return equation are uncorrelated with those that matter. This is not what is observed in practice.

Instead, we will choose a simulation mechanism for C_t that gives some correlation across the factors and across time. To that end, first consider drawing normal random numbers for each $1 \leq i \leq N$ and $1 \leq j \leq P_c$, according to

$$\bar{c}_{ij,t} = \rho_j \bar{c}_{ij,t-1} + \epsilon_{ij,t}, \ \rho_j \mathcal{U}[1/2, 1].$$

Then, define the matrix

$$B := \Lambda\Lambda' + \frac{1}{10}\mathbb{I}_n, \ \Lambda_i = (\lambda_{i1}, \ldots, \lambda_{i4})', \ \lambda_{ik} \sim N(0,1), \ k = 1, \ldots, 4,$$

which we transform into a correlation matrix W via

$$W = \mathrm{diag}^{-1/2}(W) W \mathrm{diag}^{-1/2}(W).$$

To build in cross-sectional correlation, from the $N \times P_c$ matrix $\bar{C}_t$, we simulate characteristics according to

$$\widehat{C}_t = W \overline{C}_t.$$

Finally, we can construct the "observed" characteristics for each $1 \leq i \leq N$ and for $j = 1, \ldots, P_c$

according to

$$c_{ij,t} = \frac{2}{n+1} \operatorname{rank}\left(\bar{c}_{ij,t}\right) - 1.$$

For simulation of x_t we consider a VAR model

$$x_t = A x_{t-1} + u_t,$$

where we have three separate specifications for the matrix A:

$$(1)\ A = \begin{pmatrix} .95 & 0 & 0 \\ 0 & .95 & 0 \\ 0 & 0 & .95 \end{pmatrix} \quad (2)\ A = \begin{pmatrix} 1 & 0 & .25 \\ 0 & .95 & 0 \\ .25 & 0 & .95 \end{pmatrix} \quad (3)\ A = \begin{pmatrix} .99 & .2 & .1 \\ .2 & .90 & -.3 \\ .1 & -.3 & -.99 \end{pmatrix}$$

We will consider four different functions $g(\cdot)$

$$\begin{aligned}
&(1)\ g\left(z_{i,t}\right) = \left(c_{i1,t}, c_{i2,t}, c_{i3,t} \times x_t'\right) \theta_0, \quad \text{where } \theta_0 = (0.02, 0.02, 0.02)' \\
&(2)\ g\left(z_{i,t}\right) = \left(c_{i1,t}^2, c_{i1,t} \times c_{i2,t}, \operatorname{sgn}\left(c_{i3,t} \times x_t'\right)\right) \theta_0, \quad \text{where } \ \theta_0 = (0.04, 0.035, 0.01)' \\
&(3)\ g\left(z_{i,t}\right) = \left(1[c_{i3,t} > 0], c_{i2,t}^3, c_{i1,t} \times c_{i2,t} \times 1[c_{i3,t} > 0], \operatorname{logit}\left(c_{i3,t}\right)\right) \theta_0, \quad \text{where } \ \theta_0 = (0.04, 0.035, 0.01)' \\
&(4)\ g\left(z_{i,t}\right) = \left(\hat{c}_{i1,t}, \hat{c}_{i2,t}, \hat{c}_{i3,t} \times x_t'\right) \theta_0, \quad \text{where } \theta_0 = (0.02, 0.02, 0.02)'
\end{aligned}$$

Need to work out the corresponding cr0ss-sectional R^2 in this case. We can then tune θ^0 to be this close to Gu et al. (2019), as well as the predictive R^2. This will require some work.

Follow Gu et al. (2019) in regards to the choice of N, T, P_c