Estimating Factor Models under Price Limits

ECE 5412 Course Project, Jinyuan (John) Yu, 05/2022, NetID: jy478

Abstract

The factor model is an important tool for both portfolio managers and researchers in modern finance. In this report, we illustrate a Bayesian approach using Gibbs sampling to estimate linear regression models in which the stock returns are subject to a price limit regulation. The price limit introduces a censored scheme in this case.

(The source codes: https://github.com/JohnYu0510/ECE5412/tree/main/CensoredModel*)*

1 Introduction

Linear factor models for asset returns play an essential role in modern finance. Such models describe the return on each particular asset in a market as varying linearly with some exogeneous factors. For a single asset, the model has a time-series regression form:

$$r_t = f_t\beta + \epsilon_t$$

In the above function, y_t is the vector of history asset returns, f_t is a matrix containing predictive signals (e.g., the market index), and ϵ_t is a white noise process. The intercept of the regression can be modeled by adding a constant column to the factor matrix. Here we want to estimate the coefficient β as well as the variance of the residuals, σ^2.

Traditionally methods like OLS can yield an unbiased estimate. But when there is a price limit regulation on the market, the observed stock returns will be censored and we may need to apply the Bayesian approach proposed by to infer the true distribution of them.

A price limit is a rule to restrict the fluctuation of security prices within a predetermined range in each trading day. This regulation has been imposed on the trading of futures and options in U.S., and the trading of stocks in many Asian countries. We will show how the price limits can censor the returns data and affect the estimation.

2 The Factor Model under Price Limits

2.1 The Censoring Scheme

Let r_t denote the "true" returns of a stock that is modeled by the linear form above, and z_t denote the "observed" returns that can be collected from the data. Suppose the upper bound and the lower bound on the returns' fluctuation are l_u, l_d. The observed returns follow the censoring scheme:

$$z_t = \begin{cases} l_u, & if\ r_t + E_{t-1} \geq l_u \\ r_t + E_{t-1}, & if\ l_d \leq r_t + E_{t-1} \leq l_u \\ l_d, & if\ r_t + E_{t-1} \leq l_d \end{cases}$$

Here, $E_s = \Sigma_{k=1}^{s}(r_k - z_k)$ is the unrealized residual shock, i.e., if a partition of the true return is not realized due to the price limits, it will be added to next day's observed returns. As the unrealized shocks are carried over from previous days, the returns are actually autocorrelated and the standard method like two-limit Tobit model (which is based on MLE) can not be applied to estimate the parameters in the factor model.

2.2 Clustering the observations

Before moving forward, we first split T observations into several different groups. In each group, there should be: (i) A single non-limit day. (ii) Consecutive limit day and a subsequent single non-limit day. This ensures that in each group, there will be no residual shocks at the beginning. Each observation is thus assigned a new index (i,j), which denotes the j-th observation in the i-th group.

2.3 The Full Conditional Posterior Distribution

Now we estimate the true parameters in the factor model with a Bayesian method involving Gibbs sampling. For simplicity, we assume the prior distribution of the parameters takes the following form:

$$\beta \sim N(b, B^{-1});\ \sigma^2 \sim IG(v/2, \delta/2)$$

The decision maker can choose proper numbers for b, B, v, δ to reflect his belief in the prior information.

Given the parameters in the above assumption and data of factors and observed returns, the conditional posterior distribution of β, σ^2, and the true returns r are given by:

$$\beta | r, f, \sigma^2, b, B \sim N(\hat{\beta}, \hat{\Omega}), where\ \hat{\Omega} = \left(B + \frac{f'f}{\sigma^2}\right)^{-1}, \hat{\beta} = \hat{\Omega}(Bb + f'r/\sigma^2)$$

$$\sigma^2 | \beta, r, f, v, \delta \sim IG((v+T)/2, (\delta + (r - f\beta)'(r - f\beta)/2))$$

$$r_{ij} | \beta, \sigma^2, Z_i, f, r_{i,\sim j} \sim N(f_{ij}\beta, \sigma^2) I_{A_{i,\sim j}}\ with\ T_i > 1, j = 1, \ldots, T_i - 1$$

$$r_{ij} | \beta, \sigma^2, Z_i, f, r_{i,\sim T_i} = \Sigma_{k=1}^{T_i} z_{ik} - \Sigma_{k=1}^{T_i - 1} r_{ik}\ with\ T_i > 1$$

Here $A_{i,\sim j}$ is the support of r_{ij} given the information of all other true returns in the same cluster, $r_{i,\sim j}$, which takes the form $A_{i,\sim j} = \{r_{ij} | \underline{r_{ij}} \le r_{ij} \le \overline{r_{ij}}\}$, where $\underline{r_{ij}}, \overline{r_{ij}}$ are defined as the following:

$$\underline{r_{ij}} = \begin{cases} -\infty, & if\ \Sigma_{k=j}^{T_i - 1} I_{(z_{ik} = l_u)} = 0 \\ \inf \bigcap_{k=j}^{T_i - 1} \{r_{ij} : r_{ik} + E_{i,k-1} > l_u, z_{ik} = l_u\} & otherwise \end{cases}$$

$$\overline{r_{ij}} = \begin{cases} \infty, & if\ \Sigma_{k=j}^{T_i - 1} I_{(z_{ik} = l_d)} = 0 \\ \sup \bigcap_{k=j}^{T_i - 1} \{r_{ij} : r_{ik} + E_{i,k-1} > l_d, z_{ik} = l_d\} & otherwise \end{cases}$$

So, the posterior distribution of r_{ij} follows a truncated normal if there are price limits in the subsequent days in the same group.

2.4 Gibbs Sampling Approach

Based on the results above, the procedure implementing Gibbs sampling algorithm to estimate the factor model is described below:

1. Initialize $\beta^{(0)} = (f'f)^{-1} f'Z, \sigma^{2(0)} = \frac{(Z - f\beta^{(0)})'(Z - f\beta^{(0)})}{T}, q = 0;$
2. Sequentially simulate draws from the posterior distribution:
   (i) $\beta^{(q+1)}$ from $\beta | r^{(q)}, f, \sigma^{2(q)}, b, B;$
   (ii) $\sigma^{2(q+1)}$from $\sigma^2 | \beta^{(q+1)}, r^{(q)}, f, v, \delta$
   (iii) $r_{ij}^{(q+1)}$from $r_{ij} | \beta^{(q+1)}, \sigma^{2(q+1)}, f_{ij}, \left\{r_{ik}^{(q+1)}\right\}_{k=1}^{j-1}, \left\{r_{ik}^{(q)}\right\}_{k=j+1}^{T_i - 1}$
3. Set $r_{iT_i}^{(q+1)} = \Sigma_{k=1}^{T_i} z_{ik} - \Sigma_{k=1}^{T_i - 1} r_{ik}^{(q+1)}, q = q + 1$, then go to step 2 and iterate the procedure until G times.

The procedure first iterates a large number of times, say M times, to converge to the true posterior distribution, and all other simulated samples after that will be taken to compute the estimate of parameters β, σ^2.

3 A Numerical Test with Simulation

We now evaluate the performance of the above algorithm with a simulation study. We consider a one-factor model with an intercept. The factor is the market index returns which are simulated from a normal distribution $N(\mu_f, \Omega_f)$, where the mean is 0.5% and the standard deviation is 0.89%. The market has a price limit of $l_u = 10\%, l_d = -10\%$.

A single stock with $\beta = [0\ 1]^T, \sigma = 8\%$ is considered. We simulated daily returns following the factor model for T=1,000 observations. Then the observed data are censored based on the mechanism showed above.

For the simplicity, we choose noninformative priors for the algorithm to allow the data to determine posterior conclusions. B is now a zero matrix (which also means b will no longer have effects on β), and we use $v = 1, \delta = 0.1$ to provide comparatively diffuse priors for σ^2. The Gibbs sampling algorithm iterates $G = 2000$ times and discards the first $B = 500$ samples to make sure the distribution has converged.

The following table shows the estimation results. The "Non-Censored" column shows the OLS estimates given the true returns data. We mainly evaluate the performance from the last two columns, which estimate the model based on the observed returns data.

Table 1 Estimation Results with Simulated Data

	True Value	Non-Censored	OLS estimates	Gibbs estimates
β_0	0	0.0006	0.0012	0.0003
β_1	1	0.8987	0.7733	0.9548
σ^2	0.0064	0.0063	0.0043	0.0073

As is shown above, the Gibbs sampler proposed in this report provided a more accurate estimate for the true estimates. Its deviation from the true β is much lower than that of the tradition OLS. As OLS estimation doesn't take the censored property of the data into account, it should always yield a biased lower estimate for $\beta_1 > 0$. By contrast, the Gibbs sampling method can generate asymptotic unbiased estimates for the true parameters.

4 Conclusion

In summary, when there are price limits on the market which follow a residual shocks mechanism, the observed returns of stocks will be censored, and the tradition regression to estimate the parameters in a factor model will be thus biased. In this case, Gibbs sampling can be used to draw samples from the posterior distribution of the parameters as well as the true returns, and gives a better estimate.

Reference

[1] Chou, 1997. A Gibbs sampling approach to the estimation of linear regression models under daily price limits. Pacific-Basin Finance Journal, 5 (1997), pp. 39-62