QFM Exploration Notes

April 8, 2025

1 Location-Scale Model Notes

I consider the following DGP:

$$y_{it} = \alpha_i \beta_t + \eta_i \gamma_t e_{it}$$

We require that $\eta_t \gamma_t > 0$, so we will draw them from the χ^2 distribution. All together we have:

$$\alpha_i \sim \mathcal{N}(0,1); \quad \beta_t \sim \mathcal{N}(0,1); \quad \eta_i \sim \chi^2(1); \quad \gamma_t \sim \chi^2(1); \quad e_{it} \sim \mathcal{N}(0,1)$$

In general, we let F be the cdf of e_{it}. Then the conditional quantile function is:

$$Q_\tau(y_{it}|\alpha_i, \beta_t, \eta_i \gamma_t) = \alpha_i \beta_t + \eta_i \gamma_t F^{-1}(\tau) = \alpha_i \beta_t + \eta_i \gamma_t Q(\tau)$$

We call the estimates of this conditional quantile function from the QPC algorithm $\hat{\alpha}, \hat{\beta}, \hat{\eta}, \hat{\gamma}$. If the algorithm converges to the true values (as captured by the common component), we would expect that the estimated factors span the same space as the true ones. Because the rotation of the factors is not identified, we need to measure this in a way that is invariant to rotation. To do so, we consider the R^2 of the following regressions:

$$\beta = a + b_1 \hat{\beta} + b_2 \hat{\gamma}; \quad \gamma = c + d_1 \hat{\gamma} + d_2 \hat{\gamma};$$

If the algorithm is converging properly, we would expect high values of R^2 in each of these regressions.

Note that under this DGP, beacuse $e_i t$ is $\mathcal{N}(0,1)$ we have that at the median $(\tau = 0.50)$ there will only be one factor.

For now, I will consider the errors-in-variables rotation, where with k_τ factors, the factor loadings are restricted such that $\Lambda = [I_{k_\tau} \Lambda_2]'$. This is the most simple computationally, though there are others I could consider from Bai and Ng (2013).

For now, I want to see how this DGP behaves and wether we can do any estimation consistently. My target is to replicate table 2 from Sagner (2019), shown in Figure 1.

I will not report PC estimates. In addition to the mean R^2 values for both the first and second factor, I will include the proportion of simulations where the process took a long time to converge (> 100 iterations), where it didn't converge at all (not converged at 1000

Figure 1

Table 1.2: Correlation Between Estimated and True Parameters - DGP 2

	PC			QPC ($\tau = 0.25$)			QPC ($\tau = 0.50$)			QPC ($\tau = 0.75$)		
$T \setminus N$	10	50	100	10	50	100	10	50	100	10	50	100
Panel A: First Factor												
50	0.9418	0.9908	0.9947	0.9078	0.9725	0.9900	0.9147	0.9868	0.9928	0.9129	0.9751	0.9910
100	0.9411	0.9910	0.9951	0.8441	0.8937	0.9566	0.9175	0.9862	0.9928	0.8175	0.9280	0.9748
200	0.9421	0.9909	0.9952	0.8736	0.9677	0.9874	0.9164	0.9861	0.9927	0.8824	0.9462	0.9812
1000	0.9426	0.9909	0.9950	0.8640	0.9509	0.9861	0.9118	0.9859	0.9926	0.8612	0.9478	0.9845
Panel B: Second Factor												
50				0.4640	0.6617	0.8567				0.5641	0.8209	0.9088
100				0.5194	0.7112	0.8465				0.5479	0.7142	0.8739
200				0.5561	0.8114	0.8594				0.5513	0.7484	0.8321
1000				0.5666	0.7660	0.8999				0.5151	0.7827	0.8494

This table reports the average correlation between the QPC estimators $\left\{\hat{\beta}_t^{(s)}(\tau)\right\}_{t=1}^{T}$ and $\left\{\hat{\gamma}_t^{(s)}(\tau)\right\}_{t=1}^{T}$, the PC estimators $\left\{\tilde{\beta}_t^{(s)}\right\}_{t=1}^{T}$, and their true counterparts $\{\beta_t\}_{t=1}^{T}$ and $\{\gamma_t\}_{t=1}^{T}$, respectively, for $N = \{10, 50, 100\}$, $T = \{50, 100, 200, 1000\}$, $\tau = \{0.25, 0.50, 0.75\}$ and $s = 1, \ldots, 1000$ simulations. The average correlation was computed as $\bar{\rho}_X = S^{-1} \sum_{s=1}^{S} \rho(\hat{X}^{(s)}, X)$, where $\hat{X}^{(s)}$ is an estimator (QPC or PC), and X is its true counterpart.

iterations), the proportion of simulations where the R^2 for the second factor is > 0.5, and the proportion of simulations where the R^2 of the first factor is < 0.9. For small N and T this last value may be sizable, but for larger N and T a large value of this proportion is concerning as it implies that the first factor is not being estimated consistently.

I want to first check that my estimation procedure is working correctly. To do so, I will consider 2 simple DGPs:

$$y_{it} = \alpha_i \beta_t + e_{it} \tag{1}$$

$$y_{it} = \alpha_i \beta_t + \eta_i \gamma_t + e_{it} \tag{2}$$

These DGPs are simple location models, and so we should expect a good fit for both the one factor case (Equation 1) and the two factor case (Equation 2). Results of a small simulation study are reported in.

Now that the simulation is behaving as anticipated, I will explore the question of initial values. For this study, I consider the following DGP:

$$y_{it} = \alpha_i \beta_t + \eta_i \gamma_t e_{it} \tag{3}$$

Where $\eta_i, \gamma_t \sim \chi^2_1$ to satisfy $\eta_i \gamma_t > 0$. I fit a QFM using the same procedure as before, but with a change in the initialization step. Instead of beginning with PCA to guess the initial values, I add noise from a $\mathcal{N}(0,1)$ to the initial PCA estimate. This is then rotated according to the rotation I have used. The model is fit from these new initial values, and some interesting patterns appear. Consider the example in table 1

TABLE 1: QFM FITS WITH DIFFERENT NOISE

Noise Seed	First Factor Fit	Second Factor Fit	Iterations	Objective Function Value
No Noise	0.9921	0.1600	14	3909.9828
134935	0.9912	0.7791	33	3595.2488
363439	0.9924	0.7245	18	3645.0155
880628	0.9927	0.1601	24	4005.6856
252318	0.9954	0.7844	84	3677.4222
344982	0.9965	0.7391	16	3591.6591
702677	0.9920	0.8051	54	3555.9745
116075	0.9910	0.1600	19	3915.6519
100298	0.9934	0.7755	25	3667.7547
178700	0.9943	0.1599	28	3940.8150
310893	0.9903	0.7956	28	3564.4116

References

Bai, J., and S. Ng (2013) “Principal components estimation and identification of static factors,” *Journal of Econometrics*, 176(1), 18–29.

Sagner, A. G. (2019) “Three Essays on Quantile Factor Analysis,” .