Bayesian Analysis in Financial Econometrics: Value at Risk and High-Frequency Data

Ruey S. Tsay
Graduate School of Business, University of Chicago
Chicago, IL 60637, USA
ruey.tsay@gsb.uchicago.edu

1 Introduction

This lecture consists of two topics in Bayesian analysis of financial econometrics. The first topic is Value at Risk (VaR) calculation based on a conditional extreme value theory (EVT). Consider the log return of a financial position. Conditional extreme value theory focuses on the occurrence that the return exceeds a given high threshold and the magnitude of the exceedance when it occurs. The frequency of exceeding the threshold and the exceedance are assumed to form a two-dimensional Poisson process. With a properly specified intensity function, the Poisson process provides a direct link to the extreme value theory. The two-dimensional Poisson process may be nonhomogeneous because it can depends on explanatory variables that are relevant to the financial position. Thus, the conditional approach to EVT discussed is adaptive to the changing market conditions. Details are discussed in the lecture.

The second topic of the lecture is Bayesian analysis of high-frequency financial data. By high-frequency data we mean observations taken with time intervals less than or equal to 24 hours. In the equity markets, such as the New York Stock Exchange (NYSE), trade-by-trade data commonly known as the *transactions* data are available; see the Trades and Quotes (TAQ) database of NYSE. These data are useful in the empirical study of market microstructure. In this lecture, we discuss characteristics of high-frequency financial data and econometric methods that are useful in analyzing the data. Our focus is on Bayesian analysis via hierarachical models for modeling the bivariate series of price changes and time durations between price changes.

2 Value at Risk based on conditional EVT

We begin the lecture with a brief review of extreme value theory. Both traditional and conditional EVT are discussed. The materials discussed, including empirical examples, are from Tsay (2001, chapter 7).

We use daily log returns of IBM stock to illustrate the actual calculation of all the methods discussed. The results obtained can therefore be used to compare the performance of different methods. Figure 1 shows the time plot of daily log returns of IBM stock from July 3, 1962 to December 31, 1998 for 9190 observations.

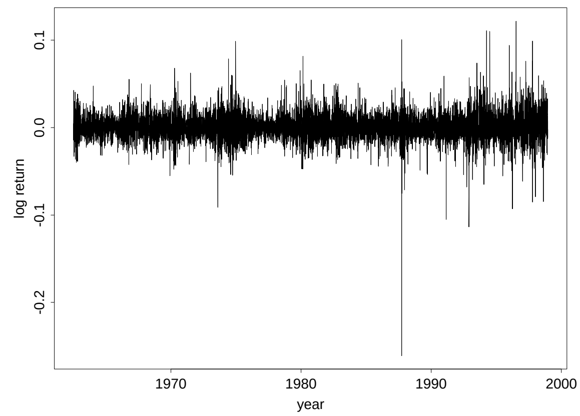

Figure 1: Time plot of daily log returns of IBM stock from July 3, 1962 to December 31, 1998.

3 Value at Risk

We define VaR under a probabilistic framework. Suppose that at the time index t we are interested in the risk of a financial position for the next ℓ periods. Let $\Delta V(\ell)$ be the change in value of the assets in the financial position from time t to $t+\ell$. This quantity is measured in dollars and is a random variable at the time index t. Denote the cumulative distribution function (CDF) of $\Delta V(\ell)$ by $F_\ell(x)$. We define the VaR of a long position over the time horizon ℓ with probability p as

$$p = Pr[\Delta V(\ell) \leq \text{VaR}] = F_\ell(\text{VaR}). \tag{3.1}$$

Since the holder of a long financial position suffers a loss when $\Delta V(\ell) < 0$, the VaR defined in Eq. (3.1) typically assumes a negative value when p is small. The negative sign signifies a loss. From the definition, the probability that the holder would encounter a loss greater than or equal to VaR over the time horizon ℓ is p. Alternatively, VaR can be interpreted as follows. With probability $(1-p)$, the potential loss encountered by the holder of the financial position over the time horizon ℓ is less than or equal to VaR.

The holder of a short position suffers a loss when the value of the asset increases [i.e., $\Delta V(\ell) > 0$]. The VaR is then defined as

$$p = Pr[\Delta V(\ell) \geq \text{VaR}] = 1 - Pr[\Delta V(\ell) \leq \text{VaR}] = 1 - F_\ell(\text{VaR}).$$

For a small p, the VaR of a short position typically assumes a positive value. The positive sign signifies a loss.

The previous definitions show that VaR is concerned with tail behavior of the CDF $F_\ell(x)$. For a long position, the left tail of $F_\ell(x)$ is important. Yet a short position focuses on the right tail of $F_\ell(x)$. Notice that the definition of VaR in Eq. (3.1) continues to apply to a short position if one uses the distribution of $-\Delta V(\ell)$. Therefore, it suffices to discuss methods of VaR calculation using a long position.

Fo any univariate CDF $F_\ell(x)$ and probability p, such that $0 < p < 1$, the quantity

$$x_p = \inf\{x | F_\ell(x) \geq p\}$$

is called the pth quantile of $F_\ell(x)$, where inf denotes the smallest real number satisfying $F_\ell(x) \geq p$. If the CDF $F_\ell(x)$ of Eq. (3.1) is known, then VaR is simply its pth quantile (i.e., VaR $= x_p$). The CDF is unknown in practice, however. Studies of VaR are essentially concerned with estimation of the CDF and/or its quantile, especially the tail behavior of the CDF.

In practical applications, calculation of VaR involves several factors:

1. The probability of interest p such as $p = 0.01$ or $p = 0.05$.
2. The time horizon ℓ. It might be set by a regulatory committee, such as 1 day or 10 days.
3. The frequency of the data, which might not be the same as the time horizon ℓ. Daily observations are often used.
4. The CDF $F_\ell(x)$ or its quantiles.
5. The amount of the financial position or the mark-to-market value of the portfolio.

Among these factors, the CDF $F_\ell(x)$ is the focus of econometric modeling. Different methods for estimating the CDF give rise to different approaches to VaR calculation.

4 Extreme value theory

4.1 Review of extreme value theory

Assume that the returns r_t are serially independent with a common cumulative distribution function $F(x)$ and that the range of the return r_t is $[l, u]$. For log returns, we have $l = -\infty$ and $u = \infty$. Then the CDF of $r_{(1)}$, denoted by $F_{n,1}(x)$, is given by

$$\begin{aligned} F_{n,1}(x) &= Pr[r_{(1)} \leq x] = 1 - Pr[r_{(1)} > x] \\ &= 1 - Pr(r_1 > x, r_2 > x, \cdots, r_n > x) \\ &= 1 - \prod_{j=1}^{n} Pr(r_j > x), \quad \text{(by independence)} \\ &= 1 - \prod_{j=1}^{n} [1 - Pr(r_j \leq x)] \\ &= 1 - \prod_{j=1}^{n} [1 - F(x)] \quad \text{(by common distribution)} \\ &= 1 - [1 - F(x)]^n. \end{aligned} \tag{4.2}$$

In practice, the CDF $F(x)$ of r_t is unknown and, hence, $F_{n,1}(x)$ of $r_{(1)}$ is unknown. However, as n increases to infinity, $F_{n,1}(x)$ becomes degenerated – namely, $F_{n,1}(x) \to 0$ if $x \leq l$ and $F_{n,1}(x) \to 1$ if $x > l$ as n goes to infinity. This degenerated CDF has no practical value. Therefore, the extreme value theory is concerned with finding two sequences $\{\beta_n\}$ and $\{\alpha_n\}$, where $\alpha_n > 0$, such that the distribution of $r_{(1*)} \equiv (r_{(1)} - \beta_n)/\alpha_n$ converges to a nondegenerated distribution as n goes to infinity. The sequence $\{\beta_n\}$ is a location series and $\{\alpha_n\}$ is a series of scaling factors. Under the independent assumption, the limiting distribution of the normalized minimum $r_{(1*)}$ is given by

$$F_*(x) = \begin{cases} 1 - \exp[-(1 + kx)^{1/k}] & \text{if } k \neq 0 \\ \\ 1 - \exp[-\exp(x)] & \text{if } k = 0 \end{cases} \tag{4.3}$$

for $x < -1/k$ if $k < 0$ and for $x > -1/k$ if $k > 0$, where the subscript $*$ signifies the minimum. The case of $k = 0$ is taken as the limit when $k \to 0$. The parameter k is referred to as the *shape parameter* that governs the tail behavior of the limiting distribution. The parameter $\alpha = -1/k$ is called the *tail index* of the distribution.

The limiting distribution in Eq. (4.3) is the *generalized extreme value distribution* of Jenkinson (1955) for the minimum. It encompasses the three types of limiting distribution of Gnedenko (1943):

- Type I: $k = 0$, the Gumbel family. The CDF is

$$F_*(x) = 1 - \exp[-\exp(x)], \quad -\infty < x < \infty. \tag{4.4}$$

- Type II: $k < 0$, the Fréchet family. The CDF is

$$F_*(x) = \begin{cases} 1 - \exp[-(1 + kx)^{1/k}] & \text{if } x < -1/k \\ 1 & \text{otherwise.} \end{cases} \tag{4.5}$$

- Type III: $k > 0$, the Weibull family. The CDF here is

$$F_*(x) = \begin{cases} 1 - \exp[-(1 + kx)^{1/k}] & \text{if } x > -1/k \\ 0 & \text{otherwise.} \end{cases}$$

Gnedenko (1943) gave necessary and sufficient conditions for the CDF $F(x)$ of r_t to be associated with one of the three types of limiting distribution. Briefly speaking, the tail behavior of $F(x)$ determines the limiting distribution $F_*(x)$ of the minimum. The (left) tail of the distribution declines exponentially for the Gumbel family, by a power function for the Fréchet family, and is finite for the Weibull family. Readers are referred to Embrechts, Kuppelberg, and Mikosch (1997) for a comprehensive treatment of the extreme value theory. For risk management, we are mainly interested in the Fréchet family that includes stable and Student-t distributions. The Gumbel family consists of thin-tailed distributions such as normal and log-normal distributions. The probability density function (pdf) of the generalized limiting distribution in Eq. (4.3) can be obtained easily by differentiation:

$$f_*(x) = \begin{cases} (1 + kx)^{1/k-1} \exp[-(1 + kx)^{1/k}] & \text{if } k \neq 0 \\ \\ \exp[x - \exp(x)] & \text{if } k = 0, \end{cases} \tag{4.6}$$

where $-\infty < x < \infty$ for $k = 0$, $x < -1/k$ for $k < 0$, and $x > -1/k$ for $k > 0$.

The aforementioned extreme value theory has two important implications. First, the tail behavior of the CDF $F(x)$ of r_t, not the specific distribution, determines the limiting distribution $F_*(x)$ of the (normalized) minimum. Thus, the theory is generally applicable to a wide range of distributions for the return r_t. The sequences $\{\beta_n\}$ and $\{\alpha_n\}$, however, may depend on the CDF $F(x)$. Second, Feller (1971, p. 279) shows that the tail index k does not depend on the time interval of r_t. That is, the tail index (or equivalently the shape parameter) is invariant under time aggregation. This second feature of the limiting distribution becomes handy in the VaR calculation.

The extreme value theory has been extended to serially dependent observations $\{r_t\}_{t=1}^n$ provided that the dependence is weak. Berman (1964) shows that the same form of the limiting extreme value distribution holds for stationary normal sequences provided that the autocorrelation function of r_t is squared summable (i.e., $\sum_{i=1}^\infty \rho_i^2 < \infty$), where ρ_i is the lag-i autocorrelation function of r_t. For further results concerning the effect of serial dependence on the extreme value theory, readers are referred to Leadbetter, Lindgren, and Rootzén (1983, Chapter 3).

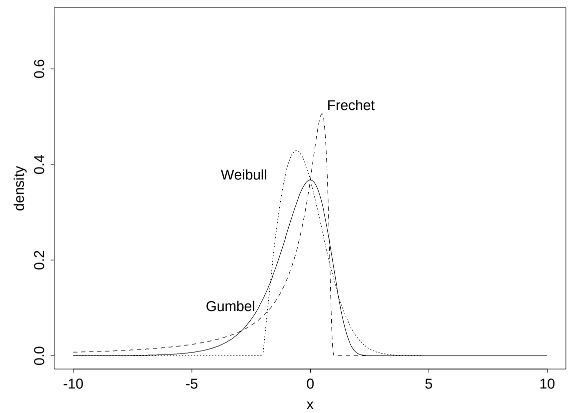

Figure 2: Probability density functions of extreme value distributions for minimum: The solid line is for a Gumbel distribution, the dotted line is for the Weibull distribution with $k = 0.5$, and the dashed line for the Fréchet distribution with $k = -0.9$.

4.2 Empirical estimation

The extreme value distribution contains three parameters – k, β_n, and α_n. These parameters are referred to as the *shape*, *location*, and *scale parameter*, respectively. They can be estimated by using either parametric or nonparametric methods. We use the maximum likelihood method to estimate the three parameters.

For a given sample, there is only a single minimum or maximum, and we can not estimate the three parameters with only an extreme observation. Alternative ideas must be used. One of the ideas used in the literature is to divide the sample into subsamples and apply the extreme value theory to the subsamples. Assume that there are T returns $\{r_j\}_{j=1}^{T}$ available. We divide the sample into g non-overlapping subsamples each with n observations, assuming for simplicity that $T = ng$. In other words, we divide the data as

$$\{r_1, \cdots, r_n | r_{n+1}, \cdots, r_{2n} | r_{2n+1}, \cdots, r_{3n} | \cdots | r_{(g-1)n+1}, \cdots, r_{ng}\}$$

and write the observed returns as r_{in+j}, where $1 \leq j \leq n$ and $i = 0, \cdots, g-1$. Notice that each subsample corresponds to a subperiod of the data span. When n is sufficiently large, we hope that the extreme value theory applies to each subsample. In application, the choice of n can be guided by practical considerations. For example, for daily returns, $n = 21$ corresponds approximately to the number of trading days in a month and $n = 63$ denotes the number of trading days in a quarter.

Let $r_{n,i}$ be the minimum of the ith subsample (i.e., $r_{n,i}$ is the smallest return of the ith subsample), where the subscript n is used to denote the size of the subsample. When n is sufficiently large, $x_{n,i} = (r_{n,i} - \beta_n)/\alpha_n$ should follow an extreme value distribution, and the collection of subsample minima $\{r_{n,i} | i = 1, \cdots, g\}$ can then be regarded as a sample of g observations from that extreme value distribution. Specifically, we define

$$r_{n,i} = \min_{1 \leq j \leq n} \{r_{(i-1)n+j}\}, \quad i = 1, \cdots, g. \tag{4.7}$$

The collection of subsample minima $\{r_{n,i}\}$ are the data we use to estimate the unknown parameters of the extreme value distribution. Clearly, the estimates obtained may depend on the choice of subperiod length n.

Maximum likelihood method Assuming that the subperiod minima $\{r_{n,i}\}$ follow a generalized extreme value distribution such that the pdf of $x_i = (r_{n,i} - \beta_n)/\alpha_n$ is given in Eq. (4.6), we can obtain the pdf of $r_{n,i}$ by a simple transformation as

$$f(r_{n,i}) =$$

$$\begin{cases} \frac{1}{\alpha_n}\left[1+\frac{k_n(r_{n,i}-\beta_n)}{\alpha_n}\right]^{1/k_n-1} \exp\left\{-\left[1+\frac{k_n(r_{n,i}-\beta_n)}{\alpha_n}\right]^{1/k_n}\right\} & \text{if } k_n \neq 0 \\ \frac{1}{\alpha_n}\exp\left\{\frac{r_{n,i}-\beta_n}{\alpha_n} - \exp\left[\frac{r_{n,i}-\beta_n}{\alpha_n}\right]\right\} & \text{if } k_n = 0, \end{cases}$$

where it is understood that $1 + k_n(r_{n,i} - \beta_n)/\alpha_n > 0$ if $k_n \neq 0$. The subscript n is added to the shape parameter k to signify that its estimate depends on the choice of n. Under the independence assumption, the likelihood function of the subperiod minima is

$$\ell(r_{n,1}, \cdots, r_{n,g}|k_n, \alpha_n, \beta_n) = \prod_{i=1}^{g} f(r_{n,i}).$$

Nonlinear estimation procedures can then be used to obtain maximum likelihood estimates of k_n, β_n, and α_n. These estimates are unbiased, asymptotically normal, and of minimum variance under proper assumptions.

4.3 Application to stock returns

We apply the extreme value theory to the daily log returns of IBM stock from July 3, 1962 to December 31, 1998. The returns are measured in percentages, and the sample size is 9190 (i.e., $T = 9190$). Figure 3 shows the time plots of extreme daily log returns when the length of the subperiod is 21, which corresponds approximately to a month. The October 1987 crash is clearly seen from the plot. Excluding the 1987 crash, the range of extreme daily log returns is between 0.5% and 13%.

We apply the maximum likelihood method to estimate parameters of the generalized extreme value distribution for IBM daily log returns. Table 1 summarizes the estimation results for different choices of the length of subperiods ranging from 1 month ($n = 21$) to 1 year ($n = 252$). From the table, we make the following observations:

- Estimates of the location and scale parameters β_n and α_n increase in modulus as n increases. This is expected as magnitudes of the subperiod minimum and maximum are nondecreasing functions of n.
- Estimates of the shape parameter (or equivalently the tail index) are stable for the negative extremes when $n \geq 63$ and are approximately -0.33.
- Estimates of the shape parameter are less stable for the positive extremes. The estimates are smaller in magnitude, but remain significantly different from zero.
- The results for $n = 252$ have higher variabilities as the number of subperiods g is relatively small.

Again the conclusion obtained is similar to that of Longin (1996), who provided a good illustration of applying the extreme value theory to stock market returns.

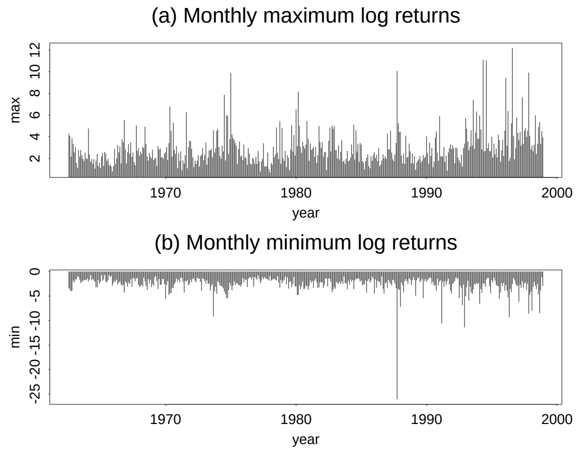

Figure 3: Maximum and minimum daily log returns of IBM stock when the subperiod is 21 trading days. The data span is from July 3, 1962 to December 31, 1998: (a) positive returns, and (b) negative returns.

Table 1: Maximum Likelihood Estimates of The Extreme Value Distribution For Daily Log Returns of IBM Stock From July 3, 1962 to December 31, 1998. Standard Errors Are in Parentheses.

Length of subperiod	Scale α_n	Location β_n	Shape Par. k_n
(a) Minimal returns			
1 mon.(n = 21, g = 437)	0.823(0.035)	−1.902(0.044)	−0.197(0.036)
1 qur(n = 63, g =145)	0.945(0.077)	−2.583(0.090)	−0.335(0.076)
6 mon.(n = 126, g = 72)	1.147(0.131)	−3.141(0.153)	−0.330(0.101)
1 year(n = 252, g = 36)	1.542(0.242)	−3.761(0.285)	−0.322(0.127)
(b) Maximal returns			
1 mon.(n = 21, g = 437)	0.931(0.039)	2.184(0.050)	−0.168(0.036)
1 qur(n = 63, g = 145)	1.157(0.087)	3.012(0.108)	−0.217(0.066)
6 mon.(n = 126, g = 72)	1.292(0.158)	3.471(0.181)	−0.349(0.130)
1 year(n = 252, g = 36)	1.624(0.271)	4.475(0.325)	−0.264(0.186)

5 An extreme value approach to VaR

In this section, we discuss an approach to VaR calculation using the extreme value theory. The approach is similar to that of Longin (1999a, 1999b), who proposed an eight-step procedure for the same purpose. We divide the discussion into two parts. The first part is concerned with parameter estimation using the method discussed in the previous subsections. The second part focuses on VaR calculation by relating the probabilities of interest associated with different time intervals.

Part I Assume that there are T observations of an asset return available in the sample period. We partition the sample period into g nonoverlapping subperiods of length n such that $T = ng$. If $T = ng + m$ with $1 \leq m < n$, then we delete the first m observations from the sample. The extreme value theory discussed in the previous section enables us to obtain estimates of the location, scale, and shape parameters β_n, α_n, and k_n for the subperiod minima $\{r_{n,i}\}$. Plugging the MLE estimates into the CDF in Eq. (4.3) with $x = (r - \beta_n)/\alpha_n$, we can obtain the quantile of a given probability of the generalized extreme value distribution. Because we focus on holding a long financial position, the lower probability (or left) quantiles are of interest. Let p^* be a small probability that indicates the potential loss of a long position and r_n^* be the p^*th quantile of the subperiod minimum under the limiting generalized extreme value distribution. Then we have

$$p^* = \begin{cases} 1 - \exp\left[-\left(1 + \frac{k_n(r_n^* - \beta_n)}{\alpha_n}\right)^{1/k_n}\right] & \text{if } k_n \neq 0 \\ 1 - \exp\left[-\exp\left(\frac{r_n^* - \beta_n}{\alpha_n}\right)\right] & \text{if } k_n = 0, \end{cases}$$

where it is understood that $1 + k_n(r_n^* - \beta_n)/\alpha_n > 0$ for $k_n \neq 0$. Rewriting this equation as

$$\ln(1 - p^*) = \begin{cases} -\left[1 + \frac{k_n(r_n^* - \beta_n)}{\alpha_n}\right]^{1/k_n} & \text{if } k_n \neq 0 \\ -\exp\left[\frac{r_n^* - \beta_n}{\alpha_n}\right] & \text{if } k_n = 0, \end{cases}$$

we obtain the quantile as

$$r_n^* = \begin{cases} \beta_n - \frac{\alpha_n}{k_n}\left\{1 - [-\ln(1 - p^*)]^{k_n}\right\} & \text{if } k_n \neq 0 \\ \beta_n + \alpha_n \ln[-\ln(1 - p^*)] & \text{if } k_n = 0. \end{cases} \tag{5.8}$$

In financial applications, the case of $k_n \neq 0$ is of major interest.

Part II For a given lower (or left tail) probability p^*, the quantile r_n^* of Eq. (5.8) is the VaR based on the extreme value theory for the subperiod minima. The next step is to make explicit the relationship between subperiod minima and the observed return r_t series.

Because most asset returns are either serially uncorrelated or have weak serial correlations, we may use the relationship in Eq. (4.2) and obtain

$$p^* = P(r_{n,i} \leq r_n^*) = 1 - [1 - P(r_t \leq r_n^*)]^n$$

or, equivalently,

$$1 - p^* = [1 - P(r_t \leq r_n^*)]^n. \tag{5.9}$$

This relationship between probabilities allows us to obtain VaR for the original asset return series r_t. More precisely, for a specified small lower probability p, the pth quantile of r_t is r_n^* if the probability p^* is chosen based on Eq. (5.9), where $p = P(r_t \le r_n^*)$. Consequently, for a given small probability p, the VaR of holding a long position in the asset underlying the log return r_t is

$$\text{VaR} = \begin{cases} \beta_n - \frac{\alpha_n}{k_n}\left\{1 - [-n\ln(1-p)]^{k_n}\right\} & \text{if } k_n \neq 0 \\ \beta_n + \alpha_n \ln[-n\ln(1-p)] & \text{if } k_n = 0. \end{cases} \tag{5.10}$$

Summary We summarize the approach of applying the traditional extreme value theory to VaR calculation as follows:

1. Select the length of the subperiod n and obtain subperiod minima $\{r_{n,i}\}$, $i = 1, \cdots, g$, where $g = T/n$.
2. Obtain the maximum likelihood estimates of β_n, α_n, and k_n.
3. Check the adequacy of the fitted extreme value model; see the next section for some methods of model checking.
4. If the extreme value model is adequate, apply Eq. (5.10) to calculate VaR.

Remark: Since we focus on holding a long financial position and, hence, on the quantile in the left tail of a return distribution, the quantile is negative. Yet it is customary in practice to use a positive number for VaR calculation. Thus, in using Eq. (5.10), one should be aware that the negative sign signifies a loss.

Example 1. (Example 7.6 of Tsay, 2001). Consider the daily log return, in percentage, of IBM stock from July 7, 1962 to December 31, 1998. From Table 1, we have $\hat{\alpha}_n = 0.945$, $\hat{\beta}_n = -2.583$, and $\hat{k}_n = -0.335$ for $n = 63$. Therefore, for the left-tail probability $p = 0.01$, the corresponding VaR is

$$\begin{aligned} \text{VaR} &= -2.583 - \frac{0.945}{-0.335}\left\{1 - [-63\ln(1-0.01)]^{-0.335}\right\} \\ &= -3.04969. \end{aligned}$$

Thus, for daily log returns of the stock, the 1% quantile is -3.04969. If one holds a long position on the stock worth \$10 million, then the estimated VaR with probability 1% is $\$10,000,000 \times 0.0304969$ = \$304,969. If the probability is 0.05, then the corresponding VaR is \$166,641.

If we chose $n = 21$ (i.e., approximately 1 month), then $\hat{\alpha}_n = 0.823$, $\hat{\beta}_n = -1.902$, and $\hat{k}_n = -0.197$. The 1% quantile of the extreme value distribution is

$$\text{VaR} = -1.902 - \frac{0.823}{-0.197}\{1 - [-21\ln(1-0.01)]^{-0.197}\} = -3.40013.$$

Therefore, for a long position of \$10,000,000, the corresponding 1-day horizon VaR is \$340,013 at the 1% risk level. If the probability is 0.05, then the corresponding VaR is \$184,127. In this particular case, the choice of $n = 21$ gives higher VaR values.

It is somewhat surprising to see that the VaR values obtained in Example 1 using the extreme value theory are smaller than those based on a GARCH(1,1) model; see Example 7.3 of Tsay (2001).

In fact, the VaR values of Example 1 are even smaller than those based on the empirical quantile; see Example 7.5 of Tsay (2001). This is due in part to the choice of probability 0.05. If one chooses probability 0.001 = 0.1% and considers the same financial position, then we have VaR = \$546,641 for the Gaussian AR(2)-GARCH(1,1) model and VaR = \$666,590 for the extreme value theory with $n = 21$. Furthermore, the VaR obtained here via the traditional extreme value theory may not be adequate because the independent assumption of daily log returns is often rejected by statistical testings. Finally, the use of subperiod minima overlooks the fact of volatility clustering in the daily log returns. The new approach of extreme value theory discussed in the next section overcomes these weaknesses.

Remark: As shown by the results of Example 1, the VaR calculation based on the traditional extreme value theory depends on the choice of n, which is the length of subperiods. For the limiting extreme value distribution to hold, one would prefer a large n. But a larger n means a smaller g when the sample size T is fixed, where g is the effective sample size used in estimating the three parameters α_n, β_n, and k_n. Therefore, some compromise between the choices of n and g is needed. A proper choice may depend on the returns of the asset under study. We recommend that one should check the stability of the resulting VaR in applying the traditional extreme value theory.

5.1 Discussion

We have applied various methods of VaR calculation to the daily log returns of IBM stock for a long position of \$10 million. Consider the VaR of the position for the next trading day. If the probability is 5%, which means that with probability 0.95 the loss will be less than or equal to the VaR for the next trading day, then the results obtained are

1. \$302,500 for the RiskMetrics,
2. \$287,200 for a Gaussian AR(2)-GARCH(1,1) model,
3. \$283,520 for an AR(2)-GARCH(1,1) model with a standardized Student-t distribution with 5 degrees of freedom,
4. \$216,030 for using the empirical quantile, and
5. \$184,127 for applying the traditional extreme value theory using monthly minima (i.e., subperiod length $n = 21$).

If the probability is 1%, then the VaR is

1. \$426,500 for the RiskMetrics,
2. \$409,738 for a Gaussian AR(2)-GARCH(1,1) model,
3. \$475,943 for an AR(2)-GARCH(1,1) model with a standardized Student-t distribution with 5 degrees of freedom,
4. \$365,709 for using the empirical quantile, and
5. \$340,013 for applying the traditional extreme value theory using monthly minima (i.e., subperiod length $n = 21$).

If the probability is 0.1%, then the VaR becomes

1. \$566,443 for the RiskMetrics,
2. \$546,641 for a Gaussian AR(2)-GARCH(1,1) model,
3. \$836,341 for an AR(2)-GARCH(1,1) model with a standardized Student-t distribution with 5 degrees of freedom,
4. \$780,712 for using the empirical quantile, and
5. \$666,590 for applying the traditional extreme value theory using monthly minima (i.e., subperiod length $n = 21$).

There are substantial differences among different approaches. This is not surprising because there exists substantial uncertainty in estimating tail behavior of a statistical distribution. Since there is no true VaR available to compare the accuracy of different approaches, we recommend that one applies several methods to gain insight into the range of VaR.

The choice of tail probability also plays an important role in VaR calculation. For the daily IBM stock returns, the sample size is 9190 so that the empirical quantiles of 5% and 1% are decent estimates of the quantiles of the return distribution. In this case, we can treat the results based on empirical quantiles as conservative estimates of the true VaR (i.e., lower bounds). In this view, the approach based on the traditional extreme value theory seems to underestimate the VaR for the daily log returns of IBM stock. The conditional approach of extreme value theory discussed in the next section overcomes this weakness.

When the tail probability is small (e.g., 0.1%), the empirical quantile is a less reliable estimate of the true quantile. The VaR based on empirical quantiles can no longer serve as a lower bound of the true VaR. Finally, the earlier results show clearly the effects of using a heavy-tail distribution in VaR calculation when the tail probability is small. The VaR based on either a Student-t distribution with 5 degrees of freedom or the extreme value distribution is greater than that based on the normal assumption when the probability is 0.1%.

5.2 Multiperiod VaR

The square root of time rule of the RiskMetrics methodology becomes a special case under the extreme value theory. The proper relationship between ℓ-day and 1-day horizons is

$$\text{VaR}(\ell) = \ell^{1/\alpha}\text{VaR} = \ell^{-k}\text{VaR},$$

where α is the tail index and k is the shape parameter of the extreme value distribution; see Danielsson and de Vries (1997a). This relationship is referred to as the α-root of time rule.

For illustration, consider the daily log returns of IBM stock in Example 1. If we use $p = 0.05$ and the results of $n = 21$, then for a 30-day horizon we have

$$\text{VaR}(30) = (30)^{0.335}\text{VaR} = 3.125 \times \$184,127 = \$575,397.$$

Because $\ell^{0.335} < \ell^{0.5}$, the α-root of time rule produces lower ℓ-day horizon VaR than does the square root of time rule.

5.3 VaR for a short position

In this subsection, we give the formulas of VaR calculation for holding short positions. Here the quantity of interest is the subperiod maximum and the limiting extreme value distribution becomes

$$F_*(r) = \begin{cases} \exp\left\{-\left[1 - \frac{k_n(r-\beta_n)}{\alpha_n}\right]^{1/k_n}\right\} & \text{if } k_n \neq 0 \\ \\ \exp\left[-\exp\left(\frac{r-\beta_n}{\alpha_n}\right)\right] & \text{if } k_n = 0, \end{cases} \tag{5.11}$$

where r denotes a value of the subperiod maximum and it is understood that $1 - k_n(r - \beta_n)/\alpha_n > 0$ for $k_n \neq 0$.

Following similar procedures as those of long positions, we obtain the $(1-p)$th quantile of the return r_t as

$$\text{VaR} = \begin{cases} \beta_n + \frac{\alpha_n}{k_n}\left\{1 - [-n\ln(1-p)]^{k_n}\right\} & \text{if } k_n \neq 0 \\ \\ \beta_n + \alpha_n \ln[-n\ln(1-p)] & \text{if } k_n = 0, \end{cases} \tag{5.12}$$

where p is a small probability denoting the chance of loss for holding a short position.

6 A new approach based on the extreme value theory

The aforementioned approach to VaR calculation using the extreme value theory encounters some difficulties. First, the choice of subperiod length n is not clearly defined. Second, the approach is unconditional and, hence, does not take into consideration effects of other explanatory variables. To overcome these difficulties, a modern approach to extreme value theory has been proposed in the statistical literature; see Davison and Smith (1990) and Smith (1989). Instead of focusing on the extremes (maximum or minimum), the new approach focuses on exceedances of the measurement over some high threshold and the times at which the exceedances occur. For instance, consider the daily log returns r_t of IBM stock used in this chapter and a long position on the stock. Let η be a prespecified high threshold. We may choose $\eta = -2.5\%$. Suppose that the ith exceedance occurs at day t_i (i.e., $r_{t_i} \leq \eta$). Then the new approach focuses on the data $(t_i, r_{t_i} - \eta)$. Here $r_{t_i} - \eta$ is the exceedance over the threshold η and t_i is the time at which the ith exceedance occurs. Similarly, for a short position, we may choose $\eta = 2\%$ and focus on the data $(t_i, r_{t_i} - \eta)$ for which $r_{t_i} \geq \eta$.

In practice, the occurrence times $\{t_i\}$ provide useful information about the intensity of the occurrence of important "rare events" (e.g., less than the threshold η for a long position). A cluster of t_i indicates a period of large market declines. The exceeding amount (or exceedance) $r_{t_i} - \eta$ is also of importance as it provides the actual quantity of interest.

Based on the prior introduction, the new approach does not require the choice of a subperiod length n, but it requires the specification of threshold η. Different choices of the threshold η lead to different estimates of the shape parameter k (and hence the tail index $\alpha = -1/k$). In the literature, some researchers believe that the choice of η is a statistical problem as well as a financial one, and it cannot be determined purely based on the statistical theory. For example, different financial institutions (or investors) have different risk tolerances. As such, they may select different thresholds even for an identical financial position. For the daily log returns of IBM stock considered in this chapter, the calculated VaR is not sensitive to the choice of η.

The choice of threshold η also depends on the observed log returns. For a stable return series, $\eta = -2.5\%$ may fare well for a long position. For a volatile return series (e.g., daily returns of a dot-com stock), η may be as low as -10%. Limited experience shows that η can be chosen so that the number of exceedances is sufficiently large (e.g., about 5% of the sample). For a more formal study on the choice of η, see Danielsson and de Vries (1997b).

6.1 Statistical theory

Again consider the log return r_t of an asset. Suppose that the ith exceedance occurs at t_i. Focusing on the exceedance $r_t-\eta$ and exceeding time t_i results in a fundamental change in statistical thinking. Instead of using the marginal distribution (e.g., the limiting distribution of the minimum or maximum), the new approach employs a conditional distribution to handle the magnitude of exceedance given that the measurement exceeds a threshold. The chance of exceeding the threshold is governed by a probability law. In other words, the new approach considers the conditional distribution of $x = r_t - \eta$ given $r_t \le \eta$ for a long position. Occurrence of The event $\{r_t \le \eta\}$ follows a point process (e.g., a Poisson process). In particular, if the intensity parameter λ of the process is time-invariant, then the Poisson process is homogeneous. If λ is time-variant, then the process is nonhomogeneous. The concept of Poisson process can be generalized to the multivariate case.

For ease in presentation, in what follows we use a positive threshold and the right-hand side of a return distribution to discuss the statistical theory behind the new approach of extreme value theory. This corresponds to holding a short financial position. However, the theory applies equally well to holding a long position if it is applied to the r_t^c series, where $r_t^c = -r_t$. This is easily seen because $r_t^c \ge \eta$ for a positive threshold is equivalent to $r_t \le -\eta$, where $-\eta$ becomes a negative threshold.

The basic theory of the new approach is to consider the conditional distribution of $r = x + \eta$ given $r > \eta$ for the limiting distribution of the maximum given in Eq. (5.11). Since there is no need to choose the subperiod length n, we do not use it as a subscript of the parameters. Then the conditional distribution of $r \le x+\eta$ given $r > \eta$ is

$$Pr(r \le x+\eta | r > \eta) = \frac{Pr(\eta \le r \le x+\eta)}{Pr(r > \eta)} = \frac{Pr(r \le x+\eta) - Pr(r \le \eta)}{1 - Pr(r \le \eta)}. \tag{6.13}$$

Using the CDF $F_*(.)$ of Eq. (5.11) and the approximation $e^{-y} \approx 1 - y$ and after some algebra, we obtain that

$$\begin{aligned} Pr(r \le x+\eta | r > \eta) &= \frac{F_*(x+\eta) - F_*(\eta)}{1 - F_*(\eta)} \\ &= \frac{\exp\left\{-\left[1 - \frac{k(x+\eta-\beta)}{\alpha}\right]^{1/k}\right\} - \exp\left\{-\left[1 - \frac{k(\eta-\beta)}{\alpha}\right]^{1/k}\right\}}{1 - \exp\left\{-\left[1 - \frac{k(\eta-\beta)}{\alpha}\right]^{1/k}\right\}} \\ &\approx 1 - \left[1 - \frac{kx}{\alpha - k(\eta-\beta)}\right]^{1/k}, \end{aligned} \tag{6.14}$$

where $x > 0$ and $1 - k(\eta - \beta)/\alpha > 0$. As is seen later, this approximation makes explicit the connection of the new approach to the traditional extreme value theory. The case of $k = 0$ is taken as the limit of $k \to 0$ so that

$$Pr(r \le x+\eta | r > \eta) \approx 1 - \exp(-x/\alpha).$$

6.2 A new approach

Using the statistical result in Eq. (6.14) and considering jointly the exceedances and exceeding times, Smith (1989) proposes a two-dimensional Poisson process to model (t_i, r_{t_i}). This approach was used by Tsay (1999) to study VaR in risk management. We follow the same approach.

Assume that the baseline time interval is T, which is typically a year. In the United States, $T = 252$ is used as there are typically 252 trading days in a year. Let t be the time interval of the data points (e.g., daily) and denote the data span by $t = 1, 2, \cdots, N$, where N is the total number of data points. For a given threshold η, the exceeding times over the threshold are denoted by $\{t_i,\ i = 1, \cdots, N_\eta\}$ and the observed log return at t_i is r_{t_i}. Consequently, we focus on modeling $\{(t_i, r_{t_i})\}$ for $i = 1, \cdots, N_\eta$, where N_η depends on the threshold η.

The new approach to applying the extreme value theory is to postulate that the exceeding times and the associated returns [i.e., (t_i, r_{t_i})], jointly form a two-dimensional Poisson process with intensity measure given by

$$\Lambda[(T_2, T_1) \times (r, \infty)] = \frac{T_2 - T_1}{T} S(r; k, \alpha, \beta), \tag{6.15}$$

where

$$S(r; k, \alpha, \beta) = \left[1 - \frac{k(r - \beta)}{\alpha}\right]_+^{1/k},$$

$0 \le T_1 \le T_2 \le N$, $r > \eta$, $\alpha > 0$, β, and k are parameters, and the notation $[x]_+$ is defined as $[x]_+ = \max(x, 0)$. This intensity measure says that the occurrence of exceeding the threshold is proportional to the length of the time interval $[T_1, T_2]$ and the probability is governed by a survival function similar to the exponent of the CDF $F_*(r)$ in Eq. (5.11). A survival function of a random variable X is defined as $S(x) = Pr(X > x) = 1 - Pr(X \le x) = 1 - \text{CDF}(x)$. When $k = 0$, the intensity measure is taken as the limit of $k \to 0$ – that is,

$$\Lambda[(T_2, T_1) \times (r, \infty)] = \frac{T_2 - T_1}{T} \exp\left[\frac{-(r - \beta)}{\alpha}\right].$$

In Eq. (6.15), the length of time interval is measured with respect to the baseline interval T.

The idea of using the intensity measure in Eq. (6.15) becomes clear when one considers its implied conditional probability of $r = x + \eta$ given $r > \eta$ over the time interval $[0, T]$, where $x > 0$,

$$\frac{\Lambda[(0, T) \times (x + \eta, \infty)]}{\Lambda[(0, T) \times (\eta, \infty)]} = \left[\frac{1 - k(x + \eta - \beta)/\alpha}{1 - k(\eta - \beta)/\alpha}\right]^{1/k} = \left[1 - \frac{kx}{\alpha - k(\eta - \beta)}\right]^{1/k},$$

which is precisely the survival function of the conditional distribution given in Eq. (6.14). This survival function is obtained from the extreme limiting distribution for maximum in Eq. (5.11). We use survival function here because it denotes the probability of exceedance.

The relationship between the limiting extreme value distribution in Eq. (5.11) and the intensity measure in Eq. (6.15) directly connects the new approach of extreme value theory to the traditional one.

Mathematically, the intensity measure in Eq. (6.15) can be written as an integral of an intensity function:

$$\Lambda[(T_2, T_1) \times (r, \infty)] = \int_{T_1}^{T_2} \int_r^\infty \lambda(t, z; k, \alpha, \beta) dt dz,$$

where the intensity function $\lambda(t, z; k, \alpha, \beta)$ is defined as

$$\lambda(t, z; k, \alpha, \beta) = \frac{1}{T} g(z; k, \alpha, \beta) \tag{6.16}$$

Table 2: Estimation Results of a Two-Dimensional Homogeneous Poisson Model For The Daily Negative Log Returns of IBM Stock From July 3, 1962 to December 31, 1998. The Baseline Time Interval is 252 (i.e., 1 Year). The Numbers in Parentheses Are Standard Errors, Where "Thr." and "Exc." Stand For Threshold And The Number of Exceedings.

Thr.	Exc.	Shape Par. k	Log(Scale) $\ln(\alpha)$	Location β
(a) Original log returns				
3.0%	175	−0.30697(0.09015)	0.30699(0.12380)	4.69204(0.19058)
2.5%	310	−0.26418(0.06501)	0.31529(0.11277)	4.74062(0.18041)
2.0%	554	−0.18751(0.04394)	0.27655(0.09867)	4.81003(0.17209)
(b) Removing the sample mean				
3.0%	184	−0.30516(0.08824)	0.30807(0.12395)	4.73804(0.19151)
2.5%	334	−0.28179(0.06737)	0.31968(0.12065)	4.76808(0.18533)
2.0%	590	−0.19260(0.04357)	0.27917(0.09913)	4.84859(0.17255)

where

$$g(z; k, \alpha, \beta) = \begin{cases} \frac{1}{\alpha}\left[1 - \frac{k(z-\beta)}{\alpha}\right]^{1/k-1} & \text{if } k \neq 0 \\ \frac{1}{\alpha}\exp\left[\frac{-(z-\beta)}{\alpha}\right] & \text{if } k = 0. \end{cases}$$

Using the results of a Poisson process, we can write down the likelihood function for the observed exceeding times and their corresponding returns $\{(t_i, r_{t_i})\}$ over the two-dimensional space $[0, N] \times (\eta, \infty)$ as

$$L(k, \alpha, \beta) = \left(\prod_{i=1}^{N_\eta} \frac{1}{T} g(r_{t_i}; k, \alpha, \beta)\right) \times \exp\left[-\frac{N}{T} S(\eta; k, \alpha, \beta)\right]. \tag{6.17}$$

The parameters k, α, β can then be estimated by maximizing the logarithm of this likelihood function. Since the scale parameter α is nonnegative, we use $\ln(\alpha)$ in the estimation.

Example 2. (Example 7.7 of Tsay, 2001). Consider again the daily log returns of IBM stock from July 3, 1962 to December 31, 1998. There are 9190 daily returns. Table 2 gives some estimation results of the parameters k, α, β for three choices of the threshold when the negative series $\{-r_t\}$ is used. We use the negative series $\{-r_t\}$, instead of $\{r_t\}$, because we focus on holding a long financial position. The table also shows the number of exceeding times for a given threshold. It is seen that the chance of dropping 2.5% or more in a day for IBM stock occurred with probability $310/9190 \approx 3.4\%$. Because the sample mean of IBM stock returns is not zero, we also consider the case when the sample mean is removed from the original daily log returns. From the table, removing the sample mean has little impact on the parameter estimates. These parameter estimates are used next to calculate VaR, keeping in mind that in a real application one needs to check carefully the adequacy of a fitted Poisson model. We discuss methods of model checking in the next subsection.

6.3 VaR calculation based on the new approach

As shown in Eq. (6.14), the two-dimensional Poisson process model used, which employs the intensity measure in Eq. (6.15), has the same parameters as those of the extreme value distribution in Eq.

(5.11). Therefore, one can use the same formula as that of the Eq. (5.12) to calculate VaR of the new approach. More specifically, for a given upper tail probability p, the $(1-p)$th quantile of the log return r_t is

$$\mathrm{VaR} = \begin{cases} \beta + \frac{\alpha}{k}\left\{1 - [-T\ln(1-p)]^k\right\} & \text{if } k \neq 0 \\ \\ \beta + \alpha \ln[-T\ln(1-p)] & \text{if } k = 0, \end{cases} \tag{6.18}$$

where T is the baseline time interval used in estimation. Typically, $T = 252$ in the United States for the approximate number of trading days in a year.

Example 3. (Example 7.8 of Tsay, 2001). Consider again the case of holding a long position of IBM stock valued at \$10 million. We use the estimation results of Table 2 to calculate 1-day horizon VaR for the tail probabilities of 0.05 and 0.01.

- Case I: Use the original daily log returns. The three choices of threshold η result in the following VaR values:

  1. $\eta = 3.0\%$: VaR(5%) = \$228,239, VaR(1%) = \$359.303.
  2. $\eta = 2.5\%$: VaR(5%) = \$219,106, VaR(1%) = \$361,119.
  3. $\eta = 2.0\%$: VaR(5%) = \$212,981, VaR(1%) = \$368.552.

- Case II: The sample mean of the daily log returns is removed. The three choices of threshold η result in the VaR values:

  1. $\eta = 3.0\%$: VaR(5%) = \$232,094, VaR(1%) = \$363,697.
  2. $\eta = 2.5\%$: VaR(5%) = \$225,782, VaR(1%) = \$364,254.
  3. $\eta = 2.0\%$: VaR(5%) = \$217,740, VaR(1%) = \$372,372.

As expected, removing the sample mean, which is positive, increases slightly the VaR. However, the VaR is rather stable among the three threshold values used. In practice, we recommend that one removes the sample mean first before applying this new approach to VaR calculation.

Discussion: Compared with the VaR of Example 1 that uses the traditional extreme value theory, the new approach provides a more stable VaR calculation. The traditional approach is rather sensitive to the choice of the subperiod length n.

6.4 Use of explanatory variables

The two-dimensional Poisson process model discussed earlier is *homogeneous* because the three parameters k, α, and β are constant over time. In practice, such a model may not be adequate. Furthermore, some explanatory variables are often available that may influence the behavior of the log returns r_t. A nice feature of the new extreme value theory approach to VaR calculation is that it can easily take explanatory variables into consideration. We discuss such a framework in this subsection. In addition, we also discuss methods that can be used to check the adequacy of a fitted two-dimensional Poisson process model.

Suppose that $\boldsymbol{x}_t = (x_{1t}, \cdots, x_{vt})'$ is a vector of v explanatory variables that are available *prior to* time t. For asset returns, the volatility σ_t^2 of r_t obtained by econometric models is an example of explanatory variables. Another example of explanatory variables in the U.S. equity markets is an indicator variable denoting the meetings of Federal Open Market Committee. A simple way to make use of explanatory variables is to postulate that the three parameters k, α, and β are time-varying and are linear functions of the explanatory variables. Specifically, when explanatory variables $\boldsymbol{x}_t$ are available, we assume that

$$\begin{aligned}
k_t &= \gamma_0 + \gamma_1 x_{1t} + \cdots + \gamma_v x_{vt} \equiv \gamma_0 + \boldsymbol{\gamma}'\boldsymbol{x}_t \\
\ln(\alpha_t) &= \delta_0 + \delta_1 x_{1t} + \cdots + \delta_v x_{vt} \equiv \delta_0 + \boldsymbol{\delta}'\boldsymbol{x}_t \\
\beta_t &= \theta_0 + \theta_1 x_{1t} + \cdots + \theta_v x_{vt} \equiv \theta_0 + \boldsymbol{\theta}'\boldsymbol{x}_t.
\end{aligned} \tag{6.19}$$

If $\boldsymbol{\gamma} = \mathbf{0}$, then the shape parameter $k_t = \gamma_0$, which is time-invariant. Thus, testing the significance of $\boldsymbol{\gamma}$ can provide information about the contribution of the explanatory variables to the shape parameter. Similar methods apply to the scale and location parameters. In Eq. (6.19), we use the same explanatory variables for all the three parameters $k_t, \ln(\alpha_t)$, and β_t. In an application, different explanatory variables may be used for different parameters.

When the three parameters of the extreme value distribution are time-varying, we have an *inhomogeneous* Poisson process. The intensity measure becomes

$$\Lambda[(T_1, T_2) \times (r, \infty)] = \frac{T_2 - T_1}{T}\left(1 - \frac{k_t(r - \beta_t)}{\alpha_t}\right)_+^{1/k_t}, \quad r > \eta. \tag{6.20}$$

The likelihood function of the exceeding times and returns $\{(t_i, r_{t_i})\}$ becomes

$$L = \left(\prod_{i=1}^{N_\eta} \frac{1}{T} g(r_{t_i}; k_t, \alpha_t, \beta_t)\right) \times \exp\left[-\frac{1}{T}\int_0^N S(\eta; k_t, \alpha_t, \beta_t)dt\right],$$

which reduces to

$$L = \left(\prod_{i=1}^{N_\eta} \frac{1}{T} g(r_{t_i}; k_t, \alpha_t, \beta_t)\right) \times \exp\left[-\frac{1}{T}\sum_{t=1}^{N} S(\eta; k_t, \alpha_t, \beta_t)\right] \tag{6.21}$$

if one assumes that the parameters k_t, α_t, and β_t are constant within each trading day, where $g(z; k_t, \alpha_t, \beta_t)$ and $S(\eta; k_t, \alpha_t, \beta_t)$ are given in Eqs. (6.16) and (6.15), respectively. For given observations $\{r_t, \boldsymbol{x}_t | t = 1, \cdots, N\}$, the baseline time interval T, and the threshold η, the parameters in Eq. (6.19) can be estimated by maximizing the logarithm of the likelihood function in Eq. (6.21). Again we use $\ln(\alpha_t)$ to satisfy the positive constraint of α_t.

6.5 Model checking

Checking an entertained two-dimensional Poisson process model for exceedance times and excesses involves examining three key features of the model. The first feature is to verify the adequacy of the exceedance rate, the second feature is to examine the distribution of exceedances, and the final feature is to check the independence assumption of the model. We discuss briefly some statistics that are useful for checking these three features. These statistics are based on some basic statistical theory concerning distributions and stochastic processes.

<u>Exceedance rate</u> A fundamental property of univariate Poisson processes is that the time durations between two consecutive events are independent and exponentially distributed. To exploit a similar

property for checking a two-dimensional process model, Smith and Shively (1995) propose to examine the time durations between consecutive exceedances. If the two-dimensional Poisson process model is appropriate for the exceedance times and excesses, the time duration between the ith and $(i-1)$th exceedances should follow an exponential distribution. More specifically, letting $t_0 = 0$, we expect that

$$z_{t_i} = \int_{t_{i-1}}^{t_i} \frac{1}{T} g(\eta; k_s, \alpha_s, \beta_s) ds, \quad i = 1, 2, \cdots$$

are independent and identically distributed (iid) as a standard exponential distribution. Because daily returns are discrete-time observations, we employ the time durations

$$z_{t_i} = \frac{1}{T} \sum_{t=t_{i-1}+1}^{t_i} S(\eta; k_t, \alpha_t, \beta_t) \tag{6.22}$$

and use the quantile-to-quantile (QQ) plot to check the validity of the iid standard exponential distribution. If the model is adequate, the QQ-plot should show a straight line through the origin with unit slope.

Distribution of excesses Under the two-dimensional Poisson process model considered, the conditional distribution of the excess $x_t = r_t - \eta$ over the threshold η is a generalized Pareto distribution (GPD) with shape parameter k_t and scale parameter $\psi_t = \alpha_t - k_t(\eta - \beta_t)$. Therefore, we can make use of the relationship between a standard exponential distribution and GPD, and define

$$w_{t_i} = \begin{cases} \frac{-1}{k_{t_i}} \ln \left(1 - k_{t_i} \frac{r_{t_i} - \eta}{\psi_{t_i}}\right)_+ & \text{if } k_{t_i} \neq 0 \\ \frac{r_{t_i} - \eta}{\psi_{t_i}} & \text{if } k_{t_i} = 0. \end{cases} \tag{6.23}$$

If the model is adequate, $\{w_{t_i}\}$ are independent and exponentially distributed with mean 1; see also Smith (1999). We can then apply the QQ-plot to check the validity of the GPD assumption for excesses.

Independence A simple way to check the independence assumption, after adjusting for the effects of explanatory variables, is to examine the sample autocorrelation functions of z_{t_i} and w_{t_i}. Under the independence assumption, we expect zero serial correlations for both z_{t_i} and w_{t_i}.

6.6 An illustration

In this subsection, we apply a two-dimensional inhomogeneous Poisson process model to the daily log returns, in percentages, of IBM stock from July 3, 1962 to December 31, 1998. We focus on holding a long position of \$10 million. The analysis enables us to compare the results with those obtained before by using other approaches to calculating VaR.

We begin by pointing out that the two-dimensional homogeneous model of Example 2 needs further refinements because the fitted model fails to pass the model checking statistics of the previous subsection. Figures 4(a) and (b) show the autocorrelation functions of the statistics z_{t_i} and w_{t_i}, defined in Eqs. (6.22) and (6.23), of the homogeneous model when the threshold is $\eta = 2.5\%$. The horizontal lines in the plots denote asymptotic limits of two standard errors. It is seen that both z_{t_i} and w_{t_i} series have some significant serial correlations. Figures 5(a) and (b) show the QQ-plots of z_{t_i} and w_{t_i} series. The straight line in each plot is the theoretical line, which passes through the origin

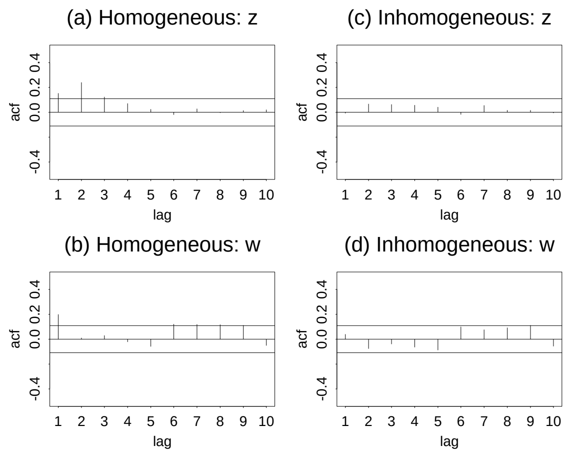

Figure 4: Sample autocorrelation functions of the z and w measures for two-dimensional Poisson models. (a) and (b) are for the homogeneous model, and (c) and (d) are for the inhomogeneous model. The data are daily mean-corrected log returns, in percentages, of IBM stock from July 3, 1962 to December 31, 1998 and the threshold is 2.5%. A long financial position is used.

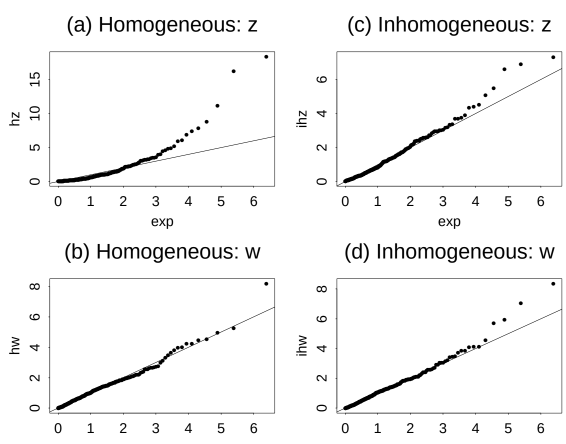

Figure 5: Quantile-to-quantile plot of the z and w measures for two-dimensional Poisson models. (a) and (b) are for the homogeneous model, and (c) and (d) are for the inhomogeneous model. The data are daily mean-corrected log returns, in percentages, of IBM stock from July 3, 1962 to December 31, 1998, and the threshold is 2.5%. A long financial position is used.

and has a unit slope under the assumption of a standard exponential distribution. The QQ-plot of z_{t_i} shows some discrepancy.

To refine the model, we use the mean-corrected log return series

$$r_t^o = r_t - \bar{r}, \quad \bar{r} = \frac{1}{9190}\sum_{t=1}^{9190} r_t,$$

where r_t is the daily log return in percentages, and employ the following explanatory variables:

1. x_{1t}: an indicator variable for October, November, and December. That is, $x_{1t} = 1$ if t is in October, November, or December. This variable is chosen to take care of the fourth-quarter effect (or year-end effect), if any, on the daily IBM stock returns.

2. x_{2t}: an indicator variable for the behavior of the previous trading day. Specifically, $x_{2t} = 1$ if and only if the log return $r_{t-1}^o \leq -2.5\%$. Since we focus on holding a long position with threshold 2.5%, an exceedance occurs when the daily price drops over 2.5%. Therefore, x_{2t} is used to capture the possibility of panic selling when the price of IBM stock dropped 2.5% or more on the previous trading day.

3. x_{3t}: a qualitative measurement of volatility, which is the number of days between $t-1$ and $t-5$ (inclusive) that has a log return with magnitude exceeding the threshold. In our case, x_{3t} is the number of r_{t-i}^o satisfying $|r_{t-i}^o| \geq 2.5\%$ for $i = 1, \cdots, 5$.

4. x_{4t}: an annual trend defined as $x_{4t} =$ (year of time $t - 1961)/38$. This variable is used to detect any trend in the behavior of extreme returns of IBM stock.

5. x_{5t}: a volatility series based on a Gaussian GARCH(1,1) model for the mean-corrected series r_t^o. Specifically, $x_{5t} = \sigma_t$, where σ_t^2 is the conditional variance of the GARCH(1,1) model

$$r_t^o = a_t, \quad a_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim N(0,1)$$

$$\sigma_t^2 = 0.04565 + 0.0807 a_{t-1}^2 + 0.9031 \sigma_{t-1}^2.$$

These five explanatory variables are all available at time $t-1$. We use two volatility measures (x_{3t} and x_{5t}) to study the effect of market volatility on VaR. Because the serial correlations in r_t are weak so that we do not entertain any ARMA model for the mean equation.

Using the prior five explanatory variables and deleting insignificant parameters, we obtain the estimation results shown in Table 3. Figures 4(c) and (d) and Figures 5(c) and (d) show the model checking statistics for the fitted two-dimensional inhomogeneous Poisson process model when the threshold is $\eta = 2.5\%$. All autocorrelation functions of z_{t_i} and w_{t_i} are within the asymptotic two standard-error limits. The QQ-plots also show marked improvements as they indicate no model inadequacy. Based on these checking results, the inhomogeneous model seems adequate.

Consider the case of threshold 2.5%. The estimation results show the following:

1. All three parameters of the intensity function depend significantly on the annual time trend. In particular, the shape parameter has a negative annual trend, indicating that the log returns of IBM stock are moving farther away from normality as time passes. Both the location and scale parameters increase over time.

2. Indicators for the fourth quarter, x_{1t}, and for panic selling, x_{2t}, are not significant for all three parameters.

3. The location and shape parameters are positively affected by the volatility of the GARCH(1,1) model; see the coefficients of x_{5t}. This is understandable because the variability of log returns increases when the volatility is high. Consequently, the dependence of log returns on the tail index is reduced.

4. The scale and shape parameters depend significantly on the qualitative measure of volatility. The signs of the estimates are also plausible.

The explanatory variables for December 31, 1998 assumed the values $x_{3,9190} = 0$, $x_{4,9190} = 0.9737$, and $x_{5,9190} = 1.9766$. Using these values and the fitted model in Table 3, we obtain

$$k_{9190} = -0.01195, \quad \ln(\alpha_{9190}) = 0.19331, \quad \beta_{9190} = 6.105.$$

Assume that the tail probability is 0.05. The VaR quantile shown in Eq. (6.18) gives VaR = 3.03756%. Consequently, for a long position of \$10 million, we have

$$\text{VaR} = \$10,000,000 \times 0.0303756 = \$303,756.$$

If the tail probability is 0.01, the VaR is \$497,425. The 5% VaR is slightly larger than that based on a Gaussian AR(2)-GARCH(1,1) model. The 1% VaR is larger than that of Case I of Example 7.3 of Tsay (2001). Again, as expected, the effect of extreme values (i.e., heavy tails) on VaR is more pronounced when the tail probability used is small.

An advantage of using explanatory variables is that the parameters are adaptive to the change in market conditions. For example, the explanatory variables for December 30, 1998 assumed the values $x_{3,9189} = 1$, $x_{4,9189} = 0.9737$, and $x_{5,9189} = 1.8757$. In this case, we have

$$k_{9189} = -0.2500, \quad \ln(\alpha_{9189}) = 0.52385, \quad \beta_{9189} = 5.8834.$$

The 95% quantile (i.e., the tail probability is 5%) then becomes 2.69139%. Consequently, the VaR is

$$\text{VaR} = \$10,000,000 \times 0.0269139 = \$269,139.$$

If the tail probability is 0.01, then VaR becomes \$448,323. Based on this example, the homogeneous Poisson model shown in Example 3 seems to underestimate the VaR.

7 Bayesian Approach

A Bayesian approach to VaR calculation that uses the conditional EVT will be discussed in the lecture.

Exercise

The file "d-csco9199.dat" contains the daily log returns of Cisco Systems stock from 1991 to 1999 with 2275 observations. Suppose that you hold a long position of Cisco stock valued at \$1 million. Compute the Value at Risk of your position for the next trading day using probability $p = 0.01$.

Table 3: Estimation Results of a Two-Dimensional Inhomogeneous Poisson Process Model For Daily Log Returns, in Percentages, of IBM Stock From July 3, 1962 to December 31, 1998. Four Explanatory Variables Defined in The Text Are Used. The Model is For Holding a Long Position on IBM Stock. The Sample Mean of The Log Returns is Removed From The Data.

(a) Threshold 2.5% with 334 exceedances				
Parameter	constant	Coef. of x_{3t}	Coef. of x_{4t}	Coef. of x_{5t}
β_t	0.3202		1.4772	2.1991
(Std.err)	(0.3387)		(0.3222)	(0.2450)
$\ln(\alpha_t)$	-0.8119	0.3305	1.0324	
(Std.err)	(0.1798)	(0.0826)	(0.2619)	
k_t	-0.1805	-0.2118	-0.3551	0.2602
(Std.err)	(0.1290)	(0.0580)	(0.1503)	(0.0461)
(b) Threshold 3.0% with 184 exceedances				
β_t	1.1569			2.1918
(Std.err)	(0.4082)			(0.2909)
$\ln(\alpha_t)$	-0.0316	0.3336		
(Std.err)	(0.1201)	(0.0861)		
k_t	-0.6008	-0.2480		0.3175
(Std.err)	(0.1454)	(0.0731)		(0.0685)

1. Use the RiskMetrics method.

2. Use a GARCH model with a conditional Gaussian distribution.

3. Use a GARCH model with a Student-t distribution. You may also estimate the degrees of freedom.

4. Use the unconditional sample quantile.

5. Use a two-dimensional homogeneous Poisson process with threshold 2%. That is, focusing on the exceeding times and exceedances that the daily stock price drops 2% or more. Check the fitted model.

6. Use a two-dimensional nonhomogeneous Poisson process with threshold 2%. The explanatory variables are (1) an annual time trend, (2) a dummy variable for October, November, and December, and (3) a fitted volatility based on a Gaussian GARCH(1,1) model. Perform a diagnostic check on the fitted model.

7. Repeat the prior two-dimensional nonhomogeneous Poisson process with threshold 2.5% or 3%. Comment on the selection of threshold.

References

Berman, S. M. (1964), "Limiting theorems for the maximum term in stationary sequences," *Annals of Mathematical Statistics*, 35, 502-516.

Danielsson, J., and De Vries, C. G. (1997a), "Value at risk and extreme returns," working paper, London School of Economics, London, U.K.

Danielsson, J., and De Vries, C. G. (1997b), "Tail index and quantile estimation with very high frequency data," *Journal of Empirical Finance*, 4, 241-257.

Davison, A. C., and Smith, R. L. (1990), "Models for exceedances over high thresholds" (with discussion), *Journal of the Royal Statistical Society, Series B*, 52, 393-442.

Embrechts, P., Kuppelberg, C., and Mikosch, T. (1997), *Modelling Extremal Events*, Berlin: Springer Verlag.

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications*, Vol. 2, New York: Wiley.

Gnedenko, B. V. (1943), "Sur la distribution limite du terme maximum of d'une série Aléatorie," *Annals of Mathematics*, 44, 423-453.

Gumbel, E. J. (1958), *Statistics of Extremes*, New York: Columbia University Press.

Jenkinson, A. F. (1955), "The frequency distribution of the annual maximum (or minimum) of meteorological elements," *Quarterly Journal of the Royal Meteorological Society*, 81, 158-171.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983), *Extremes and Related Properties of Random Sequences and Processes*, New York: Springer Verlag.

Longin, F. M. (1996), "The asymptotic distribution of extreme stock market returns," *Journal of Business*, 69, 383-408.

Longin, F. M. (1999a), "Optimal margin level in futures markets: Extreme price movements," *The Journal of Futures Markets*, 19, 127-152.

Longin, F. M. (1999b), "From value at risk to stress testing: the extreme value approach," working paper, Centre for Economic Policy Research, London, UK.

Smith, R. L. (1989), "Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone" (with discussion), *Statistical Science*, 4, 367-393.

Smith, R. L. (1999), "Measuring risk with extreme value theory," working paper, Department of Statistics, University of North Carolina at Chapel Hill.

Smith, R. L., and Shively, T. S. (1995), "A point process approach to modeling trends in troposheric ozone," *Atmospheric Environment*, 29, 3489-3499.

Tsay, R. S. (1999), "Extreme value analysis of financial data," working paper, Graduate School of Business, University of Chicago.

Tsay, R. S. (2001), *Analysis of Financial Time Series*, Wiley: New York.

8 Analysis of High-Freuqncy Financial Data

This lecture starts with basic characteristics of transactions data in the equity market. The material discussed is from Tsay (2001, Chapter 5) and McCulloch and Tsay (2001).

9 Empirical characteristics of transactions data

Let t_i be the calendar time, measured in seconds from midnight, at which the i-th transaction of an asset takes place. Associated with the transaction are several variables such as the transaction price, the transaction volume, the prevailing bid and ask quotes, and so on. The collection of t_i and the associated measurements are referred to as the *transactions data*. These data have several important characteristics that do not exist when the observations are aggregated over time. Some of the characteristics are given next.

1. Unequally spaced time intervals: Transactions such as stock tradings on an exchange do not occur at equally spaced time intervals. As such the observed transaction prices of an asset do not form an equally spaced time series. The time duration between trades becomes important and might contain useful information about market microstructure (e.g., trading intensity).

2. Discrete-valued prices: The price change of an asset from one transaction to the next only occurs in multiples of tick size. In the NYSE, the tick size was one eighth of a dollar before June 24, 1997, and was one sixteenth of a dollar before January 29, 2001. All NYSE and AMEX stocks started to trade in decimals on January 29, 2001. Therefore, the price is a discrete-valued variable in transactions data. In some markets, price change may also be subject to limit constraints set by regulators.

3. Existence of a daily periodic or diurnal pattern: Under the normal trading conditions, transaction activity can exhibit periodic pattern. For instance, in the NYSE, transactions are heavier at the beginning and closing of the trading hours and thinner during the lunch hours, resulting in a "U-shape" transaction intensity. Consequently, time durations between transactions also exhibit a daily cyclical pattern.

4. Multiple transactions within a single second: It is possible that multiple transactions, even with different prices, occur at the same time. This is partly due to the fact that time is measured in seconds that may be too long a time scale in periods of heavy tradings.

To demonstrate these characteristics, we consider first the IBM transactions data from November 1, 1990 to January 31, 1991. These data are from the Trades, Orders Reports, and Quotes (TORQ) dataset; see Hasbrouck (1992). There are 63 trading days and 60,328 transactions. To simplify the discussion, we ignore the price changes between trading days and focus on the transactions that occurred in the normal trading hours from 9:30 am to 4:00 pm Eastern Time. It is well known that overnight stock returns differ substantially from intraday returns; see Stoll and Whaley (1990) and the references therein. Table 4 gives the frequencies in percentages of price change measured in the tick size of $\$1/8 = \$0.125$. From the table, we make the following observations:

1. About two thirds of the intraday transactions were without price change.

2. The price changed in one tick approximately 29% of the intraday transactions.

Table 4: Frequencies of Price Change in Multiples of Tick Size For IBM Stock from November 1, 1990 to January 31, 1991.

Number(tick)	≤ -3	-2	-1	0	1	2	≥ 3
Percentage	0.66	1.33	14.53	67.06	14.53	1.27	0.63

3. Only 2.6% of the transactions were associated with two-tick price changes.
4. Only about 1.3% of the transactions resulted in price changes of three ticks or more.
5. The distribution of positive and negative price changes was approximately symmetric.

Consider next the number of transactions in a 5-minute time interval. Denote the series by x_t. That is, x_1 is the number of IBM transactions from 9:30 am to 9:35 am on November 1, 1990 Eastern time, x_2 is the number of transactions from 9:35 am to 9:40 am, and so on. The time gaps between trading days are ignored. Figure 6(a) shows the time plot of x_t, and Figure 6(b) the sample ACF of x_t for lags 1 to 260. Of particular interest is the cyclical pattern of the ACF with a periodicity of 78, which is the number of 5-minute intervals in a trading day. The number of transactions thus exhibits a daily pattern. To further illustrate the daily trading pattern, Figure 7 shows the average number of transactions within 5-minute time intervals over the 63 days. There are 78 such averages. The plot exhibits a "smiling" or "U" shape, indicating heavier tradings at the opening and closing of the market and thinner tradings during the lunch hours.

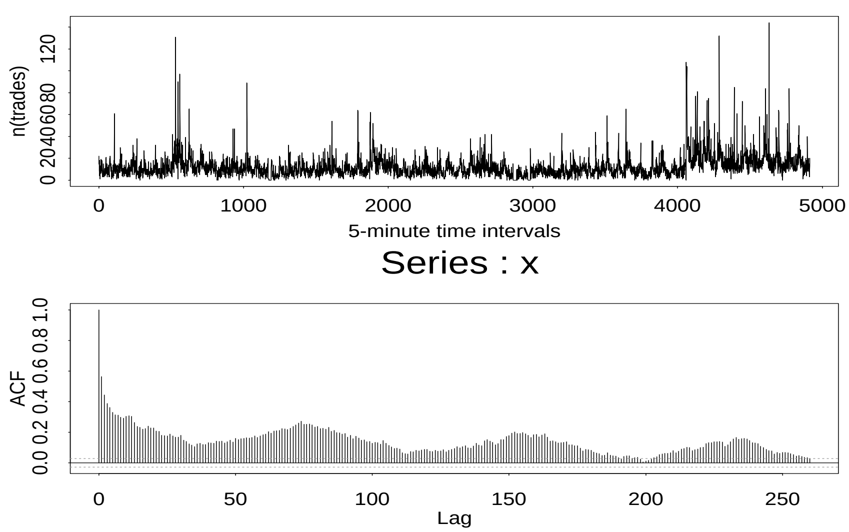

Figure 6: IBM intraday transactions data from 11/01/90 to 1/31/91: (a) The number of transactions in 5-minute time intervals, and (b) the sample ACF of the series in part(a).

Again, let t_i be the calendar time, measured in seconds from the midnight, when the i-th transaction took place. Let P_{t_i} be the transaction price. The price change from the $(i-1)$th to the ith trade

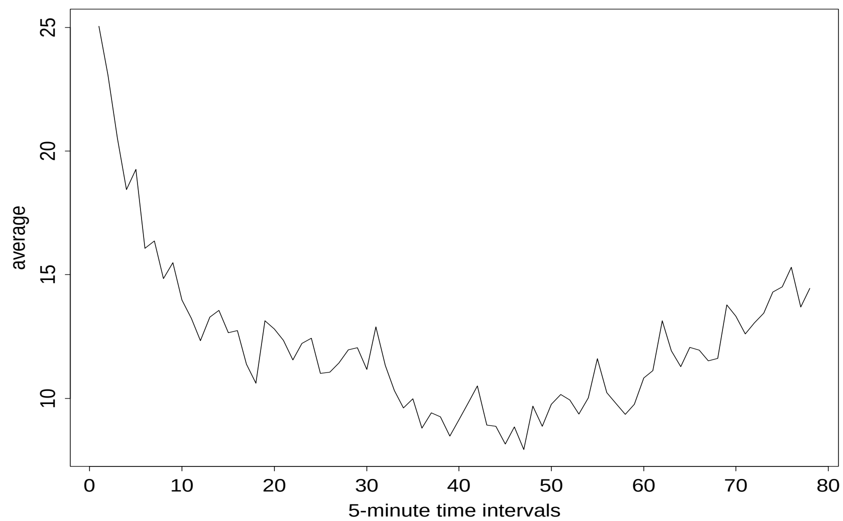

Figure 7: Time plot of the average number of transactions in 5-minute time intervals. There are 78 observations, averaging over the 63 trading days from 11/01/90 to 1/31/91 for IBM stock.

is $y_i \equiv \Delta P_{t_i} = P_{t_i} - P_{t_{i-1}}$ and the time duration is $\Delta t_i = t_i - t_{i-1}$. Here it is understood that the subscript i in Δt_i and y_i denotes the time sequence of transactions, not the calendar time. In what follows, we consider models for y_i and Δt_i both individually and jointly.

10 A Model For Price Changes

The discreteness and concentration on "no change" make it difficult to model the intraday price changes. Campbell, Lo, and MacKinlay (1997) discuss several econometric models that have been proposed in the literature. For instance, Hauseman, Lo, and MacKinlay (1992) apply a ordered probit model to study the price movements in transactions data. Here we discuss a decomposition model of McCulloch and Tsay (2001). The model is a simplified version of the model proposed by Rydberg and Shephard (1998); see also Ghysels (2000). Specifically, the price change at the ith transaction can be written as

$$y_i \equiv P_{t_i} - P_{t_{i-1}} = A_i D_i S_i, \tag{10.24}$$

where A_i is a binary variable defined as

$$A_i = \begin{cases} 1 & \text{if there is a price change at the } i\text{th trade} \\ 0 & \text{if price remains the same at the } i\text{th trade,} \end{cases} \tag{10.25}$$

D_i is also a discrete variable signifying the *direction* of the price change if a change occurs – that is,

$$D_i|(A_i = 1) = \begin{cases} 1 & \text{if price increases at the } i\text{th trade} \\ -1 & \text{if price drops at the } i\text{th trade,} \end{cases} \tag{10.26}$$

where $D_i|(A_i = 1)$ means that D_i is defined under the condition of $A_i = 1$, and S_i is size of the price change in ticks if there is a change at the ith trade and $S_i = 0$ if there is no price change at the ith trade. When there is a price change, S_i is a positive integer-valued random variable.

Note that D_i is not needed when $A_i = 0$, and there is a natural ordering in the decomposition. D_i is well defined only when $A_i = 1$ and S_i is meaningful when $A_i = 1$ and D_i is given. Model specification under the decomposition makes use of the ordering.

Let F_i be the information set available at the ith transaction. Examples of elements in F_i are Δt_{i-j}, A_{i-j}, D_{i-j}, and S_{i-j} for $j \geq 0$. The evolution of price change under model (10.24) can then be partitioned as

$$P(y_i|F_{i-1}) = P(A_i D_i S_i|F_{i-1}) = P(S_i|D_i, A_i, F_{i-1})P(D_i|A_i, F_{i-1})P(A_i|F_{i-1}). \tag{10.27}$$

Since A_i is a binary variable, it suffices to consider the evolution of the probability $p_i = P(A_i = 1)$ over time. We assume that

$$\ln(\frac{p_i}{1-p_i}) = \boldsymbol{x}_i\boldsymbol{\beta} \quad \text{or} \quad p_i = \frac{e^{\boldsymbol{x}_i\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_i\boldsymbol{\beta}}}, \tag{10.28}$$

where $\boldsymbol{x}_i$ is a finite-dimensional vector consisting of elements of F_{i-1} and $\boldsymbol{\beta}$ is a parameter vector. Conditioned on $A_i = 1$, D_i is also a binary variable, and we use the following model for $\delta_i = P(D_i = 1|A_i = 1)$,

$$\ln(\frac{\delta_i}{1-\delta_i}) = \boldsymbol{z}_i\boldsymbol{\gamma} \quad \text{or} \quad \delta_i = \frac{e^{\boldsymbol{z}_i\boldsymbol{\gamma}}}{1+e^{\boldsymbol{z}_i\boldsymbol{\gamma}}}, \tag{10.29}$$

where $\boldsymbol{z}_i$ is a finite-dimensional vector consisting of elements of F_{i-1} and $\boldsymbol{\gamma}$ is a parameter vector. To allow for asymmetry between positive and negative price changes, we assume that

$$S_i|(D_i, A_i = 1) \sim 1 + \begin{cases} g(\lambda_{u,i}) & \text{if } D_i = 1, A_i = 1 \\ g(\lambda_{d,i}) & \text{if } D_i = -1, A_i = 1, \end{cases} \tag{10.30}$$

where $g(\lambda)$ is a geometric distribution with parameter λ and the parameters $\lambda_{j,i}$ evolve over time as

$$\ln(\frac{\lambda_{j,i}}{1-\lambda_{j,i}}) = \boldsymbol{w}_i\boldsymbol{\theta}_j \quad \text{or} \quad \lambda_{j,i} = \frac{e^{\boldsymbol{w}_i\boldsymbol{\theta}_j}}{1+e^{\boldsymbol{w}_i\boldsymbol{\theta}_j}}, \quad j = u, d, \tag{10.31}$$

where $\boldsymbol{w}_i$ is again a finite-dimensional explanatory variables in F_{i-1} and $\boldsymbol{\theta}_j$ is a parameter vector.

In Eq. (10.30), the probability mass function of a random variable x, which follows the geometric distribution $g(\lambda)$, is

$$p(x = m) = \lambda(1-\lambda)^m, \quad m = 0, 1, 2, \cdots.$$

We added 1 to the geometric distribution so that the price change, if it occurs, is at least 1 tick. In Eq. (10.31), we take the logistic transformation to ensure that $\lambda_{j,i} \in [0, 1]$.

The previous specification classifies the ith trade, or transaction, into one of three categories:

1. no price change: $A_i = 0$ and the associated probability is $(1 - p_i)$;

2. a price increase: $A_i = 1$, $D_i = 1$, and the associated probability is $p_i\delta_i$. The size of the price increase is governed by $1 + g(\lambda_{u,i})$.

3. a price drop: $A_i = 1$, $D_i = -1$, and the associated probability is $p_i(1 - \delta_i)$. The size of the price drop is governed by $1 + g(\lambda_{d,i})$.

Let $I_i(j)$ for $j = 1, 2, 3$ be the indicator variables of the prior three categories. That is, $I_i(j) = 1$ if the jth category occurs and $I_i(j) = 0$ otherwise. The log likelihood function of Eq. (10.27) becomes

$$\begin{aligned}
&\ln[P(y_i|F_{i-1})] \\
= \;& I_i(1)\ln[(1-p_i)] + I_i(2)[\ln(p_i) + \ln(\delta_i) + \ln(\lambda_{u,i}) + (S_i - 1)\ln(1-\lambda_{u,i})] \\
+ \;& I_i(3)[\ln(p_i) + \ln(1-\delta_i) + \ln(\lambda_{d,i}) + (S_i - 1)\ln(1-\lambda_{d,i})],
\end{aligned}$$

and the overall log likelihood function is

$$\ln[P(y_1, \cdots, y_n|F_0)] = \sum_{i=1}^{n} \ln P(y_i|F_{i-1})], \tag{10.32}$$

which is a function of parameters $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\theta}_u$, and $\boldsymbol{\theta}_d$.

Example 4. (Example 5.2 of Tsay, 2001). We illustrate the decomposition model by analyzing the intraday transactions of IBM stock from November 1, 1990 to January 31, 1991. There were 63 trading days and 59,838 intraday transactions in the normal trading hours. The explanatory variables used are

1. A_{i-1}: The action indicator of the previous trade (i.e., the $[i-1]$th trade within a trading day),
2. D_{i-1}: The direction indicator of the previous trade,
3. S_{i-1}: The size of the previous trade.
4. V_{i-1}: The volume of the previous trade, divided by 1000,
5. Δt_{i-1}: Time duration from the $(i-2)$th to $(i-1)$th trade,
6. BA_i: The bid-ask spread prevailing at the time of transaction.

Because we use lag-1 explanatory variables, the actual sample size is 59,775. It turns out that V_{i-1}, Δt_{i-1} and BA_i are not statistically significant for the model entertained. Thus, only the first three explanatory variables are used. The model employed is

$$\begin{aligned}
\ln\left(\frac{p_i}{1-p_i}\right) &= \beta_0 + \beta_1 A_{i-1} \\
\ln\left(\frac{\delta_i}{1-\delta_i}\right) &= \gamma_0 + \gamma_1 D_{i-1} \\
\ln\left(\frac{\lambda_{u,i}}{1-\lambda_{u,i}}\right) &= \theta_{u,0} + \theta_{u,1} S_{i-1} \\
\ln\left(\frac{\lambda_{d,i}}{1-\lambda_{d,i}}\right) &= \theta_{d,0} + \theta_{d,1} S_{i-1}.
\end{aligned} \tag{10.33}$$

The parameter estimates, using the log-likelihood function in Eq. (10.32), are given in Table 5. The estimated simple model shows some dynamic dependence in the price change. In particular, the trade-by-trade price changes of IBM stock exhibit some appealing features:

Table 5: Parameter Estimates of The ADS Model in Eq. (10.33) For IBM Intraday Transactions: 11/01/90 to 1/31/91.

Parameter	β_0	β_1	γ_0	γ_1
Estimate	-1.057	0.962	-0.067	-2.307
Std.Err.	0.104	0.044	0.023	0.056
Parameter	$\theta_{u,0}$	$\theta_{u,1}$	$\theta_{d,0}$	$\theta_{d,1}$
Estimate	2.235	-0.670	2.085	-0.509
Std.Err.	0.029	0.050	0.187	0.139

1. The probability of a price change depends on the previous price change. Specifically, we have

$$P(A_i = 1|A_{i-1} = 0) = 0.258, \quad P(A_i = 1|A_{i-1} = 1) = 0.476.$$

The result indicates that a price change may occur in clusters and, as expected, most transactions are without price change. When no price change occurred at the $(i-1)$th trade, then only about one out of four trades in the subsequent transaction has a price change. When there is a price change at the $(i-1)$th transaction, the probability of a price change in the ith trade increases to about 0.5.

2. The direction of price change is governed by

$$P(D_i = 1|F_{i-1}, A_i) = \begin{cases} 0.483 & \text{if } D_{i-1} = 0 \text{ (i.e., } A_{i-1} = 0) \\ 0.085 & \text{if } D_{i-1} = 1,\ A_i = 1 \\ 0.904 & \text{if } D_{i-1} = -1,\ A_i = 1. \end{cases}$$

This result says that (a) if no price change occurred at the $(i-1)$th trade, then the chances for a price increase or decrease at the ith trade are about even; and (b) the probabilities of consecutive price increases or decreases are very low. The probability of a price increase at the ith trade given that a price change occurs at the ith trade and there was a price increase at the $(i-1)$th trade is only 8.6%. However, the probability of a price increase is about 90% given that a price change occurs at the ith trade and there was a price decrease at the $(i-1)$th trade. Consequently, this result shows the effect of bid-ask bounce and supports price reversals in high-frequency trading.

3. There is weak evidence suggesting that big price changes have a higher probability to be followed by another big price change. Consider the size of a price increase. We have

$$S_i|(D_i = 1) \sim 1 + g(\lambda_{u,i}), \quad \lambda_{u,i} = 2.235 - 0.670 S_{i-1}.$$

Using the probability mass function of a geometric distribution, we obtain that the probability of a price increase by one tick is 0.827 at the ith trade if the transaction results in a price increase and $S_{i-1} = 1$. The probability reduces to 0.709 if $S_{i-1} = 2$ and to 0.556 if $S_{i-1} = 3$. Consequently, the probability of a large S_i is proportional to S_{i-1} given that there is a price increase at the ith trade.

11 The PCD model

In this section, we introduce a model that considers jointly the process of price change and the associated duration. As mentioned before, many intraday transactions of a stock result in no price change. Those transactions are highly relevant to trading intensity, but they do not contain direct information on price movement. Therefore, to simplify the complexity involved in modeling price change, we focus on transactions that result in a price change and consider a price change and duration (PCD) model to describe the multivariate dynamics of price change and the associated time duration.

We continue to use the same notation as before, but the definition is changed to transactions with a price change. Let t_i be the calendar time of the ith price change of an asset. As before, t_i is measured in seconds from midnight of a trading day. Let P_{t_i} be the transaction price when the ith price change occurred and $\Delta t_i = t_i - t_{i-1}$ be the time duration between price changes. In addition, let N_i be the number of trades in the time interval (t_{i-1}, t_i) that result in no price change. This new variable is used to represent trading intensity during a period of no price change. Finally, let D_i be the direction of the ith price change with $D_i = 1$ when price goes up and $D_i = -1$ when the price comes down, and let S_i be the size of the ith price change measured in ticks. Under the new definitions, the price of a stock evolves over time by

$$P_{t_i} = P_{t_{i-1}} + D_i S_i, \tag{11.34}$$

and the transactions data consist of $\{\Delta t_i, N_i, D_i, S_i\}$ for the ith price change. The PCD model is concerned with the joint analysis of $(\Delta t_i, N_i, D_i, S_i)$.

To illustrate the relationship among the price movements of all transactions and those of transactions associated with a price change, we consider the intraday tradings of IBM stock on November 21, 1990. There were 726 transactions on that day during the normal trading hours, but only 195 trades resulted in a price change. Figure 8 shows the time plot of the price series for both cases. As expected, the price series are the same.

The PCD model decomposes the joint distribution of $(\Delta t_i, N_i, D_i, S_i)$ given F_{i-1} as

$$\begin{aligned} & f(\Delta t_i, N_i, D_i, S_i | F_{i-1}) \\ &= f(S_i | D_i, N_i, \Delta t_i, F_{i-1}) f(D_i | N_i, \Delta t_i, F_{i-1}) f(N_i | \Delta t_i, F_{i-1}) f(\Delta t_i | F_{i-1}). \end{aligned} \tag{11.35}$$

This partition enables us to specify suitable econometric models for the conditional distributions and, hence, to simplify the modeling task. There are many ways to specify models for the conditional distributions. A proper specification might depend on the asset under study. Here we employ the specifications used by McCulloch and Tsay (2000), who use generalized linear models for the discrete-valued variables and a time series model for the continuous variable $\ln(\Delta t_i)$.

For the time duration between price changes, we use the model

$$\ln(\Delta t_i) = \beta_0 + \beta_1 \ln(\Delta t_{i-1}) + \beta_2 S_{i-1} + \sigma \epsilon_i, \tag{11.36}$$

where σ is a positive number and $\{\epsilon_i\}$ is a sequence of iid $N(0,1)$ random variables. This is a multiple linear regression model with lagged variables. Other explanatory variables can be added if necessary. The log transformation is used to ensure the positiveness of time duation.

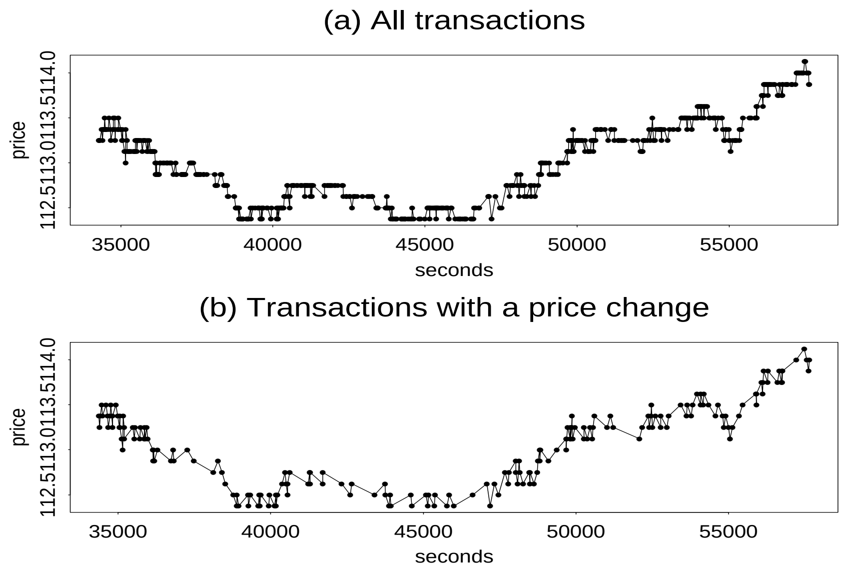

Figure 8: Time plots of the intraday transaction prices of IBM stock on November 21, 1990: (a) all transactions, and (b) transactions that resulted in a price change.

The conditional model for N_i is further partitioned into two parts because empirical data suggest a concentration of N_i at 0. The first part of the model for N_i is the logit model

$$p(N_i = 0|\Delta t_i, F_{i-1}) = \text{logit}[\alpha_0 + \alpha_1 \ln(\Delta t_i)], \tag{11.37}$$

where $\text{logit}(x) = \exp(x)/[1+\exp(x)]$, whereas the second part of the model is

$$N_i|(N_i > 0, \Delta t_i, F_{i-1}) \sim 1 + g(\lambda_i), \quad \lambda_i = \frac{\exp[\gamma_0 + \gamma_1 \ln(\Delta t_i)]}{1+\exp[\gamma_0 + \gamma_1 \ln(\Delta t_i)]}, \tag{11.38}$$

where $\sim$ means "is distributed as," and $g(\lambda)$ denotes a geometric distribution with parameter λ, which is in the interval (0,1).

The model for direction D_i is

$$D_i|(N_i, \Delta t_i, F_{i-1}) = \text{sign}(\mu_i + \sigma_i \epsilon), \tag{11.39}$$

where ϵ is a $N(0,1)$ random variable, and

$$\begin{aligned} \mu_i &= \omega_0 + \omega_1 D_{i-1} + \omega_2 \ln(\Delta t_i) \\ \ln(\sigma_i) &= \beta|\sum_{j=1}^{4} D_{i-j}| = \beta|D_{i-1} + D_{i-2} + D_{i-3} + D_{i-4}|. \end{aligned}$$

In other words, D_i is governed by the sign of a normal random variable with mean μ_i and variance σ_i^2. A special characteristic of the prior model is the function for $\ln(\sigma_i)$. For intraday transactions, a key feature is the *price reversal* between consecutive price changes. This feature is modeled by the dependence of D_i on D_{i-1} in the mean equation with a negative ω_1 parameter. However, there exists occasional local trend in the price movement. The previous variance equation allows for such a local trend by increasing the uncertainty in the direction of price movement when the past data showed evidence of a local trend. For a normal distribution with a fixed mean, increasing its variance makes a random draw have the same chance to be positive and negative. This in turn increases the chance for a sequence of all positive or all negative draws. Such a sequence produces a local trend in price movement.

To allow for different dynamics between positive and negative price movements, we use different models for the size of a price change. Specifically, we have

$$\begin{aligned} S_i|(D_i = -1, N_i, \Delta t_i, F_{i-1}) &\sim p(\lambda_{d,i}) + 1, \quad \text{with} \\ \ln(\lambda_{d,i}) &= \eta_{d,0} + \eta_{d,1} N_i + \eta_{d,2} \ln(\Delta t_i) + \eta_{d,3} S_{i-1} \end{aligned} \tag{11.40}$$

$$\begin{aligned} S_i|(D_i = 1, N_i, \Delta t_i, F_{i-1}) &\sim p(\lambda_{u,i}) + 1, \quad \text{with} \\ \ln(\lambda_{u,i}) &= \eta_{u,0} + \eta_{u,1} N_i + \eta_{u,2} \ln(\Delta t_i) + \eta_{u,3} S_{i-1}, \end{aligned} \tag{11.41}$$

where $p(\lambda)$ denotes a Poisson distribution with parameter λ, and 1 is added to the size because the minimum size is 1 tick when there is a price change.

The specified models in Eqs. (11.36)−(11.41) can be estimated jointly by either the maximum likelihood method or the Markov Chain Monte Carlo methods. Based on Eq. (11.35), the models consist of six conditional models that can be estimated separately.

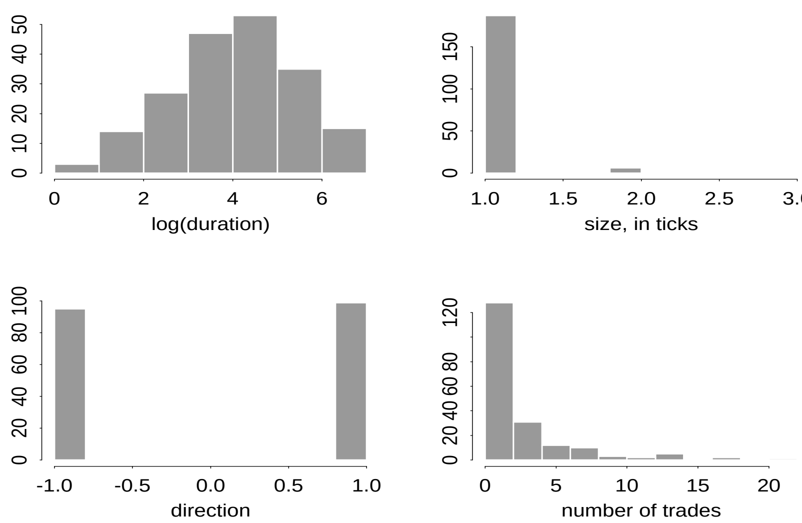

Figure 9: Histograms of intraday transactions data for IBM stock on November 21, 1990: (a) log durations between price changes, (b) direction of price movement, (c) size of price change measured in ticks, and (d) number of trades without a price change.

Example 5. (Example 5.5 of Tsay, 2001). Consider the intraday transactions of IBM stock on November 21, 1990. There are 194 price changes within the normal trading hours. Figure 9 shows the histograms of $\ln(\Delta t_i)$, N_i, D_i, and S_i. The data for D_i are about equally distributed between "upward" and "downward" movements. Only a few transactions resulted in a price change of more than 1 tick; as a matter of fact, there were seven changes with two ticks and one change with three ticks. Using Markov Chain Monte Carlo (MCMC) methods (see Chapter 10 of Tsay, 2001), we obtained the following models for the data. The reported estimates and their standard deviations are the posterior means and standard deviations of MCMC draws with 9500 iterations. The model for the time duration between price changes is

$$\ln(\Delta t_i) = 4.023 + 0.032\ln(\Delta t_{i-1}) - 0.025S_{i-1} + 1.403\epsilon_i,$$

where standard deviations of the coefficients are 0.415, 0.073, 0.384, and 0.073, respectively. The fitted model indicates that there was no dynamic dependence in the time duration. For the N_i variable, we have

$$Pr(N_i > 0|\Delta t_i, F_{i-1}) = \text{logit}[-0.637 + 1.740\ln(\Delta t_i)],$$

where standard deviations of the estimates are 0.238 and 0.248, respectively. Thus, as expected, the number of trades with no price change in the time interval (t_{i-1}, t_i) depends positively on the length of the interval. The magnitude of N_i when it is positive is

$$N_i|(N_i > 0, \Delta t_i, F_{i-1}) \sim 1 + g(\lambda_i), \quad \lambda_i = \frac{\exp[0.178 - 0.910\ln(\Delta t_i)]}{1 + \exp[0.178 - 0.910\ln(\Delta t_i)]},$$

where standard deviations of the estimates are 0.246 and 0.138, respectively. The negative and significant coefficient of $\ln(\Delta t_i)$ means that N_i is positively related to the length of the duration Δt_i because a large $\ln(\Delta t_i)$ implies a small λ_i, which in turn implies higher probabilities for larger N_i; see the geometric distribution in Eq. (10.30).

The fitted model for D_i is

$$\begin{aligned} \mu_i &= 0.049 - 0.840D_{i-1} - 0.004\ln(\Delta t_i) \\ \ln(\sigma_i) &= 0.244|D_{i-1} + D_{i-2} + D_{i-3} + D_{i-4}|, \end{aligned}$$

where standard deviations of the parameters in the mean equation are 0.129, 0.132 and 0.082, respectively, whereas that for the parameter in the variance equation is 0.182. The price reversal is clearly shown by the highly significant negative coefficient of D_{i-1}. The marginally significant parameter in the variance equation is exactly as expected. Finally, the fitted models for the size of a price change are

$$\begin{aligned} \ln(\lambda_{d,i}) &= 1.024 - 0.327N_i + 0.412\ln(\Delta t_i) - 4.474S_{i-1} \\ \ln(\lambda_{u,i}) &= -3.683 - 1.542N_i + 0.419\ln(\Delta t_i) + 0.921S_{i-1}, \end{aligned}$$

where standard deviations of the parameters for the "down size" are 3.350, 0.319, 0.599, and 3.188, respectively, whereas those for the "up size" are 1.734, 0.976, 0.453, and 1.459. The interesting estimates of the prior two equations are the negative estimates of the coefficient of N_i. A large N_i means there were more transactions in the time interval (t_{i-1}, t_i) with no price change. This can be taken as evidence of no new information available in the time interval (t_{i-1}, t_i). Consequently, the size for the price change at t_i should be small. A small $\lambda_{u,i}$ or $\lambda_{d,i}$ for a Poisson distribution gives precisely that.

In summary, granted that a sample of 194 observations in a given day may not contain sufficient information about the trading dynamic of IBM stock, but the fitted models appear to provide some sensible results. McCulloch and Tsay (2001) extend the PCD model to a hierarchical framework to handle all the data of the 63 trading days between November 1, 1990 and January 31, 1991. Many of the parameter estimates become significant in this extended sample, which has more than 19,000 observations. For example, the overall estimate of the coefficient of $\ln(\Delta t_{i-1})$ in the model for time duration ranges from 0.04 to 0.1, which is small, but significant.

12 Hierarchical Models

Hierarchical models have been extensively used in recent years to deal with this situation; see Gelman, Carlin, Stern, and Rubin (1995, ch. 5). We employ a PCD model for each trading day, and model the variation in parameters from day to day. The PCD model actually consists of six model components. There are the four basic models for Δt, N, D, and S with the models for S and N each having two components. We apply the hierarchical modeling strategy separately to each of these six components.

We first provide detailed discussion of hierarchical modeling of the simple logit component of the model for N in order to illustrate the approach. We then present results for all six model components.

12.1 The Hierarchical Modeling for the Logit Component of N

Let $\theta_j = (\alpha_{0j}, \alpha_{1j})$ from equation (11.37) (section 4.1), where j indexes the day. Thus, θ_j represents the parameters for the logit component of the N model on day j. Note that in order to roughly orthogonalize the intercept and slope parameters the overall (using all the days) mean of $\ln(\Delta t) = 3.5$ has been subtracted from all of the $\ln(\Delta t)$ values (on all days).

Figure 10 displays the time series of estimates (posterior means) of α_{0j} and α_{1j} obtained by applying the PCD to each day. The average intercept estimate is about .04 and the values range from -1.2 to .93. The average slope estimate is about 1.08 and the slopes range from .65 to 1.8. The average intercept is small (logit(.04) = exp(.04)/(1+exp(.04)) = .51). However, logit(-1.2) = .22 and logit(.93) = .74 so that the day to day variation in the intercepts is substantial. Since the .1 and .9 quantiles of $\ln(\Delta t)$ are 1.4 and 5.3 (see Figure 9(a)), we see that the slopes suggest dependence of the event $N > 0$ on $\ln(\Delta t)$ and the day to day variation in the slope estimates is substantial.

From Figure 10 we might wonder if day 28 is unusual because of the relatively large slope and small intercept. There is also the suggestion that the intercepts are larger after day 53 (inclusive). However, neither of these features are clearly distinguishable from the overall variation. In addition, we must remember that the quantities plotted are only estimates and no attempt is made in the figure to represent our uncertainty.

We would like to elaborate our model to include a description of the variation in parameters from day to day. The way in which we elaborate our model depends on the goal of our study. If our goal was prediction for subsequent days, capturing any temporal pattern (structural shift etc.) would be quite important. Instead, we ask the pair of somewhat simpler questions: (i) overall (that is, over all the days in our sample), what are the PCD parameters like, and (ii) for which days is there strong evidence that the parameter values are "different". For this goal the standard iid shrinkage model is suitable and convenient: we let $\theta_j \sim N(\theta_*, \Sigma_*)$ iid. Given choices for the prior distrubution of (θ_*, Σ_*) we can compute the posterior distribution of these quantities and each θ_j. The parameter θ_* can be regarded as the "overall mean" of the parameter θ_j across all trading days and the parameter Σ_* describes the day to day variation. This model also allows for adaptive shrinkage of the θ_j towards the overall mean θ_*. If the data for a day j suggests a vector θ_j different from the rest relative to the overall variation but the strength of the evidence in the data is weak, then the support of the

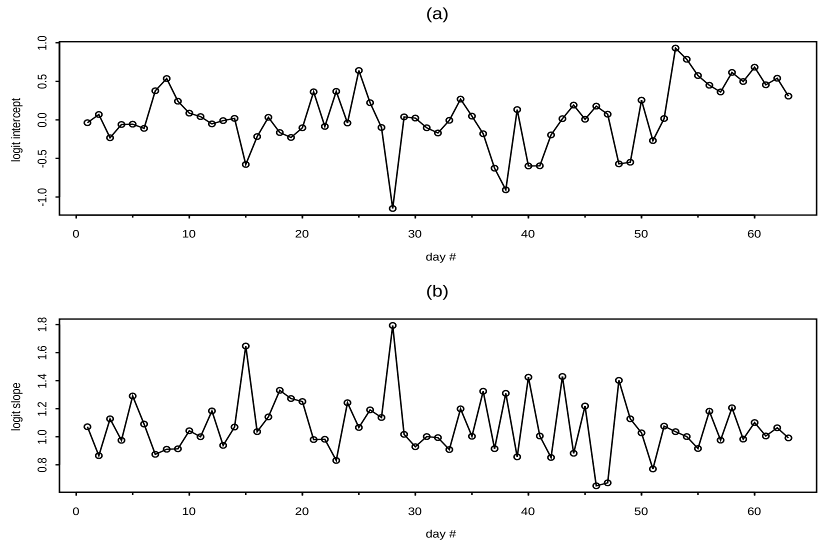

Figure 10: Time series plots of posterior means for the logit component of N by applying the PCD model to each trading day: (a) intercept α_{0j}, (b) slope α_{1j}.

posterior will be shrunk towards the overall mean. If the evidence is strong however, the shrinkage will be negligable. In this way our model strikes an adaptive compromise between the extremes of treating each day separately and lumping them all together. Thus, the posterior distribution of θ_* answers question (i) and days for which the posterior of θ_j are not shrunk to θ_* are the days corresponding to question (ii).

To implement the approach we must first choose a prior for (θ_*, Σ_*) and then compute the posterior. The approach of Barnard, McCulloch, and Meng (2000) is used. The chosen priors are extremely diffuse. θ_* and Σ_* are independent. The components of θ_* are iid normal with mean 0 and standard deviation 1000. The square roots of the diagonals of Σ_* (the standard deviations) are iid log-normal with a mean of -.5 and a standard deviation of 1.5. The diagonal elements of Σ_* are independent of the correlation matrix which is uniformly distributed on the set of positive definite matrices having unit diagonals.

Figure 11 compares the estimates (posterior means) of θ_j obtained by applying the PCD separately to each day (those displayed in figure 10) with those obtained from the hierarchical model. Panel (a) plots the daily intercept estimates obtained from the hierarchical model on the vertical axis vs the non-hierarchical estimates on the horizontal axis. Panel (b) is the corresponding plot for the slopes. The line $y = x$ is drawn through both panels (a) and (b). Panel (c) plots the time series of intercept estimates where the hierachical esimtates are connected by the line while the non-hierarchical estimates are plotted with an O. Panel (d) is the slope version of panel (c). We see that while most intercept estimates are only slightly altered by the shrinkage, the few unusually small values are substantially shrunk towards 0 so that, for example, day 28 no longer looks like a dramatic outlier. The shrinkage of the slope is quite dramatic as the shrunk values are much more tightly clustered near one. Even with the shrinkage, there is still the suggestion that the intercepts jump to larger values for the last several days.

Figure 12 displays the posterior of (θ_*, Σ_*). Panels (a)-(e) display draws from the marginal poste-

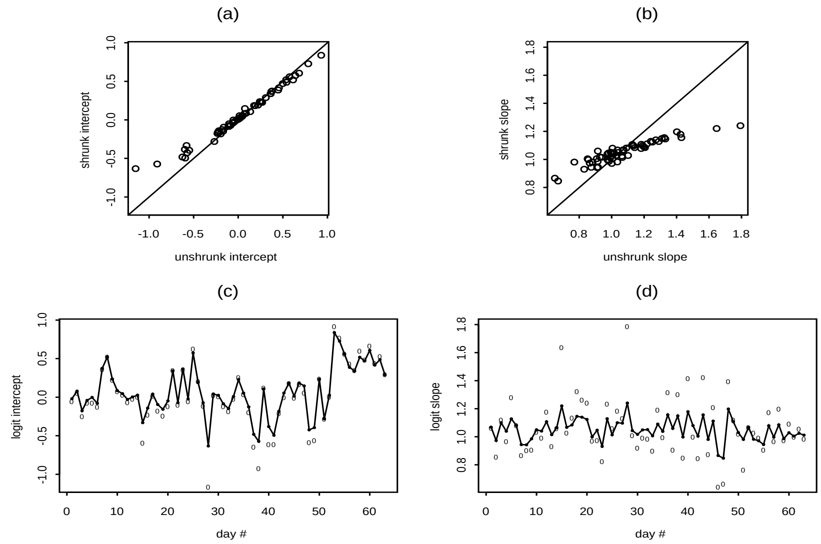

Figure 11: Comparison between hierarchical and non-hierarchical models: (a) scatterplots of posterior means of the intercept, (b) scatterplot of posterior means of the slope, (c) time series plots of intercept posterior means with solid line denoting estimates under the hierarchical model, (d) time series plots of slope posterior means with solid line denoting estimates of the hierarchical model.

riors of the first component of θ_* (the "overall" intercept), the second component of θ_* (the "overall" slope), the square root of the first diagonal of Σ_* (the standard deviation describing the day to day variation in intercepts), the square root of the second diagonal of Σ_* (the standard deviation describing the day to day variation in slopes), and single correlation from the two by two Σ_* (the day to day correlation between intercepts and slopes). The solid line in panel (f) displays a kernel estimate of the posterior of the overall intercept (same quantity displayed in panel (a)) and the dashed line displays the kernel estimate of the posterior of the intercept where only the data from the first day in our sample is used. The relative tightness of the solid kernel illustrates the effect of pooling information from all 63 days.

12.2 Hierarchical Results for all Model Components

We now present results obtained from applying the hierarchical model to each of the six components of the PCD model. For each model we present the .025, .5, and .975 posterior quantiles for the "overall mean" and the standard deviation of each parameter (components of θ_* and square roots of diagonal elements of Σ_* in section 12.1). The mean gives us an overall idea of the parameter over all 63 days and the standard deviation describes the variation of the parameter from day to day. We do not report the posterior distributions of the day to day correlations between pairs of parameters.

Table 6 contains the quantiles. The table has three columns. The first column (labeled "parameter") identifies the particular parameter of the model where the notation of section 4.1 is used. The second column and third columns (labeled "mean" and "stan dev") give the overall mean and standard deviation quantiles. The three quantiles are listed from least to largest.

So, for example, the two rows labeled α_0 and α_1 under the heading "Logit Model for N Positive"

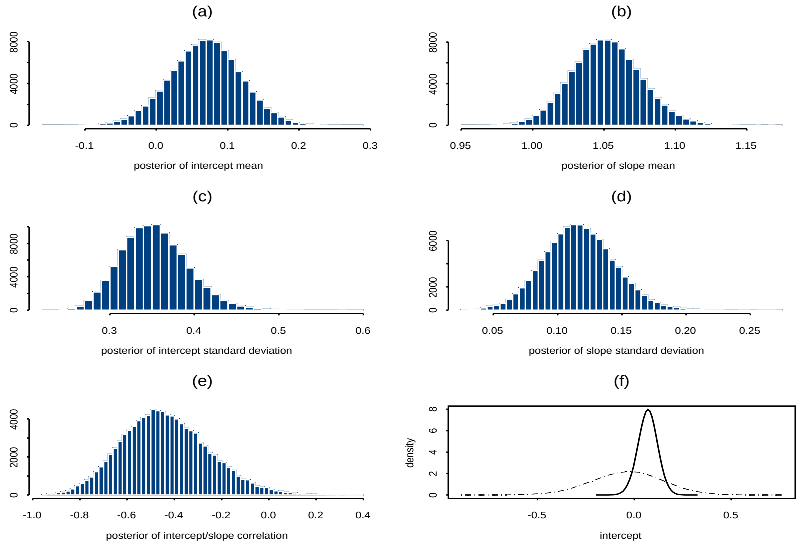

Figure 12: Posterior density functions of overall parameters of the hierarchical model.

give the intervals for the logit model discussed in detail in section 12.1. To connect the Table with our detailed discussion, the quantiles (-.029, .069, .164) and (.283, .351, .436) in the row labeled α_0 summarize panels (a) and (c) of figure 12. The quantiles (1.01, 1.05, 1.10) and (.065, .116, .172) summarize panels (b) and (d). Note than for the time duration model (equation (11.36), section 4.1) only the β's are modelled hierarchically so that the error standard deviation σ varies "freely" from day to day. In all other model components all parameters are included in the hierarchical setup.

The table reveals how spread out our priors are relative to the posteriors. For all the mean parameters the prior was $N(0, 1000^2)$. All of the posterior intervals in the mean column are extremely tight relative to this prior. For all the standard deviation parameters the prior was $\sigma \sim \exp(N(-.5, 1.5^2))$. This prior is again very diffuse relative the the posterior intervals. The prior is very skewed with a 99% quantile of about 20. The 1% and 99% prior quantiles of the $\sigma | \sigma < 1$ (the prior conditional on where most of posterior supports are) are .014 and .98 so that restricted to this more realistic interval the prior is still quite spread out.

The results for α_1 and γ_1 both reflect a strong dependence of N on the duration. Detailed results for α_1 have been presented in figure 11. The interval for the mean of γ_1 is quite tight around values which are of practical significance. If we take the median of -.77 as a point estimate and consider a change of 4 in $\ln(\Delta t)$ as possible then an increase in duration could lead to as substantial decrease in the probability parameter of the geometric distribution (λ of equation (11.38)). Since for the geometric distribution smaller probabilities make larger outcomes more likely this means that conceivable increases in $\ln(\Delta t)$ suggest larger values of N. The relatively small interval for the standard deviation (point estimate .088) tells us that the day to day variation in γ_1 is such that essentially the same result is obtained for all days.

Another parameter which clearly suggest an effect of practical importance for all days is ω_1 in the direction model. The posterior median of the overall mean is -.879. Since this parameter is the coefficient of lagged D which is always ± 1 the estimate suggests an important effect given the probit type specification in equation (11.39). The estimate (.099) and quantiles of the daily standard

Table 6: 95% Posterior Intervals for the Hierarchical Parameters

parameter	mean	stan dev
Time Duration Model		
β_0	(3.74, 3.85, 3.97)	(.319, .395, .494)
β_1	(.072, .094, .118)	(.045, .066, .090)
β_2	(-.281, -.212, -.152)	(.077, .132, .204)
Logit Model for N Positive		
α_0	(-.029, .069, .164)	(.283, .351, .436)
α_1	(1.01, 1.05, 1.10)	(.065, .116, .172)
Geometric Model for Positive N		
γ_0	(-.167, -.064, .044)	(.314, .384, .476)
γ_1	(-.807, -.771, -.737)	(.057, .088, .124)
Direction Model		
ω_0	(-.011, .012, .034)	(.010, .031, .065)
ω_1	(-.914, -.879, -.844)	(.064, .099, .139)
ω_2	(-.048, -.032, -.017)	(.004, .016, .038)
β	(-.056, -.022, .012)	(.047, .082, .125)
Poisson Model for Price Down		
$\eta_{d,0}$	(-3.04, -2.80, -2.58)	(.608, .750, .938)
$\eta_{d,1}$	(-.144, -.089, -.044)	(.096, .131, .182)
$\eta_{d,2}$	(-.232, -.175, -.114)	(.026, .106, .174)
$\eta_{d,3}$	(.526, .614, .713)	(.177, .245, .340)
Poisson Model for Price Up		
$\eta_{u,0}$	(-3.43, -3.13, -2.86)	(.773, .973, 1.21)
$\eta_{u,1}$	(-.433, -.319, -.231)	(.179, .257, .367)
$\eta_{u,2}$	(-.177, -.125, -.067)	(.041, .101, .170)
$\eta_{u,3}$	(.817, .939, 1.07)	(.315, .409, .518)

deviation of ω_1 clearly indicate that a comparable and important effect is likely to be present in most days. The negative sign indicates that this parameter captures the "bounce" behavior of the price: an increase is often followed by a decrease and vice versa.

The posteriors of other parameters suggest additional patterns in the data although not as stongly: a large change in price (large S) is more likely if the previous change was large ($\eta_{d,3}$ and $\eta_{u,3}$), a longer duration makes a price fall more likely (ω_2), and a large price change leads to a shorter duration. Overall, the only parameters whose 95% posterior intervals for the mean parameter include zero are the three intercepts α_0, γ_0, and ω_0, and β.

Some days are identified as unusual in that even estimates obtained from the shrinkage model look unusual compared to those of other days. Panel (b) of figure 13 plots the time series of daily shrunk estimates of the parameter $\eta_{u,2}$ in the Poisson model from price increases. We see that the estimate for day 3 is unusually large. Panel (a) of figure 13 plots the time series of prices for that day (with the first price of the first day subtracted off). In the price plot there are five "spikes" indicating a sharp price increase followed by an offsetting decrease. The reader may wish to compare panel (a) to panel (b) of figure 8 which is a more "typical" day in the judgement of the authors. Another unusual day is illustrate by figure 14. Panels (a) is the price series for day 25 and panel (b) is the daily shrunk estimates of the parameter ω_0 in the direction model. The estimate for day 25 is unusually small. Actually, several parameters have unusual estimates for this day. The price series for this day has a drop followed by several spikes which seem to reach back up to the pre-drop level. A plausible explanation for the sharp price changes in day 25 is the limit orders, because the transaction price of a trade associated with a sharp price increase was close to those of trades that occurred in the morning of the day.

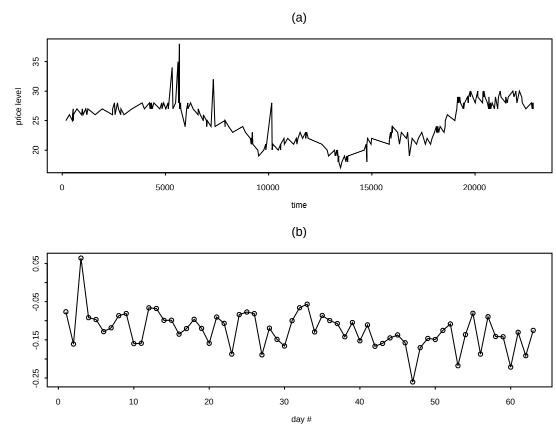

Figure 13: (a) Time plot of the price series for day 3; (b) Time plot of daily shrunk estimates of the parameter $\eta_{u,2}$ in the Poisson model from price increases.

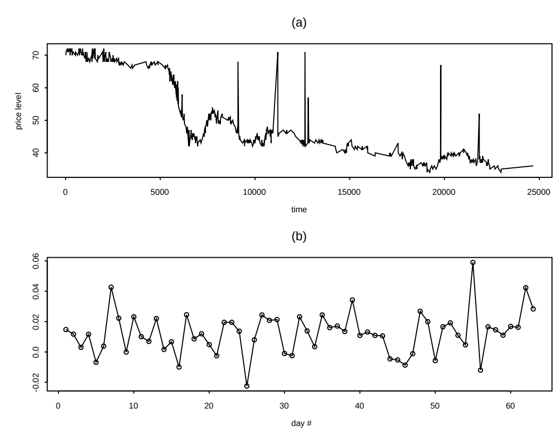

Figure 14: (a) Time plot of the price series for day 25; (b) Time plot of the daily shrunk estimates of the parameter ω_0 in the direction model.

References

Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997), *The Econometrics of Financial Markets*, Princeton University Press: New Jersey.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (1995): *Bayesian Data Analysis*, CRC Press: London.

Ghysels, E. (2000), "Some econometric recipes for high-frequency data cooking," *Journal of Business and Economic Statistics*, 18, 154-163.

Hasbrouck, J. (1992), *Using the TORQ database*, Stern School of Business, New York University.

Hauseman, J., Lo, A., and MacKinlay, C. (1992), "An ordered probit analysis of transaction stock prices," *Journal of Financial Economics*, 31, 319-379.

McCulloch, R. E., and Tsay, R. S. (2001), "Nonlinearity in high frequency data and hierarchical models," *Studies in Nonlinear Dynamics and Econometrics*, 5, (to appear).

Rydberg, T. H., and Shephard, N. (1998), "Dynamics of trade-by-trade price movements: decomposition and models," Working paper, Nuffield College, Oxford University.

Tsay, R. S. (2001), *Analysis of Financial Time Series*, Wiley: New York.