

ECONOMETRICS OF DYNAMIC CENSORED MODELS

by

Steven Xiangdong Wei

A thesis submitted in conformity with the requirements
for the Degree of Doctor of Philosophy
Graduate Department of Economics
University of Toronto

© Copyright by Steven Xiangdong Wei 1997



National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-28079-9

Canada

ECONOMETRICS OF DYNAMIC CENSORED MODELS

Degree of Doctor of Philosophy 1997
Steven Xiangdong Wei
Graduate Department of Economics
University of Toronto

Abstract

Censored models arise in various dynamic settings. Due to the intractable likelihoods of these dynamic censored models, the existing literature shows little research in this area. Drawing on all the advances of the modern Markov Chain Monte Carlo (MCMC) techniques, this dissertation considers three broad classes of dynamic censored models which are studied separately in three chapters of this book.

Chapter I starts with a simple class of dynamic censored models – dynamic Tobit models. The major obstacle of tackling these models, the intractability of their likelihood functions, is solved by developing a simulation-based estimation procedure. The major statistical innovation is a sampling scheme of the censored (latent) data, which involves the crucial step of applying the Gibbs sampler with data augmentation. The method is examined and demonstrated by means of a Monte Carlo experiment, and applied to a regression model of Japanese exports of passenger cars to U.S. subject to a non-tariff trade barrier.

Censored mechanisms can be more complex. **Chapter II** offers an empirical perspective on the open market operations (of the Federal Reserve System in the United States) by developing a dynamic friction model. Friction here denotes this kind of censored structures that depicts the behavior of reluctant interventions of the open market Desk (the Desk). Various inferential procedures are derived. Results show that: (1) the Desk's interventions are highly discretionary with the intention of market signaling; (2) the operation projections made by the Fed's staff members significantly capture the dynamic effects of the Desk's daily actions; and (3) the Fed's operating procedure experienced a major shift immediately after the stock market crash of October 1987.

Censored data can also arise in GARCH processes. Two censored GARCH models, motivated in **Chapter III**, are developed to address the significance of price limits in futures markets. I compare, through exploiting the posterior draws by the Griddy Gibbs sampler with data augmentation technique, the relative performances of the censored GARCH models and the pure GARCH counterpart in terms of both in-sample fit and out-of-sample predictability. A technique for computing Bayes factor for censored models and the modification of the concept of predictive Bayes factor are developed. Based on U.S. Treasury bill futures data over a period of high volatility and frequent limit moves, my empirical findings conclude that: (1) limit rules in futures markets do matter in GARCH processes of futures returns; (2) a pure GARCH model systematically underestimates the conditional volatilities of futures returns; and (3) there is no evidence showing a distortionary effect of the price limits and the imposition of the limit rules in futures markets seems to be empirically justified.

Acknowledgments

I am deeply indebted to my supervisor Dale J. Poirier for his excellent supervision, constant encouragement and helpful comments, and for opening me the world of Bayesian econometrics. Without him, the current work would not have existed. Also I would like to thank all the other members of my thesis committee, Tom McCurdy (Faculty of Management), Varouj Aivazian, Gary Koop and Efthymios Tsionas for their invaluable advice. Jack Carr, Angelo Melino, Melvyn Fuss, Adonis Yatchew, Luc Bauwens (CORE), Michel Mouchart (CORE), Evans Mike (Department of Statistics), Raymond Kan (Faculty of Management) and Gordon Kemp (University of Essex) also provided helpful criticism. The data sets used in this book are provided by H. Tsurumi (Chapter I), the Federal Reserve Board of the United States (Chapter II) and I.G. Morgan (Chapter III). Samita Sareen, one of my colleagues, offered her kind help on the improvement of my writing. The financial support from University of Toronto Doctor Fellowship, Simco Special Scholarship, and CORE Research Fellowship are gratefully acknowledged.

Last but not least I would like to thank my wife Limei Wang and my daughter Lucy Wei. They have been unfailingly supportive and very understanding of the inevitable disruptions to family life caused by writing this dissertation.

A Bayesian Approach to Dynamic Tobit Models

1 Introduction

Tobit (censored) models have become a class of important limited dependent variable models in the econometrics literature. Since the original work by Tobin (1958), especially after the early 1970's, the basic Tobit framework has been extended and incorporated into many areas of economics. For general references, see Amymyia (1984, 1985), Maddala (1987), Green (1992), Chib (1992) and Geweke (1992). The extension of the Tobit structure into a dynamic model is, however, not a well solved problem, at least in the likelihood paradigm. The major obstacle rooted in a dynamic Tobit model lies in high dimensional integrals, induced by correlation among censored observations, in a likelihood function. Recently, Hajivassiliou, McFadden and Ruud (1996) surveyed the simulation methods on computing multivariate normal rectangle probabilities. Although these methods might be used to facilitate the ML estimations, a general treatment of dynamic Tobit models has not been addressed. This Chapter proposes a Bayesian simulation method which differs substantially from its classical counterpart. While classical simulation methods focus on simulating the probabilities of multivariate (normal) distributions and evaluating the derivatives (with respect to parameters) of the probabilities, the developed Bayesian method here is in essence to rely on sampling from truncated multivariate normal (or Student-t if we have a Student-t version of dynamic Tobit models) distributions. The latter has a strongly computational advantage over the former because it is much easier to sample from a distribution than to evaluate the probabilities of the distribution.

Dynamic Tobit models can be found in real-world applications. The air pollution data (Zeger and Brookmeyer (1986)) subject to lower limits of detection are an example. In economics and finance, such instances can often occur when constraints and/or regulations are imposed. Peristiani (1994) adopted a simplified version of dynamic Tobit models to study the behavior of the discount window borrowing for individual banks. Kodres (1988) proposed a dynamic Tobit model, which is almost the same as the one in this Chapter, to study the effect

of price limits in currency futures markets². The estimation method she used is quite complex and less efficient (see the discussion in Morgan and Trevor (1996)). Japanese exports of passenger cars to the U.S. subject to a non-tariff trade barrier (Zangari and Tsurumi (1994)) are also a potential application of dynamic Tobit models. This application will be further explored in this Chapter.

In the existing literature, a couple of papers are related to dynamic Tobit models. Dagenais (1982) and Zeger and Brookmeyer (1986) have both explored the applicability of the maximum likelihood method to the Tobit models with autocorrelated errors. Their methods had proven very computationally limited until the recent development in computing the multivariate normal probabilities in (multidimensional) rectangles (see Hajivassiliou, McFadden and Ruud (1996)). The maximum pseudo-likelihood method proposed by Zeger and Brookmeyer is tractable but notably inefficient. An additional problem associated with the method is that the consistent estimator of the covariance matrix of all parameters is hardly obtainable. To simplify matters, Robinson (1982) proved that the maximum likelihood estimators of Tobit models are strongly consistent even if data are autocorrelated. The problem appears that the standard Tobit estimators (and/or predictors) are inefficient when censored data are serially correlated.

This Chapter proposes an exact posterior estimation inference for dynamic Tobit models. The primary contribution of the current study is to develop a practical and efficient sampling scheme for the conditional posterior distribution of the latent (i.e., unobserved) data, so that the Gibbs sampler with data augmentation (cf. Geman and Geman (1984), Tanner and Wong (1987) and Gelfand and Smith (1990)) algorithm can be applied. In particular, it has been derived that the unobserved data (conditioning on all parameters and observed data), viewed as a vector of special parameters, can be generated from a group of univariate

²In two subsequent papers by Kodres (1993), and Morgan and Trevor (1996), the basic censored structure is extended to GARCH models. Calzolari and Fiorentini (1996) also developed a classical treatment for Tobit models with GARCH errors. I am currently working on a Bayesian analysis on Tobit-GARCH models.

and multivariate truncated normal distributions. This is in sharp contrast with the classical treatment of the latent observations, which require (repeated) evaluation of the multivariate normal probabilities and their derivatives. The mean and variance of the normal distributions are analytically solved with fully taking account of the information embedded in both the dynamic and censored structures of the models. A decomposition of latent data into substrings in terms of their probabilistic links over time, as suggested by Zeger and Brookmeyer (1986), is used to facilitate the computation. The advantages of the developed approach can be further understood along the following lines.

(i) Despite that the proposed method is used to deal with dynamic Tobit models with Gaussian errors, there will be no difficulty to extend the technique to handle a Student-t version of dynamic Tobit models³ (cf. Chapter II). It seems, however, not so easy for the classical simulation methods to adapt to this fat-tailed situation. For clarity the focus of this Chapter is on white noise Gaussian errors.

(ii) Prior information, if any, can be formally incorporated into the process of making statistical inferences. For example, the linear constraints on the model parameters can be handled easily via-a-vis the Bayesian paradigm (cf. Gelfand, Smith, and Lee (1992)). On the other hand, Jeffrey's prior is also entertained for the cases with no prior information injected.

(iii) One advantage of the proposed method is that it can handle both stationary and non-stationary processes. This is analogous to the treatment of linear dynamic models in a Bayesian context.

(iv) The joint and marginal posterior densities and the moments of functions of all parameters can be easily simulated.

(v) The new method is investigated by means of a Monte Carlo experiment and is shown to

³A combination of the current Chapter and Geweke (1993) would be sufficient.

perform satisfactorily in various circumstances. It is emphasized that even with small sample sizes and/or non-stationary latent processes, the proposed procedure still delivers reasonably good results. Finally, the model is applied to a regression study of Japanese exports of passenger cars to the U.S. subject to a non-tariff trade barrier.

Recently, Zangari and Tsurumi (1994) considered the applicability of the traditional procedures, as well as the Gibbs sampler with data augmentation method, for the Bayesian estimation of a Tobit model with AR(1) errors. There are a number of important differences between the approach here and the ones discussed in their paper. *Firstly*, these models are distinct though there exist some similarities between them. This Chapter focuses attention on Tobit models with a general p th order lagged (unobserved) dependent variables, and their paper centered on Tobit models with AR(1) errors. *Secondly*, while we develop a unified approach to both stationary and non-stationary latent processes, their approaches are only confined to stationary ones. *Finally*, the present research can be easily generated to a Student-t version of the models which is not the case for Zangari and Tsurumi's approaches.

The remainder of the Chapter is organized as follows. Section 2 introduces a dynamic Tobit model and points out its computational difficulty. Section 3 briefly reviews the Gibbs sampler with data augmentation algorithm and Section 4 develops a complete conditional structure of the model with discussions centered on sampling the latent data in the model. To entertain performances of the proposed method, a Monte Carlo experiment in Section 5 is conducted. Section 6 applies the method to a regression model of Japanese car exports to the U.S. with a non-tariff trade barrier. Section 7 concludes the Chapter.

2 The Model

Consider a dynamic Tobit model in which an observation y , at time t , is given by

$$\begin{cases} y_t^* = x_1\beta_1 + \dots + x_{k-p}\beta_{k-p} + \lambda_1 y_{t-1}^* \dots + \lambda_p y_{t-p}^* + \epsilon_t \\ y_t = \max\{y_t^*, 0\} \end{cases} \quad (1)$$

where y_t^* is a latent process with a vector of covariates $x_t: (k-p) \times 1$, $\beta \in R^{(k-p)}$, and the error ϵ_t is white noise Gaussian with mean zero and variance $\sigma^2 > 0$, i.e., $N(0, \sigma^2)$ ⁴.

Prior to proceeding two remarks regarding this model are worth noting: (1) it is only a convenient normalization to set the censored limit at zero. In fact, this model can be easily adapted to account for more general cases, such as a non-zero censoring limit or interval censoring; (2) the Gaussian assumption can be easily relaxed in the direction of the Student-t family to capture thick-tailed data distributions (cf. Geweke (1993)).

In formulation (1), censoring is driven by sampling and thus should impose no impact on the latent process (compare it with the model in footnote [4]). The obstacle to be addressed, however, lies in the fact that the sampling distribution of the model is often analytically intractable. For an illustration, let y denote the uncensored observations, and z the censored ones. Let $f(y, z|\theta)$ denote the joint distribution of y and z , indexed by a parameter vector θ where $\theta = (\beta_1, \dots, \beta_{k-p}, \lambda_1, \dots, \lambda_p, \sigma^{-2})'$. The sampling distribution of this model is

$$L(\theta; y) \propto \int_{-\infty}^0 \dots \int_{-\infty}^0 f(y, z|\theta) dz.$$

An analytic evaluation of the required integrals is generally hopeless with effective approximation only available for simple cases. The exact likelihood of this model is not derived because it will not be explicitly utilized in this Chapter.

⁴ There exists another type of dynamic Tobit models as mentioned in Maddala (1987). It can be generally written as

$$\begin{cases} y_t^* = x_1\beta_1 + \dots + x_{k-p}\beta_{k-p} + \lambda_1 y_{t-1} + \dots + \lambda_p y_{t-p} + \epsilon_t \\ y_t = \max\{y_t^*, 0\} \end{cases}$$

The formulation might be useful when censoring imposes an impact on the latent process of y_t^* . Conventional (classical and Bayesian) methods are still applicable to this model because of the observability of the lagged terms.

Inference considerations: Bayesian statistics is the discipline of using data to revise beliefs. The posterior inference of interest here is to learn about the unknown parameters in the light of data and prior. For a given prior (belief) $\pi(\theta)$, by Bayes theorem, the posterior (updated belief) of interest is given by

$$p(\theta|y) \propto \pi(\theta)L(\theta; y).$$

The posterior computation can be extremely difficult if evaluated directly. To avoid this difficulty, in Section 4, we will develop a sampling scheme of the latent data conditioning on all parameters and observed data. Then the Gibbs sampler with data augmentation can be successfully applied.

3 The Gibbs Sampler with Data Augmentation

As a special case of the Markov Chain Monte Carlo methods, the Gibbs sampler with/without data augmentation has considerably advanced the practice of Bayesian statistics (for references, see Geman and Geman (1984), Tanner and Wong (1987), Gelfand and Smith (1990), and Tierney (1994)). Outstanding examples of its applications are Chib (1992), Geweke (1992), Albert and Chib (1993), to name only a few. This method is widely used to draw variates from the exact posterior distributions of all parameters and latent data. A brief description of the method is next.

Suppose that z are (censored) latent data. In order to sample from an (intractable) posterior distribution θ where the conditioning on y is suppressed for simplicity, we assume that the complete conditional distributions of z and partitioned $\theta = (\theta_1, \theta_2, \dots, \theta_B)$, i.e. $z|\theta$ and $\theta_s|\{z, \theta_j, j \neq s\}$ are available and have easily sampled forms. Then, initialize $\{\theta\}$ with $\{\theta^{(0)}\}$ in the support of the posterior distribution. The influence of the starting conditions vanishes after a certain number of draws, say, m . The Gibbs sampler with data augmentation algorithm produces posterior variates $\{\theta^{(i)}, z^{(i)}\}$ by sampling iteratively from

the following distributions:

$$z|\{\theta^{(i)}\} \quad \text{and} \quad \theta_s|\{z^{(i)}, \theta_j^{(i)}, \theta_k^{(i-1)}, j < s, k > s\} \quad s = 1, 2, \dots, B \quad (2)$$

After m iterations of the above scheme, the sample $\theta^{(m)} = (\theta_1^{(m)}, \theta_2^{(m)}, \dots, \theta_B^{(m)})$ and $z^{(m)}$ is obtained. Under weak conditions (cf. Gelfand and Smith (1990), or Tanner (1993)), $\{\theta^{(m)}, z^{(m)}\}$ converges in distribution to $\{\theta, z\}$. A particular advantage of this convergence is that given any continuous (or even measurable) function $g(\cdot)$, $g(\theta^{(m)}, z^{(m)})$ converges to $g(\theta, z)$ in distribution. For large m , this leads to approximately M (dependent) draws $\{\theta^{(j)}, z^{(j)}\}$, $j = m+1, m+2, \dots, m+M$, which are called the *Gibbs output* by Chib (1994), from the joint posterior distribution $\{\theta, z\}$. Based on the Gibbs output, posterior inference becomes straightforward. In fact, the marginal posterior densities of θ_s and z can be written as finite mixtures

$$\begin{aligned} \hat{f}(\theta_s) &= M^{-1} \sum_{j=m+1}^{m+M} f(\theta_s|z^{(j)}, \theta_i^{(j)}, i \neq s) \quad s = 1, 2, \dots, B \\ \hat{f}(z) &= M^{-1} \sum_{j=m+1}^{m+M} f(z|\theta^{(j)}) \end{aligned}$$

According to a Rao-Blackwell argument (cf. Gelfand and Smith (1990)), the desired posterior moment estimates of interest are

$$\begin{aligned} \hat{E}(\theta_s) &= M^{-1} \sum_{j=m+1}^{m+M} E(\theta_s|z, \theta_i^{(j)}, i \neq s) \\ \widehat{Var}(\theta_s) &= M^{-1} \sum_{j=m+1}^{m+M} E(\theta_s^2|z, \theta_i^{(j)}, i \neq s) - \{\hat{E}(\theta_s)\}^2. \end{aligned} \quad (3)$$

In practice, two things about the method are of concern: a convergence criterion of the Monte Carlo Markov chains and a measure of the computational accuracy of the estimates. An easily used criterion proposed by Yu and Mykland (1994) has been reported to behave relatively well (see Robert (1995) and Bauwens (1996)). This criterion is a visual inspection of CUMSUM statistics. Let N be the draws of a Monte Carlo Markov Chain (MCMC), noting the Gibbs sampler with or without data augmentation is a special case of MCMC methods. The CUMSUM statistic is given by

$$CS_t = \left(\frac{1}{t} \sum_{i=1}^t \theta^{(i)} - \mu \right) / \sigma$$

where μ and σ are the empirical mean and standard deviation of the N draws. If the MCMC converges, then the plot of CS_t against t should converge smoothly to zero. On the other hand, a long and regular excursion plot of CS_t indicates the absence of convergence of the chain.

The accuracy of these estimates are measured in numerical standard errors, which might be computed by using the well-known batch means method (cf. Ripley (1987)). To implement this method, divide the Gibbs (output) chain into b batches of length G . Denote the mean of each batch as m_i , and the average of the batches as \bar{m} . Then the standard error of the estimate is given by $\{b(b-1)\}^{-1} \sum_{i=1}^b (m_i - \bar{m})^2$.

4 Conditional Distributions

Within this section, we will derive a full conditional structure for the model. The focus of interest is developing a sampling scheme of the (conditional) latent data in the current dynamic setting. This is based on a decomposition of the latent data in terms of their probability links over time.

Prior specification: Bayesian analysis involves **formal** consideration of prior information. The prior distribution of θ describes what is known about θ before the data are actually observed. If the latent process of the model (1) was fully observed, we would have a dynamic linear regression model. In such a circumstance, a normal-gamma informative prior and Jeffrey's diffuse prior (cf. Poirier (1995) and Zellner (1971)) have been widely adopted in the literature for their conveniences⁵. Recalling that the censorship in (1) is only a sampling property, we can reasonably understand that the censoring here should impose no effect on one's prior belief about the parameters. Both types of the priors are discussed here.

⁵ $\{\beta, \sigma^{-2}\}$ is called to have a normal-gamma distribution if $\beta|\sigma^{-2}$ follows a normal distribution and σ^{-2} a gamma distribution. See Poirier (1995, p. 128) for this definition.

A normal-gamma prior for the parameters (θ, σ^{-2}) is given by

$$\pi(\beta, \sigma^{-2} | \underline{\beta}, \underline{Q}, \underline{s}^{-2}, \underline{v}) = \phi_k(\beta | \underline{\beta}, \sigma^2 \underline{Q}) \gamma(\sigma^{-2} | \underline{s}^{-2}, \underline{v}) \quad -\infty < \theta < +\infty, \quad \sigma^{-2} > 0 \quad (4)$$

where $(\underline{\beta}, \underline{Q}, \underline{s}^{-2}, \underline{v})$ are hyperparameters, and Jeffrey's prior is

$$\pi(\beta, \sigma^{-2}) \propto \sigma^2. \quad (5)$$

Conditional distributions of parameters: That the augmented posterior distribution $\theta | \{y, z\}$ is a normal-gamma distribution⁶ if either of the above priors is known. For later convenience, denote Z as the augmented data, i.e. $Z = \{y, z\}$. The augmented regression model might be simply written in a matrix form as,

$$Z = X\beta + U$$

where $\beta = (\beta_1, \dots, \beta_{k-p}, \lambda_1, \dots, \lambda_p)'$. Based on a result in Poirier (1995) or Zellner (1976), the conditional distributions of the parameters under the prior (4) are:

$$(a) \beta | \{z, \sigma^{-2}\} \sim N_K(\bar{\beta}, \sigma^2 \bar{Q})$$

where

$$\bar{\beta} = \bar{Q}[\underline{Q}^{-1}\underline{\beta} + (X'X)\hat{\beta}], \quad \hat{\beta} = (X'X)^{-1}X'Z \quad \text{and} \quad \bar{Q} = [\underline{Q}^{-1} + X'X]^{-1}$$

$$(b) \sigma^{-2} | \{z, \beta\} \sim \gamma(\sigma^{-2} | s^{*-2}, \nu^*)$$

where

$$\nu^* = \bar{\nu} + K, \quad s^{*-2} = (\nu^{*-1}[\bar{\nu}\bar{s}^2 + (\beta - \bar{\beta})'\bar{Q}^{-1}(\beta - \bar{\beta})])^{-1} \quad \text{and} \quad \bar{\nu} = \underline{\nu} + T$$

$$\bar{\nu}\bar{s}^2 = \underline{\nu}\underline{s}^2 + (Z - X\bar{\beta})'(Z - X\bar{\beta}) + (\bar{\beta} - \underline{\beta})'\underline{Q}^{-1}(\bar{\beta} - \underline{\beta})$$

Similarly, under prior (5), they are,

⁶ This means that $\beta | \sigma^{-2}$ is normal and σ^{-2} follows a gamma distribution. See Poirier (1995, p. 128) for a formal definition.

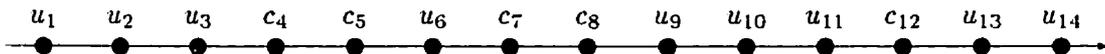
$$(a') \beta|\{z, \sigma^{-2}, \} \sim N_K(\hat{\beta}, \sigma^2(X'X)^{-1})$$

$$(b') \sigma^{-2}|\{z, \beta\} \sim \gamma(\sigma^{-2}|s^{*-2}, \bar{v})$$

where the notation here is the same as those in (a) and (b). The conditional results are a convenient tool for sampling θ from an augmented posterior. In particular, (a) represents a simple multivariate normal distribution, and (b) a gamma distribution. Both of them have a simple form and can be easily sampled.

Conditional distributions of latent data: The derivation of these distributions is the major contribution of the Chapter. The significance of so doing is obvious since the Gibbs sampler with data augmentation algorithm can be thus applied then. To start, it is convenient to have a term to describe a decomposition of the latent data into certain subgroups. Define that a *latent string* is a subset of consecutive observations which begins with a set of p consecutive uncensored observation following immediately a censored observation, and ends after the next set of p consecutive uncensored observations. This concept is similar to, but indeed different from, the censored string defined by Zeger and Brookmeyer (1986). As will be seen later in this section, a latent string consists of a minimum complete probabilistic information **unit** to learn about the censored data in this string. To help understand this concept, the following two examples are provided.

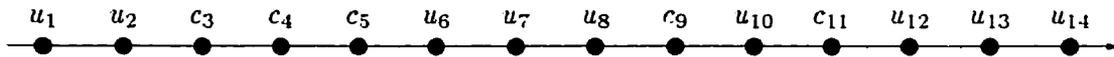
Example 4.1 In the model (1), let $p = 2$, and a sample $\{y_t\}_{t=1}^{14}$ is



where, u_t denotes an uncensored observation, and c_t a censored one. There are two censored strings in this sample: $\{u_2, u_3, c_4, c_5, u_6, c_7, c_8, u_9, u_{10}\}$ and $\{u_{10}, u_{11}, c_{12}, u_{13}, u_{14}\}$. In other words, a latent string consists of two sets of p consecutive uncensored observations as its supporting ends. The number of consecutive uncensored observations between the two sets must be less than p . It is also possible that two censored strings have a common or overlapped

supporting (uncensored) end. However, the censored elements in two distinct latent strings can **never** intersect. This is crucial in deriving conditional distributions of the latent data.

Example 4.2 Suppose that $\{y_t\}_{t=1}^{14}$ is generated from model (1) with $p = 1$. Realized y_t 's are given by



Given the same notation here as those in Example 4.1, three censored strings exist in this sample: $\{u_2, c_3, c_4, c_5, u_6\}$, $\{u_8, c_9, u_{10}\}$ and $\{u_{10}, c_{11}, u_{12}\}$. With $p = 1$, a censored string looks much simpler than that with $p = 2$. None of uncensored observations stands between the two supporting (observed) ends in a latent string. This is not generally true for $p > 1$, however.

Without loss of generality, we assume that the first and last sets of p consecutive observations are uncensored⁷. Unlike in the standard Tobit setting, $z|\theta$ can be quite complicated for the model (1). If a large, or even modest, proportion of observations are censored, it may be reasonable to expect that the conditional density is highly dimensional due to the fact that the (censored and observed) observations are serially correlated. Fortunately, this state of matter can be somewhat simplified. Similar to Zeger and Brookmeyer (1986), the following result permits an information-preserved dimensional reduction of the joint distribution $z|\theta$.

Proposition 4.1 Let $\{y, z\}$ be a sample from the model (1), where y denotes uncensored observations and z censored ones. The joint distribution $z|\theta, y$ can be decomposed into the product of the joint distributions of the latent data in each censored string.

In other words, this decomposition recommends that to sample the latent data, one can

⁷ This assumption is not necessary, but is strictly used to simplify presentation. In fact, if these first p observations are not uncensored, it is supposed that one can find first p uncensored observations somewhere in this series: $\{y_{t_1}^*, \dots, y_{t_p}^*\}$. The procedure discussed in this section requires a minor change: for $t < t_1$, equation (4.9) has to be changed into a forward-looking expression, and equations (4.10)–(4.12) should also be changed accordingly.

rely on their joint distribution in each latent string. Furthermore, it is important to emphasize that this result implies that this decomposition has no information loss. The proof of this Proposition can be analogous to that of Proposition 2.1 in Zeger and Brookmeyer (1986).

Given a latent string, say $\{y_{t-p+1}^{\pi}, \dots, y_t^{\pi}, y_{t+1}^{\pi}, \dots, y_{t+n_t}^{\pi}, y_{t+n_t+1}^{\pi}, \dots, y_{t+n_t+p}^{\pi}\}$, it is useful to derive the conditional distribution: $y_{t+1}^{\pi}, \dots, y_{t+n_t}^{\pi} | \{\theta, y_{t-p+1}^{\pi}, \dots, y_t^{\pi}\}$. This is a multivariate normal distribution. The mean A_t and variance Ω_t of this distribution can be written as follows.

$$A_t = \begin{bmatrix} Q_{t+1} \\ \vdots \\ Q_{t+n_t} \\ \vdots \\ Q_{t+n_t+p} \end{bmatrix} \quad \text{and} \quad \Omega_t = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1 \ n_t+p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2 \ n_t+p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n_t+p \ 1} & \sigma_{n_t+p \ 2} & \dots & \sigma_{n_t+p \ n_t+p} \end{bmatrix} \sigma^2 \quad (6)$$

where the recursive structure of Q_t is

$$Q_{t+\tau} = x'_{t+\tau} \beta + \sum_{v=1}^p \lambda_v Q_{t+\tau-v} \quad \tau = 1, 2, \dots, n_t + p \quad (7)$$

If $\tau - v < 1$, $Q_{t+\tau-v} = y_{t+\tau-v}$, and the recursive structure of σ_{ij} is

$$\begin{aligned} \sigma_{11} &= 1 \\ \sigma_{ii} &= \sum_{l_1=1}^p \sum_{l_2=1}^p \lambda_{l_1} \lambda_{l_2} \sigma_{i-l_1 \ i-l_2} \\ \sigma_{ij} &= \sum_{l=1}^p \lambda_l \sigma_{ij-l} \quad j > i \end{aligned} \quad (8)$$

where to ensure simplicity of exposition, an implicit assumption has been made above: if a subscript is less than 1, then the associated term is zero. The derivations of the above two recursive structures are simply based on the iteration of (1). Furthermore, the order in which the elements of the covariance matrix are computed is crucial:

$$\sigma_{11} - \begin{bmatrix} \sigma_{1j} \\ j=2,3,\dots,n_t+p \end{bmatrix} - \begin{bmatrix} \sigma_{j1} \\ \sigma_{j1}=\sigma_{1j} \end{bmatrix} - \sigma_{22} - \begin{bmatrix} \sigma_{2j} \\ j=3,4,\dots,n_t+p \end{bmatrix} - \begin{bmatrix} \sigma_{j2} \\ \sigma_{j2}=\sigma_{2j} \end{bmatrix} - \dots - \sigma_{n_t+p \ n_t+p}$$

It is also noted that a latent string includes other uncensored observations as well. These uncensored observations, through the correlation structure (6), provide additional information

about the latent data in this string. To fully utilize the information, it requires that one derive the distribution of the latent data by conditioning on the uncensored observations. To do so, rearrange the string $\{y_{t+1}^*, \dots, y_{t+n_t}^*, \dots, y_{t+n_t+p}^*\}$ into $\{\tilde{y}_{t+1}^*, \dots, \tilde{y}_{t+l}^*, \tilde{y}_{t+l+1}^*, \dots, \tilde{y}_{t+n_t+p}^*\}$, where $\{\tilde{y}_{t+1}^*, \dots, \tilde{y}_{t+l}^*\}$ are censored, and $\{\tilde{y}_{t+l+1}^*, \dots, \tilde{y}_{t+n_t+p}^*\}$ are uncensored. Define an one-to-one switching *mapping* $S : \{1, 2, \dots, n_t + p\} \rightarrow \{1, 2, \dots, n_t + p\}$, which shifts all censored observations in the string into the first l positions, and uncensored ones following after. Figure 1 illustrates one such mapping in a hypothetical case.

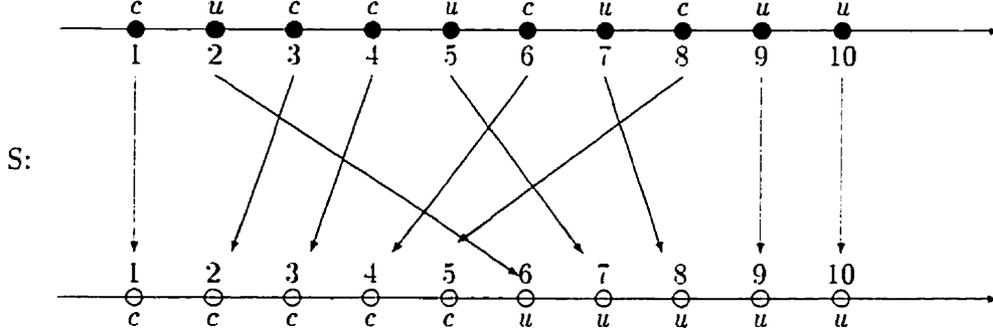


Figure 1: Switching Mapping

Again, c denotes a censored observation and u an uncensored one. The switching mapping S provides a simple tool to reorganize the string so that the formulae used to compute conditional mean and variance of the latent data in this string can be applied directly. Thus, the transformed mean and variance of the string $\{\tilde{y}_{t+1}^*, \dots, \tilde{y}_{t+l}^*, \tilde{y}_{t+l+1}^*, \dots, \tilde{y}_{t+n_t+p}^*\}$ are, respectively,

$$\tilde{A}_t = \begin{bmatrix} \tilde{Q}_{t+1} \\ \vdots \\ \tilde{Q}_{t+l} \\ Q_{t+l+1} \\ \vdots \\ Q_{t+n_t+p} \end{bmatrix} \quad \text{and} \quad \tilde{\Omega}_t = \begin{bmatrix} \tilde{\sigma}_{11} & \cdots & \tilde{\sigma}_{1l} & \tilde{\sigma}_{1,l+1} & \cdots & \tilde{\sigma}_{1n_t+p} \\ \tilde{\sigma}_{21} & \cdots & \tilde{\sigma}_{2l} & \tilde{\sigma}_{2l+1} & \cdots & \tilde{\sigma}_{2n_t+p} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \tilde{\sigma}_{n_t+p1} & \cdots & \tilde{\sigma}_{n_t+pl} & \tilde{\sigma}_{n_t+pl+1} & \cdots & \tilde{\sigma}_{n_t+pn_t+p} \end{bmatrix} \sigma^2 \quad (9)$$

where $\tilde{Q}_{t+i} = Q_{t+S^{-1}(i)}$ and $\tilde{\sigma}_{ij} = \sigma_{S^{-1}(i)S^{-1}(j)}$ and S^{-1} is the inverse of S . The distribution of the latent data $\{\tilde{y}_t^*, \tilde{y}_{t+1}^*, \dots, \tilde{y}_{t+l}^*\}$ conditioning on $\{\tilde{y}_{t+l+1}^*, \tilde{y}_{t+l+2}^*, \dots, \tilde{y}_{t+n_t+p}^*\}$ is again a normal distribution. Suppose the conditional mean and variance are \tilde{A}_t and $\tilde{\Omega}_t$. These

can be computed easily. Then, the following Proposition summarizes and extends the above discussion.

Proposition 4.2 Given a latent string $\{y_{t-p-1}^*, \dots, y_t^*, y_{t+1}^*, \dots, y_{t+n_t}^*, y_{t+n_t+1}^*, \dots, y_{t+n_t+p}^*\}$, the latent data in this string, say $\{\bar{y}_{t+1}^*, \bar{y}_{t+2}^*, \dots, \bar{y}_{t+l}^*\}$ have the following conditional distribution

$$\bar{y}_{t+1}^*, \dots, \bar{y}_{t+l}^* \mid \left\{ \begin{array}{l} \theta, \bar{y}_{t+1} = 0, \dots, \bar{y}_{t+l} = 0. \\ \bar{y}_{t+l+1}, \dots, \bar{y}_{t+n_t+p} \end{array} \right\} \sim \phi(\bar{y}_{t+1}^*, \dots, \bar{y}_{t+l}^* \mid \bar{\Lambda}_t, \bar{\Omega}_t) I_{[0, +\infty)} \quad (10)$$

This is a truncated (multivariate or univariate) normal distribution.

With the complete conditionals available for sampling, the Gibbs sampler with data augmentation algorithm described in (2) may now be run by iterating (a), (b) and (10) in Section 4. The algorithm may be initialized by using the estimates of least squares for the censored data, i.e. treating the censored data as uncensored. Note that this choice is not necessary, but only for speeding up the convergence of the generated Markov chain in the initial states. Based on the resulting posterior draws, $(\beta^{(j)}, (\sigma^{-2})^{(j)})$, $j = m + 1, \dots, m + M$ and for some suitable values of m and M , the posterior mean and standard deviation of θ and σ^2 , as well as z , can be obtained using (3) and (3) since the conditional mean and variance of each are available in closed forms.

Notes on implementation: (1) The normal distribution (a) and the gamma distribution (b) should be easily sampled in any applications. Difficulty, however, may arise when making draws from the truncated multivariate normal distribution (10). The “rejection/acceptance” method could be extremely inefficient if the truncation occurs in a low probability region. In such a case, Geweke’s (1991) method has proven quite useful. The algorithm developed by Geweke depends on an exponential rejection method for tail regions. (2) As a practical issue, instances where the dimension of the latent variables in (10) is large are problematic. Similar to dealing with the model parameters, it is also possible to partition the latent data in this

string into sub-blocks. The estimation procedure proposed above requires little change. It is emphasized that such a partition should proceed cautiously. There can be an efficiency cost resulting from the action. This indicates that sampling sequentially the latent data is not recommended; (3) The recursive structures of (7) and (8), the switching mapping S , and (9) can be all readily coded by adding a few more lines in our computer program. Based on all above, it should be fairly to say that the developed method is indeed practical and easy to implement for a general dynamic Tobit model; (4) According to a recent study by Gelfand, Smith and Lee (1992), it is simple to impose prior linear constraints on model parameters. This idea can be borrowed for dealing with the model discussed here with little change. In the application study of Section 6, we will be back to this issue again.

Remark: (1) The latent data generating procedure (see Proposition 4.2) is somewhat complicated because of the correlation among the censored observations. This situation can be simply understood by comparing it with the standard Tobit model (see Chib (1992)). One way of “simplifying” the likelihood is to use so-called **pseudo-likelihood** as suggested in Zeger and Brookmeyer (1986). The idea behind the treatment is to approximate the true likelihood by not fully taking account of the correlations among observations. While this likelihood simplification is more tractable, it generally leads to an efficiency loss for parameter estimates. On the other hand, the proposed method above is in fact to fully utilize the information about the correlations among observations, as well as the uncertainties of the latent data. Although Bayesian estimation can be also based on the pseudo-likelihood, the method proposed in this Chapter is more efficient in terms of using data information; (2) There is a major difference between the proposed Bayesian method and its classical counterpart. The crucial step of the Bayesian method is to draw variates from a truncated normal distribution (see Proposition 4.2). The key step of classical simulation methods is to evaluate the multivariate normal probabilities and their derivatives in a rectangle (see Hajivassiliou, MacFadden and Ruud (1996)). It is widely known that sampling from a distribution is often computationally much faster than

evaluating its probabilities. Furthermore, classical estimation methods also require evaluation of the derivatives of the probabilities with respect to the model parameters. Therefore our Bayesian method has at least a computational advantage over classical simulation methods.

5 A Monte Carlo Study

A Monte Carlo study is conducted in this section to evaluate practical performance of the proposed method. With the true data generating process (DGP) known, it becomes possible to observe the adequacy of the method in alternative circumstances. The analysis is conducted with vague prior (5). The programs are written in Fortran-77 and run on a Silicon Graphics 4D-320 UNIX machine with intensive use of IMSL Math/Library and Stat/Library.

The regression DGP is designed through

$$\begin{aligned}
 y_t^* &= \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \lambda_1 y_{t-1}^* + \cdots + \lambda_p y_{t-p}^* + \epsilon_t \quad t = 1, 2, \dots, T \\
 y_t &= \max\{y_t^*, 0\} \\
 \epsilon_t &\sim N(0, \sigma^2)
 \end{aligned} \tag{11}$$

where $\beta_1 = -1.0$, $\beta_2 = 1.0$ and $\beta_3 = 10.0$. Analogous to Geweke (1992), the covariates x_{2t} and x_{3t} are generated according to the following

$$\begin{aligned}
 z_{1t}, z_{2t} &\quad \text{i.i.d. uniform (0,1)} \\
 x_{2t} &= 5.0z_{1t} - 1.0 \\
 x_{3t} &= 1.5z_{2t} - C
 \end{aligned}$$

The experiments consist of two examples with sample sizes of $T=50$, 100, and 200.

Example 5.1 Consider model (11) with $p = 1$. In this case, λ_1 is designed to take values .1, .5, 1.0 and 1.2. This strategy is intended to investigate how the proposed method responses to changes of λ_1 . The changes of censoring level are realized by adjusting $C \in [-1, 0]$. Note

that $\lambda_1 = 1.0$ represents a non-stationary latent process (unit root) and $\lambda_1 = 1.2$ represents an explosive latent process.

Example 5.2 Set $p = 2$ in model (11). Design $(\lambda_1, \lambda_2) = (.2, .1), (.7, .2)$ and $(.7, .5)$. Due to the fact that latent strings in this case are more complicated than those in Example 5.1, this design is used to show the implementation and performance of the proposed Bayesian procedure in a complicated framework. Note that $(\lambda_1, \lambda_2) = (.7, .5)$ represents a non-stationary situation.

For each designed model (i.e. fixing the parameters β, λ , sample size T , and the values of x 's), the “**expected**” censoring level is chosen as a measure of unobservability of the underlying model⁸. Such a value, however, is difficult to compute. Instead of a direct approach, we estimate this value (EECL) by averaging the censored ratios of the model over different choices of seeds in coding. The results are reported in all of Tables 1-4.

In the implementation of the proposed procedure, the first 200 draws are discarded and the algorithm is run to 5,000 draws from the posterior. The convergence of each chain is detected by using the method of the visual inspection of CUMSUM statistics as mentioned in Section 3. The results are not reported since it should be easy to see the quality of our parameter estimates, given that the designed values of the parameters are known. As mentioned in Section 3, numerical standard errors are calculated by the method of batching. According to Ripley (1987), the batch size b is selected such that the first-order correlation between the batch means is at most .05. In each Table, the numerical standard error for each parameter estimate is reported in square brackets, and the lag 1 correlation of the Gibbs run is in curly brackets. The time spent on the estimation of each model varies and depends largely upon the models' degree of observability (i.e. their censoring levels) and its sample size. For example, as $EECL = 20\%$ and $T = 50$, 27 seconds are required for the estimation of Example 5.1, and

⁸ Theoretically, the expected censoring level can be computed according to the formula $\frac{1}{T} \sum_{t=1}^T t \times \text{Prob}(\text{number of censored observation} = t | \text{Model})$.

about 1 minute of Example 5.2. If *EECL* increases to 80%, then computing times rise to 20 and 25 minutes, respectively.

The estimation results of Examples 5.1 and 5.2 are reported in Tables 1–4, where the posterior means and standard deviations are calculated using (3). For comparison purposes, three types of estimates are presented in each Table. Line *a* stands for OLS estimates when the latent structure is fully “observed” (this is possible since the data are artificially generated and censored); line *b* OLS estimates with the censored data (i.e., this is designed to demonstrate distortions from OLS procedure to dynamic censored data); and line *c* the proposed Bayesian estimates with the generated censored data.

From these Tables, one can ascertain that the Bayes estimates are very close to the true values that are used to generate the data. Comparing them with the richer information-based estimates of line *a*, we can conclude that our method performs quite well. The batch standard errors show that our estimates are very reliable. On the other hand, OLS estimates with the censored data perform poorly, and get worse with the increase of censoring level (see line *b* in each Table). Part I of Tables 2 and 4 suggests that OLS estimates of the model in both examples, although maybe consistent, are much inefficient for finite samples. Further, the estimated σ^2 in row *b* is always associated with a much larger value than its designed one. Intuitively, this means that a linear (dynamic) model cannot fit (dynamic) censored data in general. We can thus conclude that OLS estimation is indeed inappropriate for dynamic Tobit models. OLS estimates, however, can be used to initialize our sampling algorithm, although this is not necessary. In addition, it is seen from Tables 1 and 3 that the lower a model’s censoring level is, the better its estimates are. This “censoring level” effect is intuitive because an increase in the “expected” censoring level results in less information available in the data. Next, the estimated results regarding changes of λ_1 and λ_2 are reported in Part II of Tables 1 and 3 for both examples. With $EECL = 50\%$, the Bayesian procedure works very well, simply by comparing our Bayesian estimates with the OLS estimates in line *a*.

Part I of Tables 2 and 4 suggests that our Bayesian estimates are consistent. When λ_1 , λ_2 and censoring level are all high, our Bayesian method still performs fairly well (see Part II of Tables 2 and 4), but at the cost of a large increase in computing time. Table 1 (II) and Table 3 (II) show that a nonstationary process of y_t^* would not be a problem with the proposed procedure.

6 An Application

One real-world example is given here to illustrate what kind of economic applications this Bayesian procedure allows. The data were also studied by Zangari and Tsurumi (1994). They examine a demand model of the Japanese exports of passenger cars to the U.S. using annual and quarterly data from 1974 to 1992. The U.S. automakers suffered huge losses in the end of 1970s because of the gasoline crisis (see Tsurumi and Tsurumi (1983)). In the meantime, Japanese passenger car exports to the U.S. increased dramatically. As a compromise, the United States negotiated with Japan Voluntary Export Restraints (VERs) to curb Japanese car imports. The VERs took effect in 1981 with 1.68 million units per year. It was raised to 1.85 in April 1984, further to 2.3 in April 1985, and then lowered to 1.65 in April 1992. The data indicate that VERs are binding during 1981 and 1986.

The question addressed here is how to model the demand of the U.S. for Japanese cars when the observations are constrained by the quotas? First of all, it is necessary to take this VER (censored) effect into account since otherwise the estimates would be distorted, as discussed before. Next, the evidence from the data shows that the demand is autocorrelated. Therefore, the data possess both censored and dynamic features, which makes the dynamic Tobit model the appropriate one to use. To simplify the analysis, assume that there is no effect of the VERs on the demand decision-making. This implies that VERs do not change the desired demand process.

To capture the above features of the demand function, the following model specification is introduced

$$\begin{aligned}
y_t^* &= \beta_1 + \beta_2 x_{2,t} + \beta_3 x_{3,t} + \lambda_1 y_{t-1}^* + \dots + \lambda_p y_{t-p}^* + \epsilon_t \\
y_t &= \min\{y_t^*, L_t\} \\
\epsilon_t &\sim i.i.d. N(0, \sigma^2)
\end{aligned} \tag{12}$$

where: y_t^* is the logarithm of **desired** per capita demand for Japanese cars; y_t is the logarithm of **observed** per capita demand for Japanese cars; $x_{2,t}$ is the logarithm of per capita real disposable income; $x_{3,t}$ is the logarithm of price ratio of Japanese car to domestic car; L_t is the logarithm of VER at time t .

This model is a modified demand regression function of Tsurumi and Tsurumi (1983) and Zangari and Tsurumi (1994). Briefly, the economic justification of the independent variables chosen is as follows. Real but nominal per capita disposable income is chosen because the purchases of cars are treated as investment in economics. A positive coefficient for this variable is thus expected. Following Zangari and Tsurumi (1994), relative price, i.e. price ratio of the Japanese car to the domestic car is chosen as a forcing variable to capture price and cross-price effects. A negative coefficient for this variable is therefore expected. To model the dynamic effect, it is important to ask whether y_{t-1}^* or y_{t-1} has effect on the process of y_t^* . From consumer theory, the demand function is derived from utility-maximization subject to a budget constraint. Certainly, VERs do not affect an individual's budget constraint. Does an individual's preference for Japanese cars depend on VERs? Apparently not. This allows one to assume that VERs do not affect an individual's utility function. Consequently, it should be clear that the lagged observed demand of y_t is free to explain the desired demand y_t^* because they depend on VERs. On the other hand, the lagged demand of y_t^* is independent of VERs. Hence it is reasonably believed that they are better candidates than the lagged y_t 's as an explanatory variable. From an economic point of view, the lagged y_t^* 's reflect the "inertial"

effect of an individual's desired consumption. Thus positive coefficients for the lagged variables are expected.

The selection of the lagged order is simply conducted according to the AIC (Akaike Information Criterion)⁹ as defined to minimize, with respect to p .

$$AIC(k) = \ln \hat{\sigma}_k^2 + \frac{2k}{T}$$

where k is the number of parameters, T is the sample size, and $\hat{\sigma}_k^2$ is the estimated regression error (see Judge, et. al., p. 244). Note that if the t th observation is latent, the regression error for the observation is computed as the average of the errors by using the simulated "observations" of the dependent variable. The calculated results overwhelmingly support $p = 1$. For an illustration, the estimation results regarding both one and two period lags are reported.

Based on quarterly data from 1974 to 1992, our estimated results are reported in Table 3. The convergence check of the MCMCs are the plot of CUMSUM statistics against t (see Figure 2). The basic message from the plot is that all of the chains converge after a short run of initial draws.

With a 30% censoring level, computation of the posterior means and standard deviations of the parameters uses up approximately 10 minutes. The discarded initial draws are $m = 200$ and the returned posterior draws are $M = 5000$. This example is somewhat special since there is only one long latent string. Zangari and Tsurumi (1994) failed to obtain their results from the Gibbs sampler with data augmentation by claiming that the long latent string and the large AR(1) coefficient in their error term contaminate the iterative procedure of the Gibbs sampler. Our simulation in Section 4 produces fairly good results even with long latent strings and large value(s) of λ_1 (and λ_2). The computation performed here does not suffer from the problems encountered by Zangari and Tsurumi.

⁹Other methods such as SC and FPE in Judge, et. al., pp. 241 - 247 will consistently derive similar results for the current problem.

The estimated results are reported in Table 5. Case I in that Table refers to the model (12) with $p = 1$. Case II refers to the model (12) with $p = 2$. The results are basically consistent with our expectations except for the coefficient of the price ratio variable. The large standard deviation of this estimate indicates that the data information cannot narrow down enough our belief on this parameter. Using a truncated prior for β_3 , i.e. $\beta_3 < 0$ can improve the results. This is done by simply following the notes on implementation (4) in Section 4. These results are reported in line d of Table 5. For both cases, this constraint from economic theory improves the parameter estimates in terms of efficiency. The relatively (very) low value of estimated λ_2 in case II shows a very fast decay of the aggregate consumption memory. A simple comparison of the OLS estimates (line b) with our Bayesian estimates (line c) indicates that the OLS method over-estimates the real income effect and under-estimates the desired demand “inertial” effect. The marginal prior (solid curves) and posterior (dotted curves) densities of all parameters for case II are shown in Figures 3.

7 Conclusion

This Chapter has developed a simulation-based Bayesian method for the estimation of a dynamic Tobit model. Due to the intractable likelihood function of such a model, traditional Bayesian and classical estimation methods are not applicable in this situation. In particular, this is because the high dimensional integrals, driven by unobserved data, in the likelihood function, make it extremely difficult, if not impossible, to directly evaluate both the likelihood and the posterior. The solution to the problem proposed in this Chapter is to develop a sampling scheme for the conditional posterior distribution of the latent data so that the Gibbs sampler with data augmentation algorithm is successfully applied. The concept of **latent string** plays a role in this analysis since it provides an easy way to learn about the “unobserved data”. The advantages of this approach are: (1) it provides a unifying approach to both stationary and non-stationary latent processes; (2) it has proven attractive from both

theoretical and practical viewpoints; (3) both informative and non-informative prior beliefs can be easily incorporated into the process of making estimation inferences; (4) a Monte Carlo experiment shows that this method performs satisfactorily in various circumstances, for example, with small sample sizes and/or nonstationarity; and (5) the method can be easily extended to handle a dynamic Tobit model with Student-t errors. The proposed procedure is applied to a regression study of Japanese exports of passenger cars to the US subject to a non-tariff trade barrier. This application shows: (1) a dynamic Tobit model is indeed appropriate in such a situation; and (2) how this model is applied in a real-world problem.

Note on Tables 1 - 4

(1) EECL represents Estimated Expected Censoring Level.

(2) The numbers just below parameters are their designed values. Line *a* refers to OLS estimates when the latent structure is “fully observed”; line *b*, OLS estimates with the designed censored data; and line *c*, estimates of the proposed method. Standard deviation is in bracket. The numerical standard error of the posterior mean is in square bracket; the lag 1 correlation of the Gibbs run is in curly bracket. The estimates are based on $m=200$ and $M=5000$.

Note on Table 5

Line *b* refers to OLS estimates with the censored data; and line *c*, estimates of the proposed method; line *d*, estimates of the proposed method with the parameter constraint $\beta_3 < 0$. Standard deviation is in bracket. The numerical standard error of the posterior mean is in square bracket; the lag 1 correlation of the Gibbs run is in curly bracket. The estimates are based on $m=200$ and $M=5000$.

Table 1: Simulation Results
(cf. Example 5.1)

I. Changes of EECL (Sample Size = 50)

Parameter —		β_1	β_2	β_3	λ_1	σ^2
EECL ↓		-1.0	1.0	10.0	0.5	1.0
20%	a	-1.25 (.20)	1.03 (.02)	9.95 (.23)	.49 (.02)	1.01 (.21)
	b	-.23 (.64)	.83 (.07)	6.48 (.69)	.50 (.05)	8.79 (.23)
	c	-1.10 (.20)	1.00 (.03)	10.15 (.34)	.49 (.01)	.87 (.23)
		[.003]	[.0004]	[.007]	[.0002]	[.003]
		{-.25}	{-.15}	{-.07}	{-.22}	{-.01}
50%	a	-1.30 (.20)	1.03 (.02)	9.95 (.23)	.49 (.02)	1.01 (.20)
	b	.30 (.67)	.69 (.07)	4.41 (.75)	4.41 (.75)	.50 (.07)
	c	-1.05 (.24)	1.00 (.03)	9.98 (.48)	.49 (.02)	.93 (.29)
		[.005]	[.0008]	[.014]	[.0004]	[.005]
		{.16}	{.18}	{.01}	{.04}	{.08}
80%	a	-1.50 (.30)	1.03 (.02)	9.95 (.23)	.48 (.02)	.99 (.21)
	b	-.40 (.62)	.29 (.06)	1.37 (.56)	.45 (.11)	5.85 (1.22)
	c	-.48 (.58)	.98 (.06)	10.90 (1.04)	.47 (.03)	1.04 (.38)
		[.018]	[.004]	[.079]	[.001]	[.008]
		{-.07}	{.29}	{.27}	{.25}	{.01}

II. Changes of λ_1 (Sample Size = 50 ; EECL =50%)

Parameter —		β_1	β_2	β_3	λ_1	σ^2
λ_1 ↓		-1.0	1.0	10.0	1.0	1.0
.1	a	-1.32 (.20)	1.03 (.02)	9.95 (.24)	.09 (.02)	1.01 (.20)
	b	1.52 (.54)	.53 (.06)	3.84 (.59)	.25 (.08)	6.42 (1.34)
	c	-1.07 (.36)	1.01 (.04)	10.18 (.58)	.09 (.02)	.98 (.29)
		[.013]	[.002]	[.025]	[.0005]	[.004]
		{.15}	{.13}	{.16}	{.02}	{-.03}
1.0	a	-1.34 (.20)	1.03 (.02)	9.86 (.24)	.99 (.01)	.96 (.20)
	b	-.49 (.78)	.57 (.09)	3.90 (.90)	.93 (.05)	13.86 (2.89)
	c	-1.20 (.23)	.97 (.03)	9.39 (.38)	.99 (.01)	.96 (.29)
		[.0047]	[.0007]	[.0097]	[.0001]	[.0047]
		{-.15}	{.33}	{. - .22}	{.05}	{-.25}
1.2	a	-.92 (.15)	1.00 (.02)	10.00 (.001)	1.19 (.01)	.70 (.15)
	b	-.09 (1.02)	.49 (.11)	10.00 (.01)	1.30 (.11)	21.90 (4.57)
	c	-.95 (.20)	1.00 (.03)	10.00 (.001)	1.19 (.01)	.90 (.29)
		[.0027]	[.0007]	[.0000]	[.0003]	[.0053]
		{-.06}	{.12}	{-.05}	{.11}	{-.13}

Table 2: Simulation Results
(cf. Example 5.1)

I. Changes of Sample Size (EECL = 50%)

Parameter —		β_1	β_2	β_3	λ_1	σ^2
Sample Size ↓		-1.0	1.0	10.0	.5	1.0
100	a	-1.15 (.13)	1.01 (.01)	9.86 (.16)	.48 (.01)	.82 (.12)
	b	.65 (.48)	.63 (.05)	4.80 (.53)	.51 (.04)	8.92 (1.29)
	c	-.96 (.19)	1.00 (.02)	9.81 (.31)	.48 (.01)	.91 (.20)
		[.003]	[.0004]	[.007]	[.0002]	[.004]
	{-.11}	{-.11}	{.19}	{-.27}	{-.05}	
200	a	-1.03 (.11)	1.00 (.01)	10.05 (.12)	.50 (.01)	.96 (.10)
	b	1.27 (.37)	.55 (.03)	5.02 (.37)	.45 (.03)	8.78 (.89)
	c	-.85 (.15)	.99 (.02)	10.16 (.21)	.48 (.01)	.99 (.13)
		[.003]	[.0004]	[.005]	[.0002]	[.002]
	{-.03}	{.10}	{.02}	{.18}	{-.20}	

II. Large λ and EECL (Sample Size = 100)

Parameter —		β_1	β_2	β_3	λ_1	σ^2
EECL ↓		-1.00	1.00	10.00	.90	1.00
80%	a	-1.02 (.15)	1.00 (.01)	10.02 (.12)	.90 (.004)	.96 (.10)
	b	.66 (.30)	.13 (.03)	1.42 (.27)	.68 (.05)	5.02 (.51)
	c	-.72 (.22)	.93 (.03)	9.57 (.23)	.90 (.005)	.90 (.17)
		[.006]	[.002]	[.021]	[.0001]	[.003]
	{.23}	{.15}	{.13}	{.06}	{-.02}	

Table 3: Simulation Results
(cf. Example 5.2)

I. Changes of EECL (Sample Size = 50)

Parameter	—	β_1	β_2	β_3	λ_1	λ_2	σ^2
EECL ↓		-1.0	1.0	10.0	0.5	0.2	1.0
20%	a	-1.21 (.22)	1.03 (.02)	9.96 (.24)	.49 (.02)	.20 (.02)	1.03 (.22)
	b	-.10 (.65)	.83 (.06)	7.34 (.65)	.50 (.06)	.20 (.06)	7.75 (1.63)
	c	-1.07 (.20)	1.01 (.02)	9.94 (.26)	.50 (.02)	.19 (.02)	.78 (.20)
		[.003]	[.0004]	[.004]	[.0002]	[.0002]	[.004]
		{-.22}	{-.20}	{.21}	{.08}	{-.12}	{-.08}
50%	a	-1.32 (.21)	1.03 (.03)	10.00 (.25)	.51 (.02)	.18 (.02)	1.15 (.24)
	b	.14 (.75)	.64 (.08)	3.80 (.79)	.49 (.09)	.14 (.09)	11.62 (2.45)
	c	-1.00 (.24)	.98 (.03)	9.80 (.43)	.52 (.03)	.17 (.02)	.97 (.32)
		[.005]	[.001]	[.013]	[.0005]	[.0004]	[.007]
		{.19}	{.05}	{-.01}	{.12}	{-.13}	{-.01}
80%	a	-1.49 (.25)	1.03 (.02)	9.97 (.23)	.49 (.02)	.19 (.02)	.99 (.21)
	b	0.50 (.66)	.37 (.06)	2.00 (.66)	.43 (.12)	-.04 (.12)	7.91 (1.67)
	c	-0.91 (.28)	.93 (.03)	9.38 (.40)	.49 (.03)	.17 (.03)	.51 (.27)
		[.007]	[.001]	[.018]	[.001]	[.009]	[.019]
		{.36}	{.16}	{-.1}	{.11}	{.26}	{.18}

II. Changes of (λ_1, λ_2) (Sample Size = 50); EECL=50%

Parameter	—	β_1	β_2	β_3	λ_1	λ_2	σ^2
(λ_1, λ_2) ↓		-1.0	1.0	10.0			1.0
(.5, .2)	a	-1.31 (.20)	1.03 (.02)	9.96 (.24)	.49 (.02)	.19 (.02)	1.01 (.21)
	b	.23 (.76)	.67 (.08)	4.06 (.81)	.47 (.09)	.15 (.09)	12.34 (2.60)
	c	-1.13 (.22)	1.02 (.03)	10.05 (.37)	.48 (.02)	.19 (.02)	.78 (.25)
		[.004]	[.0005]	[.009]	[.0004]	[.0004]	[.005]
		{-.18}	{-.17}	{-.18}	{-.21}	{-.27}	{.11}
(.7, .2)	a	-1.35 (.20)	1.03 (.02)	9.92 (.23)	.69 (.02)	.20 (.02)	.98 (.21)
	b	-.15 (.76)	.60 (.08)	3.57 (.84)	.69 (.10)	.11 (.10)	12.76 (2.69)
	c	-1.22 (.20)	1.00 (.03)	9.55 (.34)	.66 (.02)	.22 (.02)	.75 (.25)
		[.003]	[.0006]	[.008]	[.0004]	[.0004]	[.005]
		{-.16}	{.08}	{.11}	{.14}	{.09}	{-.28}
(.7, .4)	a	-1.34 (.19)	1.04 (.02)	9.87 (.14)	.68 (.01)	.40 (.01)	.96 (.20)
	b	-.26 (.99)	.59 (.11)	8.86 (.69)	.75 (.06)	.41 (.10)	21.73 (4.58)
	c	-1.23 (.20)	1.01 (.04)	9.89 (.16)	.68 (.01)	.41 (.02)	.93 (.33)
		[.004]	[.0009]	[.0025]	[.0002]	[.0003]	[.0064]
		{-.05}	{-.16}	{-.01}	{.10}	{-.12}	{-.04}

Table 4: Simulation Results
(cf. Example 5.2)

I. Changes of Sample Size (EECL = 50%; $(\lambda_1, \lambda_2) = (.5, .2)$)

Parameters	—	β_1	β_2	β_3	λ_1	λ_2	σ^2
Sample Size ↓		-1.0	1.0	10.0	0.5	0.2	1.0
100	a	-1.17 (.13)	1.01 (.01)	9.86 (.16)	.48 (.01)	.20 (.01)	.83 (.12)
	b	.52 (.55)	.62 (.05)	4.97 (.58)	.46 (.06)	.17 (.06)	11.27 (1.63)
	c	-1.07 (.16)	1.01 (.02)	9.86 (.23)	.47 (.01)	.21 (.01)	.75 (.15)
		[.003]	[.0004]	[.006]	[.0002]	[.0002]	[.003]
		{-.34}	{-.35}	{-.25}	{-.12}	{-.005}	{.104}
200	a	-1.03 (.11)	1.00 (.01)	10.05 (.12)	.49 (.01)	.21 (.01)	.96 (.10)
	b	0.80 (.42)	.53 (.04)	4.76 (.41)	.44 (.04)	.18 (.04)	10.75 (1.08)
	c	-0.98 (.14)	.99 (.02)	10.01 (.19)	.50 (.01)	.20 (.01)	.91 (.12)
		[.002]	[.0003]	[.005]	[.0002]	[.0002]	[.002]
		{-.02}	{-.02}	{.29}	{-.21}	{-.21}	{.02}

II. Large (λ_1, λ_2) and Large EECL (EECL=80%; Sample Size = 100)

Parameter	—	β_1	β_2	β_3	λ_1	λ_2	σ^2
$(\lambda_1, \lambda_2) ↓$		-1.0	1.0	10.0			1.0
(.7, .2)	a	-1.22 (.14)	1.01 (.01)	9.83 (.15)	.68 (.01)	.21 (.01)	.82 (.12)
	b	.22 (.58)	.58 (.06)	4.58 (.60)	.61 (.06)	.17 (.06)	12.74 (1.84)
	c	-1.11 (.14)	1.00 (.02)	9.70 (.21)	.65 (.01)	.23 (.01)	.62 (.12)
		[.002]	[.0004]	[.005]	[.0002]	[.0002]	[.002]
		{-.10}	{-.32}	{-.13}	{-.03}	{-.27}	{.15}

Table 5: VER Trade Example

Case I: Demand for Japanese Cars in the U.S.
 (Sample Size 75 and Censoring Level 27%)

Parameter --	β_1	β_2	β_3	λ_1	σ^2
b	-9.55 (4.14)	.86 (.42)	.04 (.35)	.74 (.08)	.020 (.003)
c	-5.82 (4.32)	.48 (.45)	.20 (.42)	.77 (.09)	.023 (.004)
	[.070]	[.007]	[.006]	[.002]	[.0006]
	{-.023}	{-.26}	{-.28}	{-.06}	{-.10}
d	-6.19 (.89)	.52 (.10)	-.09 (.07)	.77 (.02)	.022 (.004)
	[.024]	[.002]	[.004]	[.0004]	[.0002]
	{.50}	{.42}	{-.69}	{.24}	{-.22}

Case II: Demand for Japanese Cars in the U.S.
 (Sample Size 75 and Censoring Level 27%)

Parameter --	β_1	β_2	β_3	λ_1	λ_2	σ^2
b	-10.00 (4.15)	.89 (.42)	.02 (.35)	.67 (.12)	.04 (.12)	.020 (.003)
c	-5.40 (4.32)	.43 (.44)	.18 (.41)	.66 (.13)	.10 (.12)	.02 (.004)
	[.066]	[.007]	[.006]	[.002]	[.002]	[.0006]
	{.02}	{.005}	{-.08}	{-.07}	{-.007}	{-.07}
d	-5.91 (1.04)	.49 (.10)	-.10 (.08)	.66 (.04)	.10 (.12)	.021 (.004)
	[.040]	[.004]	[.003]	[.001]	[.001]	[.0001]
	{-.05}	{-.04}	{.11}	{.04}	{-.07}	{-.01}

Figure 1: CUMSUM Plots

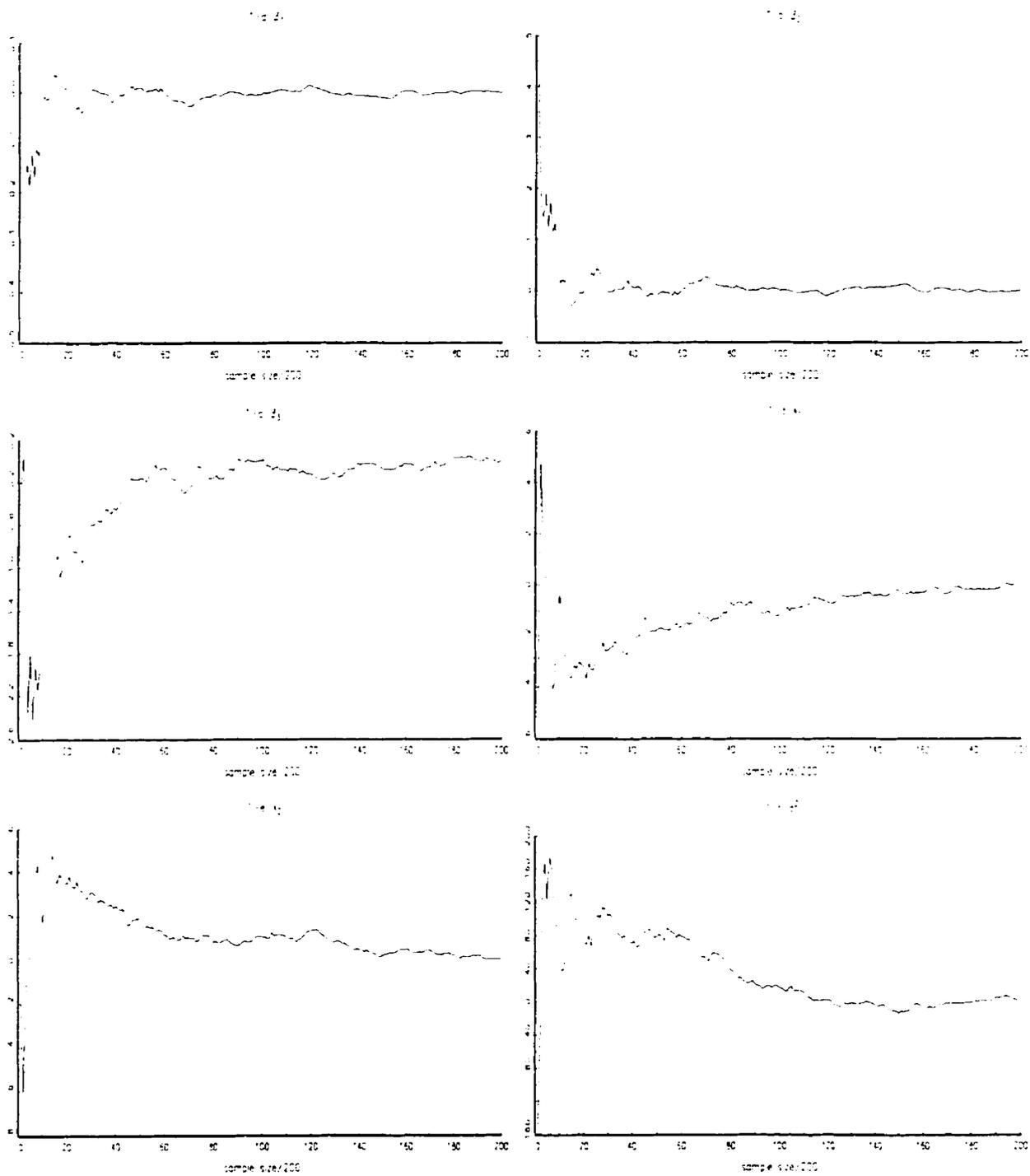


Figure 2

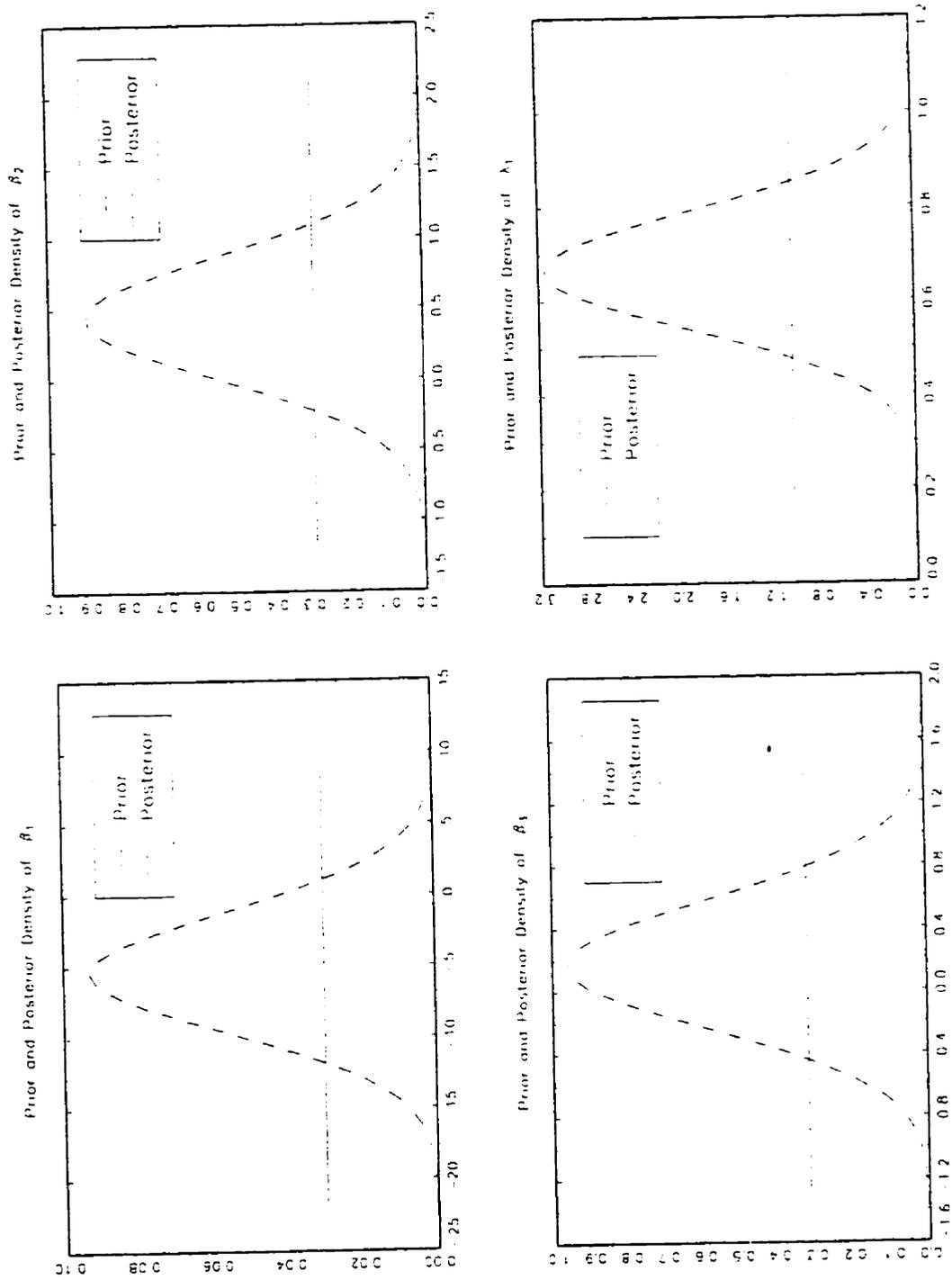
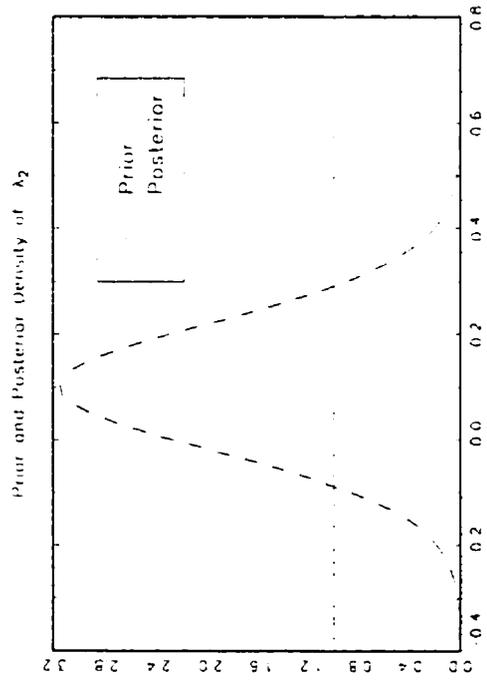
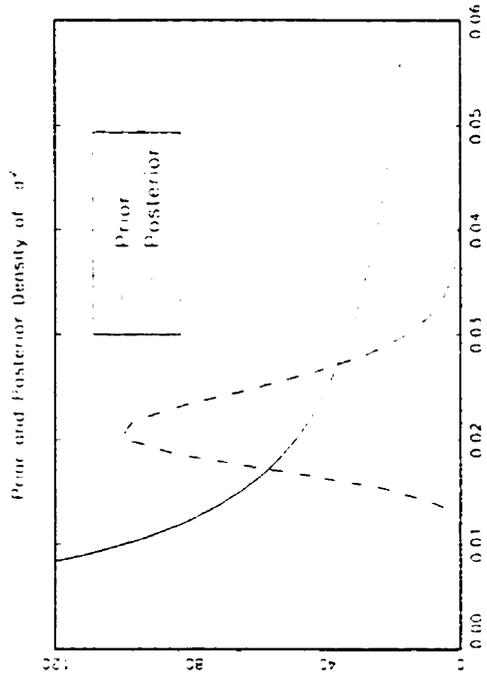


Figure 3



References

- [1] Albert, J.H. and Chib, S. (1993): "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association* **88**, 657-667.
- [2] Amemiya, T. (1984): "Tobit Models: A Survey." *Journal of Econometrics* **24**, 3-61.
- [3] Amemiya, T. (1985): *Advanced Econometrics*. Cambridge, MA: Harvard University Press.
- [4] Bauwens, L. and Lubrano, M. (1996): "Bayesian Inference on GARCH Models Using the Gibbs Sampler." *Working Paper*, Center for Operational Research & Econometrics, Université Catholique de Louvain.
- [5] Casella, G. and George, E.I. (1992): "Explaining the Gibbs Sampler." *The American Statistician* **46**, 167-174.
- [6] Chib, S. (1992): "Bayes Inference in the Tobit Censored Regression Model." *Journal of Econometrics* **51**, 79-99.
- [7] Chib, S. (1994): "Marginal Likelihood from the Gibbs Output." *Journal of the American Statistical Association* **90**, 1313-1321.
- [8] Calzolari, G. and Fiorentini, G. (1996): "A Tobit Model with GARCH errors." *Working Paper*, Università di Firenze.
- [9] Dagenais, M.G. (1982): "The Tobit Model with Serial Correlation." *Econometrics Letters* **10**, 263-267.
- [10] Geman, S. and Geman, D.J. (1984): "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721-741.

- [11] Gelfand, A.E. (1991): "Gibbs Sampling for Marginal Posterior Expectations." *Communications in Statistics B* **20**, 1747-1766.
- [12] Gelfand, A.E. and Smith, A.F.M. (1990): "Sampling Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association* **85**, 398-409.
- [13] Gelfand, A.E., Smith, A.F.M. and Lee, T. (1992): "Bayesian Analysis of Constrained Parameter and Truncated Data Problems Using Gibbs Sampling," *Journal of the American Statistical Association* **87**, 523-532.
- [14] Geweke, J. (1989): "Bayesian Inference in Econometrics Models Using Monte Carlo Integration," *Econometrica* **57**, 1317-1339.
- [15] Geweke, J. (1991): "Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints," in E.M. Keramidas (ed.), *Computing Science and Statistics: Proceedings of the Twenty-Third Symposium of the Interface*, 571-578.
- [16] Geweke, J. (1992): "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," in J.M. Bernardo, J.O. Berger, A.P. David and A.F.M. Smith, eds *Bayesian Statistics 4* (Oxford University Press), 169-193.
- [17] Geweke, J. (1993): "Bayesian Treatment of the Independent Student-t Linear Model." *Journal of Applied Econometrics*, **8**, 19-40.
- [18] Gouriierroux, C., Monfort, A. and Trognon, A. (1985): "A General Approach to Serial Correlation," *Econometrics Theory* **1**, 315-340.
- [19] Greene, W.H. (1992): *Econometrics Analysis*. New York: Macmillan.
- [20] Grether, D.M., and Maddala, G.S. (1982): "A Time Series Model with Qualitative Variables," in M. Diestler, E. Furst, and G. Schwodiauer eds, *Games, Economic Dynamics and Time Series Analysis*, Vienna-Wurzburg: Physica Verlag 291-305.

- [21] Hajivassiliou, V., McFadden, D. and Ruud, P. (1996): "Simulation of Multivariate Normal Rectangle Probabilities and Their Derivatives: Theoretical and Computational Results," *Journal of Econometrics* **72**, 85-143.
- [22] Judge, G.G., Griffiths, W.E., Hill, C., Lutkepohl, H. and Lee, T. (1985): *The Theory and Practice of Econometrics*, second edition, New York: Wiley.
- [23] Kodres, L.E. (1988): "Tests of Unbiasedness in Foreign Exchange Future Markets: The Effects of Price Limits," *Review of Futures Markets*, **7**, 138-166.
- [24] Kodres, L.E. (1993): "Tests of Unbiasedness in the Foreign Exchange Futures Markets: An Examination of Price Limits and Conditional Heteroscedasticity," *Journal of Business* **66**, No. 3, 463-490.
- [25] Maddala, G.S. (1987): *Limited Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press.
- [26] Morgan, I.G. and Trevor, R.G. (1996): "Limit Moves as Censored Observations of Equilibrium Futures Price in GARCH Processes," *Working Paper*, School of Business, Queen's University.
- [27] Peristiani, S. (1994): "An Empirical Investigation of the Determinants of Discount Window Borrowing: A Disaggregate Analysis", *Journal of Banking and Finance* **18**, 183-197.
- [28] Poirier, D.J. (1988): "Frequentist and Subjectist Perspectives on the Problems of Model Building in Economics (with discussion)," *Journal of Economic Perspective* **2** (Winter), 121-170.
- [29] Poirier, D.J. and Ruud P.A. (1988): "Probit with Dependent Observations." *Review of Economic Studies* **LV**, 593-614.
- [30] Poirier, D.J. (1995): *Intermediate Statistics and Econometrics: A Comparative Approach*. Cambridge: MIT Press.

- [31] Ripley, R.D. (1987): *Stochastic Simulation*. New York: Wiley.
- [32] Robert, C.P. (1995): "A Pathological MCMC Algorithm and Its Use as a Benchmark for Convergence Assessment Techniques." *Document de Travail CREST 9555*. INSEE. Paris.
- [33] Robinson, P.M. (1982a): "Analysis of Time Series from Mixed Distributions." *The Annals of Statistics* **10**, 915-925.
- [34] Robinson, P.M. (1982b): "On the Asymptotic properties of Estimators of Models Containing Limited Dependent Variables." *Econometrica* **50**, 27-41.
- [35] Smith, A.F.M. and Gelfand, A.E. (1992): "Bayesian Statistics without Tears: A Sampling- Resampling Perspective." *The American Statistician* **46**, 84-88.
- [36] Sweeting, T.J. (1987): "Approximate Bayesian Analysis of Censored Survival Data." *Biometrika* **74**, 809-916.
- [37] Tanner, M.A. (1993): *Tools for Statistical Inference*. New York: Springer-Verlag.
- [38] Tanner, M.A. and Wong, W.-H. (1987): "The Calculation of Posterior Distributions by Data Augmentation." *Journal of the American Statistical Association* **82**, 528-550.
- [39] Tierney, L. (1994): "Markov Chains for exploring posterior distributions." *Annals of Statistics* **22**, 1701-1762.
- [40] Tobin, J. (1958): "Estimation of Relationship for Limited Dependent Variables." *Econometrica* **26**, 24-36.
- [41] Tsurumi, H. and Tsurumi, Y. (1983): "US-Japan automobile trade: a Bayesian test of a product life cycle," *Journal of Econometrics* **23** , 193-210.
- [42] Yu, B. and Mykland, P. (1994): "Looking at Markov Samplers through Cusum Path Plots: A Simple Diagnostic Idea," *Technical Report 413*, Department of Statistics, University of California at Berkeley.

- [43] Zangari, P.J. and Tsurumi, H. (1994): "A Bayesian Analysis of Censored Autocorrelated Data on Exports of Japanese Passenger Cars to the U.S.." *Advances in Econometrics*. forthcoming.
- [44] Zeger, S.L. and Brookmeyer, R. (1986): "Regression Analysis with Censored Autocorrelated Data." *Journal of the American Statistical Association* **81**, 722-729.
- [45] Zellner, A. (1971): *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley.

**An Empirical Analysis of the Fed's
Open Market Operations**

1 Introduction

The practice of monetary policy involves two closely-linked decision processes: the process of policy formulation and the process of policy implementation. In the United States, the Federal Reserve Open Market Committee (the FOMC) is the policy-making body of the Federal Reserve System (the Fed) and the Federal Reserve's domestic trading Desk (the Desk) is the major implementing arm of the FOMC. The monetary policy strategy delivered by the FOMC is thus largely carried on the Desk's actions. The behavior of the Desk's decisions has thereby attracted much attention to both economists and market practitioners. Although some descriptive analyses about the Desk's behavior have been discussed in the literature, such as Mulendyke (1989) and Lombra (1993), an empirical investigation of the Desk's open market reactions has not been taken in the previous studies except Feinman (1993). As a potential extension of Feinman's work, this Chapter offers an exact posterior perspective on the Desk's open market operations.

This study is important for several reasons. *Firstly*, the Desk's decision behavior which attracts the interests of both theoretical economists and market practitioners can be summarized in a posterior distribution of model parameters. The uncertainty associated with model parameters and model specifications can thus be fully taken account of in a Bayesian framework. The importance of doing so is to help understand the Desk's daily behavior. *Secondly*, the developed econometric framework provides empirical evidence on whether there was a change in the implementation of monetary policy after the stock market crash in October, 1987. This issue has been debated in the literature for years. This is understandable because a change of the Desk's reaction pattern pertains to a change in the implementation of monetary policy, and vice versa. *Thirdly*, predictability of the proposed model can shed light on some useful suggestions on how to improve the implementation of monetary policy. *Finally*, the attempt to analyzing the economic problem results in an innovation of developing a dynamic

friction model and raises a series of interesting inferential questions.

The recent paper by Feinman (1993), to the knowledge of the author, is the only study in the literature using the daily data of the Desk's open market operations. The spike at zero, a very peculiar characteristic of the data distribution (see Figure 1), is the major focus in Feinman's paper. It is quite clear that any continuous distribution cannot fit the data. Smartly, he found that Rosett's (1959) friction model — an extension of the Tobit model (Tobin (1958)) — could be borrowed to fit the spike. This friction structure (*i.e.*, the spike) can be justified by an **operational cost**, but not discussed by Feinman, that will be explored in this study. The similarity between the open market operations and portfolio management provides an insight. In fact, the Desk's interventions are pertaining to managing the Fed's portfolio account, though this management is much more complicated than managing a simple portfolio. Both of these two management behaviors generate spikes at zero in their data distributions. In the management of a portfolio, this spike is due to a transaction cost (see Rosett (1959)). A fund manager needs to balance his/her portfolio adjustment and the associated transaction cost. The word "friction" was thus heuristically used to name a special censored model — a friction model — to describe the behavior of a portfolio adjustment in Rosett's paper. In the Desk's operations, the decision on whether to intervene the funds' market depends apparently on something else besides a transaction cost. For example, the Desk would put the FOMC's policy strategy on its top priority. All of the factors that generate the spike for the Desk's operational data will be discussed, and I call the overall effect of them an operational cost. It is thus the operational cost, rather than a transaction cost per se, that underlies the observed spike for the Desk's operational data.

Indeed, introduction of a friction model can be promising in analyzing the Desk's behavior. This, however, implies by no means that the simple friction model of Rosett (1959) is flexible enough to model the Desk's daily intervention behavior. As will be seen in this Chapter, fitting the spike relies also on other characteristics of the data. For example, ignoring the fat

tails of the data, possibly due to *unobserved heterogeneity*, can result in a poor fit of the spike, even if Rosett's (1959) friction structure is adopted. This leads to considering a fat-tailed data distribution, rather than the widely-used normal distribution as in Rosett (1959) and Feinman (1993). Lack of the examination of other data characteristics is a major drawback of Feinman's analysis, consequently yielding a poor fit of the spike (see Figure 3). The link between fat tails and unobserved heterogeneity can be explained approximately and statistically. Especially in a Bayesian world, this is because a Student-t (fat-tailed) linear model is equivalent to a linear mixture normal model (*i.e.*, with mixing variances), under mild conditions. The sources of unobserved heterogeneity might be various. For example, Rouse (1961) said that "...cold statistics do not provide sufficient basis for the conduct of day-to-day operations. We also rely heavily on the specialists who work on our Trading Desk, which serves as the listening post of the Federal Reserve System to the nation's money and securities markets." Then Lombra (1993) described the open market operations as partly "science" and partly "art". They seem to claim that there is an unpatterned factor involved in the Desk's behavior.

The statistical effect of this factor is heteroscedasticity. Because of its unobservability, this unpatterned factor is called unobserved heterogeneity in economics' terms. From another point of view, the real economy is surrounded by various shocks such as technology shock, utility shock, etc. These shocks would be sooner or later transmitted to the reserves market. To implement the FOMC's monetary policy and stabilize financial markets, the Desk can take discretionary actions to respond to these shocks. Again, this generates unobserved heterogeneity. As discussed before, ignoring this feature (Feinman (1993)) can greatly affect the model fit of the spike. Even for a simple linear model, it is known that selecting a "wrong" data distribution would distort parameter estimates and other inferences. This distortion can be worse when combined with a friction structure. Therefore, unobserved heterogeneity will be fully taken into account in this Chapter.

Operating on a daily basis, the open market operations should be analyzed in a dynamic

setting. It is clear that today's market intervention will influence the Desk's decision tomorrow. This consideration might be partly grounded by Mulendyke (1989) who noted that the Desk tried to smooth its actions to some extent. Smoothing reflects a positive autocorrelation. Roughly, looking at the sample statistics of the data (Table 1) confirms the point. The dynamic characteristic of the data is not even mentioned in Feinman's work, probably because adding dynamics makes a friction model intractable in terms of conventional statistical approaches. The major obstacle of a dynamic friction model, as is realized by Wei (1995) in a simpler version of this class of models, is due to its likelihood function involving multiple integrals. This is, however, no longer a problem with recent advances of the Markov Chain Monte Carlo (MCMC) methods, and fast computing technology. Based on Wei (1995), this Chapter will show the merits of these simulation-based statistical methods in solving such a dynamic friction model. Ignoring the dynamics may also distort parameter estimates, and at least results in efficiency loss. Two types of dynamics — observed and unobserved — are introduced in this Chapter in the hope of identifying the Desk's "true" behavior. The differences between these two types of behaviors are justified in terms of economics.

This Chapter adopts a Bayesian perspective for several reasons: (1) it obtains exact inference results without relying on any asymptotic approximation; (2) it can focus on any quantity of interest and easily derive its full posterior distribution; (3) both the parameter and model uncertainty can be readily taken into account; (4) with recent advances in faster computers, there is no computational burden for all of my inferences; and (5) as the posterior distributions of parameters deviate (significantly) from the normal family, it might be misleading to use posterior moments as summary statistics, because the existence of these quantities is questionable. This is clearly a disastrous situation for frequentists, but won't be a problem for Bayesians. Instead of using the quadratic loss, the absolute loss structure can be adopted and consequently the posterior medians of the parameters, which always exist, are chosen as summary statistics. Highest Probability Density (HPD) intervals (or regions)

can be used to measure the dispersion of the posterior distributions.

The outline of this work is as follows. After an introduction in Section 1, Section 2 provides and discusses competing specifications of the Desk's daily reactions in the Fed's funds market. Section 3 describes the data, Section 4 specifies the priors, and Section 5 derives the likelihood function. The estimation procedure is developed in Section 6, and the estimation results are reported in the section too. The performances of the models are compared in Section 7, and Section 8 contains the predictive analysis. Finally, Section 9 concludes the Chapter.

2 The Models of the Desk's Behavior

This section discusses the characteristics of the Desk's operational data, and proposes econometric models to capture them.

2.1 The Models

A striking characteristic of the Desk's daily actions, observed by Feinman (1993), is that the Desk took no action for a quite number of days during his sample period. In my sample period, which is close to Feinman's, this characteristic is illustrated in the data histogram (see Figure 1 (a)) as a **spike** at zero. The reason why the Desk's behavior generates such a spike, not discussed in Feinman (1993), is explained now. Apparently, the Desk's market operations are associated with a cost — **operational cost** which might be of several sorts: (1) similarly to a simple portfolio management problem (See Rosett (1958)), a *transaction cost* is immediately evident; (2) there is an *implementation cost*. Mulendyke (1989, p. 154), the open market Desk's manager, stated that “a market operation, in contrast, is rather cumbersome. While a market go-around is in progress, it may be difficult for dealers to trade since they do not know the results of their bids or offers with the Fed. Consequently, the desk prefers to limit such operations to times when the prospective reserve excess or shortage is relatively large.”:

(3) a *signalling cost* is also in play. The Desk's actions are assumed to convey the FOMC's policy signals on to the reserves market (see, for example, Simon (1991)). These signals exert influences on the expectations of economic agents and thereby affect their economic decisions. They are costly because the Desk takes extra effort to avoid confusing actions; and (4) an *adverse action cost* is the one for the Desk to take adverse actions in consecutive days. To avoid the problem, the Desk prefers conservative (less) actions. Analogously to Rosset's (1959) transaction cost story, this operational cost also induces a "friction" for the Desk's behavior. It is clear that any continuous distribution would not fit the spike of the data distribution (Figure 1). Following Rosett (1959) (see also Maddala (1987) and Desarbo, Rao, Steckel, Wind and Colombo (1987) for additional references), Feinman took a friction model specification for the Desk's daily behavior. Intuitively, this implies that there is a latent (*i.e.*, unobserved) pressure indexed by a latent (or desired) variable. This latent variable triggers the Desk to act once this latent pressure moves beyond sufficiently high/low threshold levels. These threshold levels represents a tolerance band for the Desk's "*inert or rigidity action area.*" This inert or rigidity area gives the spike at zero for the data histogram (Figure 1 (a)). The parameters of the tolerance band are of particular interest because they represents a measure of friction of the Fed's viewing the operational cost. For a general consideration, a friction regression model can be written as

$$y_t = \begin{cases} y_t^* - \delta^+ & \text{if } y_t^* > \delta^+ \\ 0 & \text{if } \delta^- \leq y_t^* \leq \delta^+ \\ y_t - \delta^- & \text{if } y_t^* < \delta^- \end{cases} \quad (1)$$

$$y_t^* = x_t' \gamma + z_t' \lambda + u_t$$

where y_t is the Desk's observed actions (buying or selling the government securities); y_t^* stands for the latent (desired) actions, and is thus unobserved; the transformation from y_t^* to y_t is often called the "observation rule"; the threshold parameters δ^+ and δ^- form the "inert band" $[\delta^-, \delta^+]$; x_t 's are covariates, which will be described in the next section; z_t 's represent lagged y_t or y_t^* ; capturing the dynamics; the u_t is i.i.d. error term; and the parameters of interest

are $\{\gamma, \sigma^2, \delta^+, \delta^-\}$. $y_t > 0$ implies the Desk's purchase of securities, $y_t < 0$ the Desk's sale of securities, and $y_t = 0$ no action. A constant term in this model is excluded for two reasons: (1) the constant term is unidentified; and (2) all covariates are just the deviation (from their targets) or projection variables so that it is reasonable to set the constant term to be zero when all forcing variables are zeros. Feinman's specification is given by

Model 1. (1) with $\lambda = 0$, and $u_t \sim i. i. d. N(0, \sigma^2)$.

The focus of this model is to explain the friction (most importantly, the spike in the data distribution (see Figure 1 (a)) for the Desk's behavior.

Like many financial and economic data, unobserved heterogeneity can be an important characteristic for the open market operations data. As will be shown later, it (*i.e.*, the unobserved heterogeneity) reflects a fat-tailed data distribution. There are various sources for this data feature. Most prominently, the Desk's open market operations are a discretionary tool in implementing monetary policy. The aspect of the discretion is directly contributed to the unobserved heterogeneity. Furthermore, while the Fed can by no means understand the real economy completely, the Desk's implementation is often called to be partly "science" and partly "art". The "artistic" aspect of the Desk's operations assisted by many of its staff members may also result in the unobserved heterogeneity. Finally, the economy is surrounded by various shocks such as technology shock, utility shock, etc. These shocks will be sooner or later transmitted to the reserves market that alter the equilibrium position of the demand and supply in the market. Clearly, similar equilibrium positions in the reserves market may be derived by different shocks from the various aspects of the economy. The Desk may thus respond somewhat differently due to the different natures of shocks. This would be also added to the unobserved heterogeneity. To model this feature of the Desk's behavior, a mixing parameter ω_t for each observation in the regression model (1) is introduced. Combining with the friction structure mentioned above, this leads to another model specification of the Desk's

behavior.

Model 2. (1) with $\lambda = 0$, $u_t = \omega_t^{1/2}\epsilon_t$ and $\epsilon_t \sim \text{i.i.d. } N(0, \sigma^2)$.

The data are thus assumed to be drawn from a mixture normal distribution. the σ^2 represents the common variance of all observations, and the ω_t stands for the unobserved heterogeneity. From a statistical point of view, this specification is not complete yet since there are more parameters than observations (the same remark is applicable to Model 3 and Model 4 below). I will be back to discuss the model in the next subsection.

The third characteristic of the data is **dynamics**. Operating in a daily basis. the Desk's actions should be analyzed in a time domain. This is consistent with Mulendyke (1989) (see the discussion in Section 1 of this Chapter). In fact, the autocorrelation coefficient. .2. of the data in Table 1 roughly confirms this point. Following the censored model literature (see Maddala (1987) and Wei (1985), generally, there are two ways to introduce the dynamics. They can be formally written as

Model 3. (1) with $z_t = y_{t-1}$, $u_t = \omega_t^{1/2}\epsilon_t$ and $\epsilon_t \sim \text{i.i.d. } N(0, \sigma^2)$.

Model 4. (1) with $z_t = y_{t-1}^*$, $u_t = \omega_t^{1/2}\epsilon_t$ and $\epsilon_t \sim \text{i.i.d. } N(0, \sigma^2)$.

The difference between Model 3 and Model 4 is whether the lagged latent actions or lagged observed actions explain the Desk's behavior. In this Chapter, we take only the first order of these lags into account, and the higher orders can be handled without further difficulty. Referring to a discussion in Wei (1995), Model 3 is only a special case of Model 4 from a statistical point of view. Further, I denote the parameter λ as a scaler representing the coefficient of the lagged y_t or y_t^* . Clearly, Model 1 and Model 2 are nested in both Model 3 and Model 4. In fact, Model 4 stands for the most general one among the four models (this will be seen later). Comparing with Model 4, other three models are far more easy to be dealt with. This is because standard Bayesian and classical methods are still applicable to the three

models. Unfortunately, these methods are too difficult to solve Model 4. if not impossible. To complete the specification of Model 2. and develop the estimation procedure of Model 4. an equivalent consideration of the models is required.

2.2 An Equivalent Consideration

The problem associated with the mixing parameters ω_t 's deserves a special attention. The literature has provided a technique in dealing with the problem in a Bayesian framework. Following Geweke (1993), under the prior specification $\nu/\omega_t \sim \chi^2(\nu)$. Model 4 is equivalent to

Model 4'. (1) with $z_t = y_{t-1}^*$ and $u_t \sim t(0, \sigma^2, \nu)$.

where $t(0, \sigma^2, \nu)$ is a Student-t distribution with mean zero, variance parameter σ^2 and degree of freedom ν . The importance of doing so is: (1) reducing the model parameters (recalling that in Model 2 — Model 4, the number of parameters is more than observations!); and (2) providing an insight on the unobserved heterogeneity, *i.e.*, a fat tailed data distribution. The mixing parameters ω_t 's are controled by the degree of freedom parameter (in Model 4') which thus measures the degree of deviation of the data distribution from the normal family. A large ν is associated with a small deviation, and a small ν indicates a large deviation. It is also interesting to note that the degree of the unobserved heterogeneity introduced this way can be automatically estimated. The parameter vector of interest is

$$\theta = \{\gamma, \sigma^{-2}, \omega, \nu, \delta^+, \delta^-\}$$

where $\gamma = (\beta, \lambda)$, $\omega = (\omega_1, \omega_2, \dots, \omega_T)$. The precision parameter σ^{-2} , rather than variance parameter σ^2 , is used for convenience.

3 Data Description

The data used in this study are provided by the Federal Reserve Board in Washington D.C., U.S.A. The sample spans from January 2, 1986 to July 7, 1989 with daily frequency. All weekends and national holidays are excluded from the data since the reserves market and the open market operations are both closed on these days. The definitions and descriptions of the variables, whose units are in million of dollars except that the variable x_5 is in percentage point, are discussed below.

3.1 Dependent Variable

The dollar effect of the Desk's market operations on the maintenance period-average level of reserves (y_t)

As is known, the Desk intervenes in the reserves market mainly by selling or buying various Treasury securities. The dollar effect of the Desk's transactions on the maintenance period-average level of reserves is the Desk's major concern on its operations [Mulendyke (1989, 148)]. According to Feinman (1993), this variable is manipulated as follows. First, all desk operations, temporary or permanent, with market participants or with foreign accounts, are included. Second, for a temporary transaction², y_t equals the par value of the securities multiplied by the number of days spanned by the transaction (including weekends and holidays) and divided by the number of days in the maintenance period (fourteen). This can be simply written as

$$y_t = \frac{\text{par value of the securities} \times \# \text{ of days spanned by the transaction}}{14} \quad (2)$$

²Temporary transactions include repurchase agreements (RPs) on the System's own account, matched-sale transactions (MSPs), and customer-related RP (CRP) [Mulendyke (1989, pp. 155 - 157)]. RPs are temporarily add reserves, MSPs are temporarily drain reserves, and CRP is also used to temporarily add reserves.

The signs of y_t are determined by whether the transaction adds or drains reserves. For example, a \$1.4 billion overnight repurchase agreement (RP), conducted by the Desk on a Tuesday, would add \$100 million ($100 \text{ million} = 1.4 \text{ billion} / 14$) to the maintenance period-average level of reserves. Consequently, $y_t = 100$ is recorded on the Tuesday. If the same transaction is conducted on a day preceding a two-day holiday, like a Friday, then we have $y_t = 300$ on the day ($300 \text{ million} = 1.4 \text{ billion} \times 3 / 14$). On the contrary, if a \$1.4 billion, three-day, matched sale-purchase (MSP) is conducted, then we have $y_t = -300$ on the day. The minus sign indicates that the transaction drains reserves from the market. In fact, the Desk relies heavily on temporary reserve operations in dealing with the uncertainties that affect bank reserves. Finally, an outright transaction³ is assumed to add or drain reserves permanently during the maintenance period. Consequently, the number of days spanned by the transaction in equation (2) is equal to the number of days remaining (including the current day) in the maintenance period for an immediate delivery of this transaction. For example, if a \$1.4 billion outright purchase of Treasury securities is conducted by the Desk on the 9th day of a maintenance period for an immediate delivery, then $y_t = 600$ is recorded ($600 \text{ million} = 1.4 \text{ billion} \times 6 / 14$). If the delivery is delayed for this transaction, then the number of remaining days in the period is counted from the delivery day. For instance, if the above outright transaction is delivered on the next day, then $y_t = 500$ is recorded on the 9th day. On the contrary, if the same amount of sale or redemption transaction is conducted on the 9th day for an immediate delivery, then $y_t = -600$ is denoted. If this sale or redemption transaction is delivered one day late, then we simply have $y_t = -500$.

³Outright transactions include outright purchases of Treasury securities, which would add reserves, and outright sale or redemption of Treasury securities, which would drain reserves [see Mulendyke (1989, pp. 151 - 155)].

3.2 Independent Variables

The maintenance period-average reserve need (x_{1t})

This variable is the projection of the Fed's staff members (including both the Federal Reserve Board and the Federal Reserve Bank of New York with equal weight on consideration) on the daily operations. It is obtained as follows. First, the borrowed reserve assumptions on the on-going maintenance period is specified in terms of the FOMC's Directive. Then the staff members estimate the maintenance period-average demand for reserves by projecting required reserves against deposits and the desired excess reserves of the banking system. Next, the non-borrowed reserve path is formed by subtracting the borrowing assumption from the period-average demand for reserves. Finally, the non-borrowed reserve path minus certain technical factors⁴ yields the "reserve need".

To maintain consistency with the dependent variable y_t , the "reserve need" variable is divided by the number of days remaining in the period (including the current day). The informal adjustments made by the Desk also include: (1) changes in the projection of excess reserve demand owing to incoming data on carryover, etc.; (2) special situation borrowing arising from computer problems or natural disasters that disturb the normal wire transfer systems and, for the purpose of conducting open market operations, are treated by the Desk as nonborrowed reserves; (3) expected withdrawals from outstanding term RPs. At all times, we use only information that was available to the Desk each morning at the standard intervention time.

⁴See the discussions of the technical factors in the Fed (1985, pp. 36-37), Mulendyke (1989, pp. 141 - 147) and Partlan, Hamdani, and Camilli (1986).

The intraperiod distribution of the reserve need in the earlier period (x_{2t}) and late period (x_{3t})

In determining the size and timing of its operation, the Desk also takes into account the intraperiod distribution of reserve deficiencies (surfeits). This is done to avoid extraordinary daily swings in reserve availability that could potentially induce inadequate reserves for clearing purposes and sharp gyrations in borrowing and the funds rate.

To capture this aspect of the Desk's behavior, we subtract, on each day of our sample, the staff's forecast of nonborrowed reserve availability on the day from the period-average reserve objective to obtain the projected daily deficiency (surfeit). To the extent that the Desk tries to smooth the day-to-day pattern of reserve availability, this variable should be an important determinant of open market operations. That is, the Desk should be more likely to address a given period-average add need on a day in which reserves are expected to be deficient than on a day in which reserves are apt to be in surplus.

Estimated cumulative reserve deficiency (surplus) to date (x_{4t})

The staff prepares estimates each day of actual free reserves accumulated since the start of the period. These estimates point to pent-up pressures in the reserve market that the Desk often takes into consideration when choosing its operations. If, for example, free reserves appear to be running well below the path assumption, the Desk might defer meeting a period-average drain need for a few days, thereby allowing a reserve cushion to build up. Alternatively, if the Desk faced a period - average add need that was not expected to manifest itself until later in the period - that is, if free reserves to date were in surplus and there was no deficiency on the day - the Desk would be less apt to address this need immediately than if the reserve need had already materialized or was expected to appear that day.

The interacting effects of the period-average need and the distribution of that need on the

Desk's behavior may vary over the course of the period. As settlement day approaches, for example, the Desk's scope for smoothing the intraperiod reserve profile diminishes because reserves on the previous days of the period have already been determined. As a result, the coefficients on the intraperiod distribution of the need have been allowed to differ over different parts of the period.

Funds rate at Fed time minus the funds rate "target" (x_{5t})

Each morning the Desk monitors the prevailing funds rate relative to the rate expected to be consistent with the reserve objective specified by the FOMC. The Desk's responsiveness to deviations of the funds rate from expectations is a key gauge of the Fed's desire to keep that rate near the anticipated level. Structural breaks in the Desk's reaction function owing to shifts in the estimated coefficient on the funds rate deviation are evidence to the extent to which operations are keyed to that rate. These changes, in turn, presumably reflect shifts in the FOMC's emphasis on the funds rate as an operating instrument. This is a very important variable, since it captures information of the reserve pressure monitored by the Fed.

4 The Prior Specification

Bayesian analysis uses formal consideration of prior belief on model parameters. The prior belief is presumably formed before the data are actually observed. In most cases, this belief demands a probabilistic description, but diffuse priors are often accepted for the purpose of estimation. Unfortunately, non-informative priors are almost useless in conducting model comparison by using Bayes factor. I will mainly focus on informative priors, leaving non-informative priors considered only in my sensitivity analysis for estimations.

For convenience, I start with a decomposition of the prior distribution of the model pa-

rameters.

$$p(\theta) = p(\gamma, \sigma^{-2}, \omega, \nu, \delta^+, \delta^-) = p(\gamma, \sigma^{-2})p(\omega|\nu)p(\nu)p(\delta^+)p(\delta^-)$$

This decomposition facilitates our belief formulation in a probabilistic manner.

First, I consider possible values of the tolerance band parameters. Since δ^+ represents the minimum equilibrium amount of buying government securities, and δ^- the maximum equilibrium amount of selling government securities, I can thus assume that δ^+ takes on positive values and δ^- on negative values. In particular, I assume that δ^+ and δ^- are independent and uniformly distributed over $(0, \bar{\delta})$ and $(\underline{\delta}, 0)$, respectively. The hyperparameters $\underline{\delta}$ and $\bar{\delta}$ are chosen to be -5000, and 5000 so as to ensure that the δ^+ and δ^- fall safely in their prior supports⁵.

Next, as can be easily seen, conditioning on δ^+ and δ^- , Model 4 collapses into a dynamic linear model with an independent Student-t distribution. Noting the equivalence argument in Subsection 2.2, or in Geweke (1993), I take an independent inverted $\chi^2(\nu)$ prior for ω_t : $\frac{\omega_t}{\nu}|\nu \sim \chi^2(\nu)$. This is the condition that guarantees the model equivalence in that section. The ν is the degree of freedom parameter, and it should not take a non-informative prior as argued by Geweke (1993) since otherwise it amounts to imposing a normal distribution on the model error term. Following Geweke (1993), an exponential power prior distribution for ν

$$p(\nu|\alpha) = \alpha (-\alpha\nu)^{\nu-1} \quad \nu > 0$$

is conveniently taken.

Third, I use a standard normal-gamma prior for the regression coefficients and the common precision parameter σ^{-2} .

$$p(\gamma, \sigma^{-2}|\underline{\gamma}, \underline{Q}, \underline{s}^{-2}, \underline{\nu}) = \phi_k(\underline{\gamma}, \sigma^2 \underline{Q})\gamma(\sigma^{-2}|\underline{s}^{-2}, \underline{\nu}) \quad -\infty < \gamma < +\infty, \quad \sigma^{-2} > 0$$

⁵There can be at least two ways for the choice of these hyperparameters. One is to choose them large enough and another one is to form further prior specifications on them. Both of the methods lead to almost same results.

where $\{\underline{\gamma}, \underline{Q}, \underline{\sigma}^{-2}, \underline{\nu}\}$ are hyperparameters. To extend my readership, Jeffrey's diffuse prior $p(\gamma, \sigma^{-2}) \propto \sigma^{-2}$ is also considered.

The results presented in Tables 4-8 are based on a prior that is reasonably flat (the hyperparameters are presented in Table 2) and that the various models have basically the same prior.

5 The Likelihood Function

This section derives the likelihood functions of the models in Section 2. The focus will be given to Model 4 (or Model 4'). Based on Wei (1995), the idea of this derivation is the following. Obtain the sampling distribution of all observable and unobservable data and then integrate out the unobserved data since they are viewed as random variables. Zeger and Brookmeyer (1986) found that the traditional Markov property of dynamic data no longer holds when they are also censored. The difficulty induced by the failure of the Markov property is the multi-dimensional integrals in the model's likelihood function. This difficulty continuously spreads its impact on making various statistical inferences. For exposition purpose, I set only one period lag for Model 4. For more lags, simply combine the following procedure with Wei (1995). Then I define the following notations.

(1) $I_1 = \{t | y_t > 0, 1 \leq t \leq T\}$, *i.e.* the subset of time points corresponding to the positive observations of y_t ;

(2) $I_2 = \{t | y_t = 0, 1 \leq t \leq T\}$, *i.e.* the subset of time points corresponding to the zero observations of y_t ;

(3) $I_3 = \{t | y_t < 0, 1 \leq t \leq T\}$, *i.e.* the subset of time points corresponding to the negative

observations of y_t ;

(4) given any $t \in I_1 \cup I_3$, n_t stands for the number of consecutive latent observations following t .

This means that if $y_{t+1} \neq 0$, then I have $n_t = 0$; if $y_{t+1} = 0$ and $y_{t+2} \neq 0$, then I have $n_t = 1$; more generally, if $y_{t+1} = 0, y_{t+2} = 0, \dots, y_{t+p} = 0$, and $y_{t+p+1} \neq 0$ (including the case that there is no more observation beyond y_{t+p}), then $n_t = p$. Following Zeger and Brookmeyer's (1986) notation, I call each run of consecutive censored observations a *latent string*. Therefore, n_t is the length of the latent string following t .

(5) $r_t = n_t + 1$.

With the above notations, I start to derive the joint distribution of each latent data string.

Iterating Model 4 yields

$$y_{t+h}^* = \left(\sum_{i=1}^h \lambda^{h-i} x'_{t+i} \right) \beta + \lambda^h y_t^* + \sum_{i=1}^h \lambda^{h-i} u_{t+i}$$

where $t \in I_1 \cup I_3$ and $h = 1, 2, \dots, T$. Given observable y_t , the above expression can be generally written as

$$y_{t+h}^* = \left(\sum_{i=1}^h \lambda^{h-i} x'_{t+i} \right) \beta + \lambda^h \left[y_t + \delta^+ I_{\{y>0\}}(y_t) + \delta^- I_{\{y<0\}}(y_t) \right] + \sum_{i=1}^h \lambda^{h-i} u_{t+i} \quad (3)$$

where $I_A(x)$ is an indicator function defined as

$$I_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

Note that all variables involved in the right hand side of equation (3) are observable, except

the error terms. This idea simplifies the approach to the joint distribution of each latent string. To see this, I further define

$$\eta_{t+h} = \sum_{i=1}^h \lambda^{h-i} u_{t+i}$$

$$Q_{t+h} = \left(\sum_{i=1}^h \lambda^{h-i} x'_{t+i} \right) \beta + \lambda^h \left[y_t + \delta^+ I_{\{y>0\}}(y_t) + \delta^- I_{\{y<0\}}(y_t) \right]$$

and note

$$\text{Cov}(\eta_{t+h}, \eta_{t+h+l}) = \sigma^2 \sum_{i=1}^h \lambda^{2(h-i)+l} w_{t+i}$$

where $h = 1, 2, \dots, T$ and $l = 0, 1, \dots, T-h$. Denote $\sigma_{ji} \equiv \sigma_{ij} = \text{cov}(\eta_{t+i}, \eta_{t+j})$, $\phi(\cdot) =$ density function for a normal distribution. Now given any $t \in I_1 \cup I_3$, if $r_t > 1$, $y_{t+1}^*, y_{t+2}^*, \dots, y_{t+r_t}^* | y_t \sim \phi(y_{t+1}^*, y_{t+2}^*, \dots, y_{t+r_t}^* | A_{r_t}, \Omega_{r_t})$ where

$$A_{r_t} = \begin{pmatrix} Q_{t+1} \\ \vdots \\ Q_{t+r_t} \end{pmatrix} \quad \Omega_{r_t} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1r_t} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2r_t} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{r_t 1} & \sigma_{r_t 2} & \cdots & \sigma_{r_t r_t} \end{pmatrix}. \quad (4)$$

If $r_t = 1$, then $y_{t+r_t}^* | y_t = y_{t+1} | y_t$ has an univariate normal distribution, i.e. $\phi(y_{t+1} | x'_t \beta + \lambda y_t, \sigma^2 \omega_t)$. Otherwise, it is a multivariate normal distribution.

Proposition 5.1 With the above notations, the likelihood function of Model 4 can be written as

$$\begin{aligned} L(y_1, y_2, \dots, y_T | \gamma, \sigma^2, \delta^+, \delta^-) &= \prod_{t \in I_1 \cup I_3} f(y_{t+1}, \dots, y_{t+r_t} | y_1, \dots, y_t) \\ &= \prod_{t \in I_1 \cup I_3} f(y_{t+1}, \dots, y_{t+r_t} | y_t) \\ &= \prod_{t \in I_1 \cup I_3} f_t \end{aligned}$$

where, from the conditional distributions discussed before, for $r_t > 1$.

$$\begin{aligned}
f_t &= \int_{\delta^-}^{\delta^+} \cdots \int_{\delta^-}^{\delta^+} \phi(y_{t+1}^*, \dots, y_{t+n_t}^*, y_{t+r_t}^* | y_t) dy_{t+1}^* \cdots dy_{t+n_t}^* \\
&= \int_{\delta^-}^{\delta^+} \cdots \int_{\delta^-}^{\delta^+} \phi(\eta_{t+1}, \dots, \eta_{t+n_t}, \eta_{t+r_t} | y_t) d\eta_{t+1} \cdots d\eta_{t+n_t} \\
&= \int_{\delta^-}^{\delta^+} \cdots \int_{\delta^-}^{\delta^+} \phi(\eta_{t+1}, \dots, \eta_{t+n_t} | \eta_{t+r_t}, y_t) d\eta_{t+1} \cdots d\eta_{t+n_t} \phi(\eta_{t+r_t} | y_t)
\end{aligned}$$

where $t(\eta_{t+1}, \dots, \eta_{t+n_t}, \eta_{t+r_t} | y_t)$ is a multivariate normal density function in r_t dimensions with mean vector 0, covariance matrix Ω_{r_t} , and $\phi(\eta_{t+1}, \dots, \eta_{t+n_t}, | \eta_{t+r_t}, y_t)$ is the derived conditional normal density of $\eta_{t+1}, \dots, \eta_{t+n_t}$, given η_{t+r_t} and y_t .

For $r_t = 1$, f_t becomes

$$f_t = \phi(\eta_{t+1} | y_t) = \phi(y_{t+1} - x'_{t+1}\beta - \lambda y_t | y_t)$$

where ϕ is an univariate Student-t density function with mean 0, variance $\sigma^2\omega_t$. To be careful here, if the final (Tth) observation is censored (i.e. $y_T = 0$), then for the largest $t \in I_1 \cup I_3$, I have

$$\begin{aligned}
f_t &= \int_{\delta^-}^{\delta^+} \cdots \int_{\delta^-}^{\delta^+} \phi(y_{t+1}^*, \dots, y_{t+n_t}^* | y_t) dy_{t+1}^* \cdots dy_{t+n_t}^* \\
&= \int_{\delta^-}^{\delta^+} \cdots \int_{\delta^-}^{\delta^+} \phi(\eta_{t+1}, \dots, \eta_{t+n_t} | y_t) d\eta_{t+1} \cdots d\eta_{t+n_t}
\end{aligned}$$

where the f_t a probability rather than a density function.

The likelihood of Model 4 can be also expressed in t-distributions. The derivation is similar to the above, except for (1) changing each normal distribution into a corresponding t-distribution, and (2) changing the covariance matrix in (3) to the following scale matrix,

$$\Omega'_{r_t} = \begin{pmatrix} \sigma'_{11} & \sigma'_{12} & \cdots & \sigma'_{1r_t} \\ \sigma'_{21} & \sigma'_{21} & \cdots & \sigma'_{2r_t} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma'_{r_t 1} & \sigma'_{r_t 2} & \cdots & \sigma'_{r_t r_t} \end{pmatrix}$$

where

$$\sigma_{j,i'} \equiv \sigma_{i,j'} = \sigma^2 \sum_{i=1}^h \lambda^{2(h-i)+l}$$

If $\nu > 2$, the covariance matrix of the latent string $\{y_{t+1}^{\pi}, y_{t+2}^{\pi}, \dots, y_{t+r_t}^{\pi}\}$ is $\frac{\nu}{\nu-2} \Omega'_{r_t}$. This consideration is useful for predictions, which will be seen clearer in Section 8.

6 Estimation

Bayesian inference is based on Bayes' rule

$$p(\theta|y) = \frac{\pi(\theta)L(y, \theta)}{p(y)} \quad (5)$$

where $\pi(\theta)$ is the prior density; θ , a vector of parameters; y , a vector or matrix of data; $L(y, \theta)$, the likelihood function; $p(y)$, the marginal density of the data; $p(\theta|y)$, the posterior density. The estimational difficulty of Model 4 comes from the intractability of its likelihood as derived in Section 5. Following Wei's (1995) idea, this difficulty is avoided by using the Markov Chain Monte Carlo methods. This section reviews the Markov Chain methods, derives the estimation algorithm for Model 4 and reports the estimation results.

6.1 The Markov Chain Methods

In recent years the Markov Chain Monte Carlo Methods have greatly advanced the ability of the Bayesian community to handle non-standard or complex distribution problems. As an important special case, the Gibbs sampler is widely recognized. The Metropolis-Hasting (MH) sampler is a more general case. For references, see Tierney (1994), Chib and Greenberg (1994), Evans and Swartz (1995) and references therein. The data augmentation method developed by Tanner and Wong (1987) is often combined with these Markov Chain method to deal with latent data problems. In fact, the data augmentation method allows a symmetric

treatment of the latent data and the parameters of the model, except that no prior needs to be specified for the former. For my purposes, a brief and general discussion is provided below.

To draw a variate from a *target distribution* $\pi(\theta)$ (for example, a posterior or predictive distribution), given any *transitional distribution*⁶ $q(\theta, x)$, set

$$\alpha(\theta, x) = \begin{cases} \min \left[\frac{\pi(x)q(x, \theta)}{\pi(\theta)q(\theta, x)}, 1 \right] & \text{if } \pi(\theta)q(\theta, x) > 0 \\ 1 & \text{otherwise.} \end{cases}$$

With any feasible initial vector θ^0 , MH algorithm is to make such an iteration. Suppose θ^j has been sampled, then generate x from $q(\theta^j, \cdot)$ and u from $U(0, 1)$. if $u \leq \alpha(\theta^j, x)$, set $\theta^{j+1} = x$; else, set $\theta^{j+1} = \theta^j$. Then it has been proved that $\theta^n \rightarrow \theta$ in distribution as $n \rightarrow \infty$. Thus, as long as n is selected sufficiently large, the MH algorithm shows how to sample a variate from the target distribution $\pi(\theta)$.

In practice, however, it is often difficult to find an appropriate *transitional distribution* $q(\theta, x)$ when θ represents a high dimensional parameter vector. In such a case, the MH algorithm can be applied to the sub-blocks of θ . This is called “*block at a time algorithms*” in Chib and Greenberg (1994, p. 19). The key idea is rather simple. Partition θ into sub-blocks, say $(\theta_1, \theta_2, \dots, \theta_g)$. Suppose that the conditional transition kernels $g(\theta_i, x | \theta_j, j \neq i)$ for the conditional distribution $\pi(\theta_i | \theta_j, j \neq i)$ are available. If each of the full conditional distributions $\pi(\theta_i | \theta_j, j \neq i)$ can be sampled directly, then this becomes the Gibbs sampler. If part of the full conditional distributions are known, then a hybrid method can be applied. That is, sample a variate from each conditional distribution $\pi(\theta_i | \theta_j, j \neq i)$. If $\pi(\theta_i | \theta_j, j \neq i)$ is known, sample directly. Otherwise, apply the general MH algorithm to this conditional distribution. Sometimes the general “*block at a time algorithm*” is also called the Gibbs sampler in the literature. In this Chapter, our estimation strategy is the same as the latter case.

⁶ *Transitional distribution* determines the transitional probability from one state to another state. The choice of transitional probability is a case-based problem. A well selected transitional distribution, though sometimes this is not easy, makes the produced Markov Chain converge fast.

Finally, if latent data Z exist, they can be treated symmetrically to the parameter θ . This is based on the data augmentation method (Tanner and Wong (1987)). In this case, the above MH method is still applicable, noting that the target distribution now is $\pi(\theta, z)$, rather than $\pi(\theta)$.

6.2 Conditional Distributions

Full conditional distributions must be derived before the “*block at a time algorithm*” can be applied. The partition of the parameter vector is rather natural, since I simply follow the prior specification blocks in Section 4. The full conditional distribution of each block is derived below.

6.2.1 The Conditional Distributions of Latent Data

In Wei (1995), I developed an idea to derive the conditional distributions of the latent data for a dynamic Tobit model. This idea is also applied here to Model 4. Recalling the notations defined in Section 5, given any $t \in I_1 \cup I_3$, if $n_t = 1$, the latent string is y_{t+1}^* ; if $n_t > 1$, then the latent string is $y_{t+1}^*, y_{t+2}^*, \dots, y_{t+n_t}^*$. It is easy to see that the distribution of each latent string, $y_{t+1}^*, y_{t+2}^*, \dots, y_{t+n_t}^*$ (n_t may be equal to or greater than 1) conditional on all other y_τ^* which includes all uncensored observations, other latent strings, and θ , is the same as its distribution conditional on y_t and y_{t+r_t} ($r_t = n_t + 1$). As indicated in Wei (1995), this is based upon the special structure of the Markov chain property of Model 4. From the discussion of the likelihood function in Section 5, I can see that $y_t^*, y_{t+1}^*, \dots, y_{t+r_t}^* | \theta$ is distributed as a multivariate normal distribution. Therefore, I have the following conditional distribution,

$$y_{t+1}^*, y_{t+2}^*, \dots, y_{t+n_t}^* | \{y_t^*, y_{t+r_t}^*, \theta\} \sim \phi \left(y_{t+1}^*, \dots, y_{t+n_t}^* | A_{n_t|r_t}, \Omega_{n_t|r_t} \right) \quad (6)$$

where $A_{n_t|r_t}$ and $\Omega_{n_t|r_t}$ are corresponding conditional mean (vector) and conditional variance (matrix) (see (4)). Since y_t^* and $y_{t+r_t}^*$ are both observable, $y_{t+1}^*, \dots, y_{t+n_t}^* | \{y_t^*, y_{t+r_t}^*, \theta\}$ is

exactly same as $y_{t+1}^*, \dots, y_{t+n_t}^* | \{y_t, y_{t+r_t}, \theta\}$. Incorporating the information about $\{y_{t+1} = 0, y_{t+2} = 0, \dots, y_{t+n_t} = 0\}$ in the conditional distribution (6), I now have the following latent data generation mechanism.

Proposition 6.1 Given a latent string $y_t^*, y_{t+1}^*, \dots, y_{t+n_t+1}^*$, the latent data in the string follows a truncated normal distribution

$$y_{t+1}^*, \dots, y_{t+n_t}^* | \{y_t, y_{t+r_t}, \theta\} \sim \phi \left(y_{t+1}^*, \dots, y_{t+n_t}^* | A_{n_t|r_t}, \Omega_{n_t|r_t} \right) I_{\{\delta^- \leq y_{t+\tau}^* \leq \delta^+, 1 \leq \tau \leq n_t\}}$$

where $A_{n_t|r_t}$ and $\Omega_{n_t|r_t}$ are conditional mean vector and variance matrix, respectively. Note that this is a truncated multivariate normal distribution if $n_t > 1$, and a truncated univariate normal distribution if $n_t = 1$.

Equivalently, if I employ Model 4' to generate the latent data, then truncated Student t-distributions are required.

Proposition 6.2 Given a latent string $y_t^*, y_{t+1}^*, \dots, y_{t+n_t+1}^*$, the latent data in the string follows a truncated Student-t distribution

$$y_{t+1}^*, \dots, y_{t+n_t}^* | \{y_t, y_{t+r_t}, \theta\} \sim t \left(y_{t+1}^*, \dots, y_{t+n_t}^* | A_{n_t|r_t}, \Omega'_{n_t|r_t}, \nu \right) I_{\{\delta^- \leq y_{t+\tau}^* \leq \delta^+, 1 \leq \tau \leq n_t\}}$$

where the notations here are similar to those in Proposition 6.1.

6.2.2 The Conditional Distributions of the Parameters

Given threshold parameters δ^+ and δ^- , and augmented data Z (if $y_t > 0$, $z_t = y_t - \delta^+$; if $y_t < 0$, $z_t = y_t - \delta^-$; and if $y_t = 0$, z_t is generated by Proposition 6.1), Model 4 becomes a linear normal regression model. According to Geweke (1993) and Poirier (1988), the following results can be derived.

Under the previous assumption and prior specified in Section 4, the following conditional

posterior distributions can be derived.

$$(1) \gamma | \{Z, \sigma^{-2}, \omega, \nu, \delta^+, \delta^-\} = \gamma | \{Z, \sigma^{-2}, \omega\} \sim N(\bar{\gamma}, \sigma^2 \bar{Q})$$

where

$$\bar{\gamma} = \bar{Q}[\underline{Q}^{-1}\underline{\gamma} + (X'\Omega^{-1}X)\hat{\gamma}] \text{ and } \hat{\gamma} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Z,$$

$$\bar{Q} = [\underline{Q}^{-1} + X'\Omega^{-1}X]^{-1}, \Omega = (\omega_1, \omega_2, \dots, \omega_T) \text{ and } \omega = (\omega_1, \omega_2, \dots, \omega_T)'$$

$$(2) \sigma^{-2} | \{Z, \gamma, \omega, \nu, \delta^+, \delta^-\} = \sigma^{-2} | \{Z, \gamma, \omega\} \sim \gamma(\sigma^{-2} | s^{*-2}, \nu^*)$$

where

$$\nu^* = \bar{\nu} + K, \quad s^{*-2} = \{\nu^{*-1}[\bar{\nu}\bar{s}^2 + (\gamma - \bar{\gamma})'\bar{Q}^{-1}(\gamma - \bar{\gamma})]\}^{-1}, \quad \bar{\nu} = \underline{\nu} + N \text{ and}$$

$$\bar{s}^2 = \bar{\nu}^{-1} \left[\underline{\nu}\underline{s}^2 + (Z - X\hat{\gamma})'\Omega^{-1}(Z - X\hat{\gamma}) + (\hat{\gamma} - \underline{\gamma})'\underline{Q}^{-1}(\hat{\gamma} - \underline{\gamma}) - (\hat{\gamma} - \bar{\gamma})'\bar{Q}^{-1}(\hat{\gamma} - \bar{\gamma}) \right]$$

and $\bar{\gamma}$, \bar{Q} and $\hat{\gamma}$ are given in (1).

$$(3) [(\sigma^2 u_t^2 + \nu)/\omega_t] | \{Z, \gamma, \omega, \nu, \delta^+, \delta^-\} = [(\sigma^2 u_t^2 + \nu)/\omega_t] | \{Z, \gamma, \omega, \nu\} \sim \chi^2(\nu + 1)$$

where $u_t = y_t^* - (x_t'\beta + \lambda y_{t-1}^*)$.

$$(4) \nu | \{Z, \gamma, \omega, \delta^+, \delta^-\} = \nu | \{Z, \gamma, \omega\} \sim (\nu/2)^{n\nu/2} \Gamma(\nu/2)^{-n} \exp(-\eta\nu)$$

where $\eta = \frac{1}{2} \sum_{t=1}^n [\ln(\omega_t) + \omega_t^{-1}] + \alpha$.

The intuition behind these distributions is the following. Once latent data are sampled, Model 4 becomes a complete dynamic linear regression model. As is known, a normal-gamma prior for $\{\gamma, \sigma^{-2}\}$ is a conjugate prior. Consequently, (1) and (2) are just posterior conditional distribution for γ and σ^{-2} . With unobserved heterogeneity parameter ω , the derivation of

this posterior normal-gamma distribution is similar to Poirier (1988). (3) and (4) are similar to the results obtained by Geweke (1993, p. S26-S27) and see also explanations there.

It is noted that $\{\gamma, \sigma^{-2}\}$ follows a normal-gamma distribution $NG_K(\gamma, \sigma^{-2} | \hat{\gamma}, \bar{Q}, \bar{s}^{-2}, \bar{v})$. This means that a normal-gamma prior is still conjugate for the parameters $\{\gamma, \sigma^{-2}\}$ for this model. Therefore, conditional distributions of (1) and (2) are derived from the normal-gamma posterior distribution (see Poirier (1995)).

If Jeffrey's diffuse prior is adopted for $\{\gamma, \sigma^{-2}\}$, then the above conditional distributions can be written as, respectively,

$$(1') \gamma | \{Z, \sigma^2, \omega, \nu, \delta^+, \delta^-\} = \gamma | \{Z, \sigma^2, \omega\} \sim N(\hat{\gamma}, \sigma^2 (X' \Omega^{-1} X)^{-1}).$$

$$(2') [\sum_{t=1}^T (u_t^2 / \omega_t) / \sigma^2] | \{Z, \gamma, \omega, \nu, \delta^+, \delta^-\} = [\sum_{t=1}^T (u_t^2 / \omega_t) / \sigma^2] | \{Z, \gamma, \omega\} \sim \chi^2(T).$$

For the threshold parameters, the following results hold.

Proposition 6.3 Given priors $\pi(\delta^+)$ and $\pi(\delta^-)$,

$$(1) \delta^+ | \{Z, \gamma, \sigma^{-2}, \nu, \omega, \delta^-\} \sim \pi(\delta^+) \exp \left[-\frac{1}{2} \left(\frac{\delta^+ - \sum_{t \in I_1} u_t \omega_t^{-1} / \sum_{t \in I_1} \omega_t^{-1}}{\sigma (\sum_{t \in I_1} \omega_t^{-1})^{-1/2}} \right)^2 \right] I_{\{\delta^+ \geq \{y_t^*, t \in I_2\}\}}$$

where $u_t = -(y_t - X_t' \beta - \lambda y_{t-1}^*)$, $t \in I_1$.

$$(2) \delta^- | \{Z, \gamma, \sigma^{-2}, \omega, \nu, \delta^+\} \sim \pi(\delta^-) \exp \left[-\frac{1}{2} \left(\frac{\delta^- - \sum_{t \in I_1} u_t \omega_t^{-1} / \sum_{t \in I_1} \omega_t^{-1}}{\sigma (\sum_{t \in I_1} \omega_t^{-1})^{-1/2}} \right)^2 \right] I_{\{\delta^- \leq \{y_t^*, t \in I_2\}\}}$$

where $u_t = -(y_t - X_t' \beta - \lambda y_{t-1}^*)$, $t \in I_3$.

See Appendix for the proof of this proposition.

6.3 The Estimation Algorithm

With the above preparation, according to the MH method, especially the *block at a time algorithm*, the estimation algorithm for Model 4 can be described as follows.

- *step 1*: starting point of the chain at a value $\{\gamma^0, (\sigma^2)^0, \omega_t^0, 1 \leq t \leq T, \nu^0, (\sigma^+)^0, (\sigma^-)^0\}$.
- *step 2*: loop starting at $i=1$.
- *step 3*: generate the latent data z_t^i according to Proposition 6.1
- *step 4*: generate the parameters $\{\gamma, \sigma^2, \omega_t, \nu\}$ according to Proposition 6.3. i.e.,
 - (i) generate γ^i from a multivariate normal distribution (see (1) or (1') of Proposition 6.3);
 - (ii) generate $(\sigma^{-2})^i$ from a γ distribution (see (2) or (2') of Proposition 6.3).
 - (iii) generate $\{\omega_t^i, 1 \leq t \leq T\}$ from a χ^2 distribution (see (3) of Proposition 6.3);
 - (iv) generate ν^i from a specified distribution (see (4) of Proposition 6.3).
- *step 5*: generate $(\delta^+)^i$ and $(\delta^-)^i$ according to the distribution specified in Proposition 6.4.
- *step 6*: increment i by 1 and go to step 3 unless $n > m + M$.
- *step 7*: discard initial m draws and return with last M draws.

The posterior output $\{Z^{(m)}, \theta^{(m)}\}$ from this algorithm can be written as

$$\theta^{(i)} = \left\{ \gamma^{(i)}, (\sigma^2)^{(i)}, \omega^{(i)}, \nu^{(i)}, (\delta^+)^{(i)}, (\delta^-)^{(i)} \right\}, \quad i = m + 1, m + 2, \dots, m + M \quad (7)$$

This output represents an approximate sample from the joint distribution of $p(\theta, Z|Y)$. The

marginal posterior densities can be easily written as

$$f(\theta_j|Y_T) = \frac{1}{M} \sum_{i=m+1}^{m+M} f(\theta_j|Y_T, z^{(i)}, (\theta_j)_{-1}^{(i)})$$

where θ_j can be any one of elements of θ and the subscript -1 indicates the parameters associated with it are the complementary ones of θ_j in parameters in θ . If an informative prior γ is used, then the posterior moments of the parameter θ exist. This can be proved by being analogous to Geweke (1993, Theorem 3). Further, according to a Rao-Blackwell argument (cf. Gelfand and Smith (1990)), the posterior means and variances of these parameters and latent data can be simply written as

$$\begin{aligned} \hat{E}(\theta_j|Y_T) &= \frac{1}{M} \sum_{i=m+1}^{m+M} E(\theta_j|Y_T, z^{(i)}, (\theta_j)_{-1}^{(i)}) \\ \hat{E}(z|Y_T) &= \frac{1}{M} \sum_{i=m+1}^{m+M} E(z|Y_T, \theta^{(i)}) \\ \hat{V}(\theta_j|Y_T) &= \frac{1}{M} \sum_{i=m+1}^{m+M} E((\theta_j)^2|Y_T, z^{(i)}, (\theta_j)_{-1}^{(i)}) - \{\hat{E}(\theta_j|Y_T)\}^2 \\ \hat{V}(z|Y_T) &= \frac{1}{M} \sum_{i=m+1}^{m+M} E(z^2|Y_T, \theta^{(i)}) - \{\hat{E}(z|Y_T)\}^2 \end{aligned} \tag{8}$$

The posterior estimates are more efficient than sampling means and variances of the posterior output (7).

In practice, two things about the method are mostly concerned: a convergence criterion of the Monte Carlo Markov chains and a measure of the computational accuracy of the estimates. An easily used criterion proposed by Yu and Mykland (1994) has been reported to behave relatively well (see Robert (1995) and Bauwens (1996)). This criterion is a visual inspection of CUMSUM statistics. Let N be the draws of a Monte Carlo Markov chain (MCMC). The CUMSUM statistic is given by

$$CS_t = \left(\frac{1}{t} \sum_{i=1}^t \theta^{(i)} - \mu \right) / \sigma$$

where μ and σ are the empirical mean and standard deviation of the N draws. If the MCMC converges, then the plot of CS_t against t should converge smoothly to zero. On the other

hand, a long and regular excursion plot of CS_t indicates the absence of convergence of the chain.

The accuracy of these estimates are measured in numerical standard errors, which might be computed by using the well-known batch means method (cf. Ripley (1987)). To implement this method, divide the Gibbs (output) chain into b batches of length G . Denote the mean of each batch as m_i , and the average of the batches as \bar{m} . Then the standard error of the estimate is given by $\{b(b-1)\}^{-1} \sum_{i=1}^b (m_i - \bar{m})^2$.

However, if a diffuse prior of γ is taken, then the existence of the posterior moments of the θ would be questionable. Geweke (1993) has proven that under certain conditions, these moments do exist. For a more general consideration, in this Chapter, I adopt symmetric linear loss function (see Poirier (1995, p. 302)), and consequently posterior point estimates are the posterior medians of the parameter (vector) θ . The Highest Probability Density (HPD) intervals are used to measure the variability of each estimate. As long as the posterior distribution exists, the posterior medians and HPD intervals always exist.

Remark: The sampling methods involved in the algorithm need further exploitation. The degree of freedom parameter ν can be sampled by using the Rejection-Acceptance sampler as discussed in Geweke (1993). For any specified priors, the threshold parameters can be generated according to the M-H method. i.e.,

- *step 1:* generate δ^+ from a truncated normal distribution with the kernel as

$$\exp \left[-\frac{1}{2} \left(\frac{\delta^+ - \sum_{t \in I_1} M_t \omega_t^{-1} / \sum_{t \in I_1} \omega_t^{-1}}{\sigma(\sum_{t \in I_1} \omega_t^{-1})^{-1/2}} \right)^2 \right] I_{\{\delta^+ \geq \{y_t^*, t \in I_2\}\}} \quad (9)$$

- *step 2:* compute

$$\alpha((\delta^+)^{i-1}, \delta^+) = \min \left\{ \frac{\pi(\delta^+)}{\pi((\delta^+)^{i-1})}, 1 \right\} \cdot$$

- *step 3:* (iii) Generate u from the uniform distribution $U(0,1)$.

- *step 4: (iv)* if $u \leq \alpha((\delta^+)^{i-1}, \delta^+)$, set $(\delta^+)^i = \delta^+$; Otherwise set $(\delta^+)^i = (\delta^+)^{i-1}$.

Similarly, $(\delta^-)^i$ can be generated as well.

6.4 Estimation Results

The estimation programs are written according to the algorithm stated in Section 6 in Fortran-77 and run on a Silicon Graphics 4D-320 UNIX machine with intensive use of IMSL Math/Library and Stat/Library. The iterations of the Markov chain method started with initial values as follows: $\{\beta, \sigma^{-2}\}$ takes OLS estimates. 200 and -500 as the initial values of $\{\delta^+, \delta^-, \nu\} = \{200, -500, 10\}$, and $\omega_t = 1, \forall t = 1, 2, \dots, T$. The change of initial values are insensitive to the final results after first 200 draws are discarded. This is based on the visual inspection of CUMSUM statistics as shown in Figure 2. The message from the figure is that all chains converge after 200 iterations. I have tried 500, 1000, 2000, and 5000 draws from the posterior distribution, and there is little change in the quantities of interest. The final results that follow are based on 2000 draws. The hyperparameters, shown in Table 2, are designed to put not much information in the parameters to be estimated. The prior sensitivity analysis, considered later, includes the case of non-informative case as discussed in Section 4. For the present time, I will concentrate on the estimation and comparative interpretations of all four models. The results are reported in Tables 5-7 with different sample periods.

Firstly, read the results of Model 1 in Tables 5-7, they are basically consistent with those obtained by Feinman (1993). The estimate of β_5 shows a substantial (10 times) difference on emphasis of federal funds rate target before and after the stock market crash in October, 1987. The difference for the estimates of threshold parameter and model standard deviation parameter for the two period are also strong when comparing with that for other parameter estimates (i.e., $\beta_1, \beta_2, \beta_3, \beta_4$).

Secondly, look at the parameter estimates of Model 2 in Tables 5-7. It seems that the

estimated results between Model 2 and Model 1 have a similar pattern. But the estimated values of the parameters for these two models differ substantially. The source of the difference comes from the choices of data distributions. For model 2, a Student-t error is chosen⁷, the degree of freedom parameter is automatically estimated. Model 1 takes a normal error, which is equivalent to Model 2 with a prior constraint $\nu = +\infty$. Based on the estimated value of $\nu \approx 1$, a Cauchy distribution is evident from the data, and thus there is no surprise to see the strong difference between the results delivered from these two models. The adequacy of Model 1 thus becomes questionable. A formal comparison between models will be discussed in the next section. From an economic point of view, a small value of ν represents a strong degree of unobserved heterogeneity in the Desk's behavior. In the literature, this is the first time to report the characteristic of the Desk's reaction behavior in the reserves market. A statistical implication of this discovery is that a normal version of the friction model would deliver distorted inferential results.

The choice between a fat-tailed distribution and a normal distribution has strongly a different effect on parameter estimates when "aberrant" observations exist. In Feinman's analysis, some "outliers" were dropped. These "outliers" have proven to bring very different maximum likelihood estimates for certain parameters from those conducted without the "outliers" (detailed results are not reported). On the other hand, our estimation results with a Student-t version of the friction model are quite stable to the inclusion and exclusion of those "outliers". This is a further evidence for the justification of choosing a fat-tailed data distribution. It should be emphasized that dropping "aberrant" observations is not a healthy practice in an econometric analysis. While some important information contained in these special observations could be thrown out, it becomes more troublesome in time series analysis because the dynamic structure of the data is artificially broken down at these time points. One way of solving the "aberrant" observations, just as have done in this Chapter, is to choose a fat-

⁷As have been proved in Subsection 2.2, a scaled mixture normal with inverted gamma prior for the mixing parameters is equivalent to a Student-t distribution.

tailed data distribution, which properly downweights the "outliers" no matter what (classical or Bayesian) estimation methods are employed.

Thirdly, consider the estimation results of Models 3 and 4. Because the estimated λ 's are quite small in all of Tables 5 - 7, there is no wonder why the results are quite similar to those from Model 2. This implies that after taking account of covariates, there is no much (either observed or latent) dynamics left in the data. This is possibly because the Fed's staff members have done a good job in their daily projections for the open market operations. Thus high order lags of the friction model are not considered since a decreasing pattern for the lagged coefficients is expected.

Fourthly, think of the possible effect of stock market crash of October 1987 on the implementation of monetary policy. The sample is divided into two parts: (1) January 2, 1986 - September 1, 1987 and (2) February 22, 1987 - July 7, 1989. For this case, two major conclusions can be drawn from Tables 6 and 7. There is indeed a shift of the implementation of monetary policy — the Desk's reaction pattern — after the stock market crash. This shift is mainly reflected in the substantial differences between the two periods for the estimates of parameter β_5^8 and threshold parameters δ^+ and δ^- . The Desk has fought much more against the deviation of the federal funds rate from its target after October, 1987. And the operational cost viewed from the Fed seems to be increased dramatically. I have a wider threshold interval for the second period than that for the first period. This is reasonable because the Fed put more effort on closely monitoring the market due to the stock market crash.

The implied histograms for four models in Figures 3 - 6 illustrates intuitively the evidence of model adequacy from the data. The focus here is on the spike at zero. Clearly, Models 2 - 4 are overwhelmingly better than Model 1. The implied histograms are plotted by plugging the estimated parameters and independent variables into relevant models and discarding the

⁸This parameter reflects the "lean against wind" effect of the Fed's monetary policy.

noise terms. The introduction of (both observed and latent) dynamics does not improve the fit, again due to the reason stated above.

Finally, prior sensitivity analysis is conducted. Table 4 provides another specification of the hyperparameters. Table 8 reports the estimation results for the whole sample period with the new set of hyperparameters. Comparing it with Table 4, I conclude that the estimated results are insensitive to the choice of priors. Similar conclusions can be drawn for the two separated periods. Results are not reported to save space. The sensitivity analysis is also conducted with Jaffrey's diffuse prior for the parameter γ . The estimated results are reported in Table 9, and are still fairly close to those reported in Table 4.

7 Model Comparisons

The Desk's behavior has been formulated in competing models in Section 2. To compare them, **Bayes factors** need to be computed. Bayes factors summarize probabilistically the information in the data on how the sample was generated. They have been widely discussed and used in the Bayesian literature. For example, Verdinelli and Wasserman (1995) propose a method based on the Savage-Dickey generalized density ratio. This method is, however, applicable only if the hypotheses are nested. My models are not all nested. Chib (1995) proposes another method based on the Gibbs sampled output (with or without data augmentation). His procedure is simple, but still restrictive because the full conditional densities (not only kernels) must be retained. In my estimation procedure, some of the full conditional distributions are given only in the forms of their kernels. This requires a further development of the computation of Bayes factors for my current models. According to the "Basic Marginal Likelihood Identity" (see Chib (1995))

$$m(Y_T) = \frac{f(Y_T|\theta)\pi(\theta)}{\pi(\theta|Y_T)}$$

Tsionas (1996) extended the previous methods to a more general context. The idea is rather simple. The marginal likelihood is estimated by evaluating the right hand side of this identity at a (any) fix point of θ^* . The posterior density $\pi(\theta|Y_T)$ is estimated *nonparasitically* (see Terrell (1990) or Terrell and Scott (1985)). Actually, an improvement of the estimate of $m(Y_T)$ is possible. This can simply take the (either parametric or nonparametric) estimator of the posterior $\pi(\theta|Y_T)$ as an importance function and then take the Monte Carlo integration

$$m(Y_T) = \int f(Y_T|\theta)\pi(\theta)d\theta = \int \frac{f(Y_T|\theta)\pi(\theta)}{\pi(\theta|Y_T)}\pi(\theta|Y_T)d\theta \approx \frac{1}{M} \sum_{i=1}^M \frac{f(Y_T|\theta^{(i)})\pi(\theta^{(i)})}{\pi(\theta^{(i)}|Y_T)}$$

It is easy to understand that the estimated posterior mimics the “true” posterior. Even if the term $\frac{f(Y_T|\theta)\pi(\theta)}{\pi(\theta|Y_T)}$ may be associated with a large variance, $m(Y_T)$ can be approximated at any accuracy by simply increasing M. This has proven to be extremely easy in practice since M can be increased by running more iterations as in Section 6.

The challenge of computing $m(Y_T)$ also lies in the evaluation of the likelihood function $f(Y_T|\theta)$ at a fixed point, say, $\theta^{(i)}$. This evaluation can be rather simple in Models 1, 2 and 3, but can be burdensome to Model 4 since I need to compute the multivariate Student-t probability in a rectangle. This is the second layer of computational difficulty associated with dynamic friction models (recalling the first layer of difficulty mentioned in the derivation of their likelihood in Section 5). This issue has not been addressed in the existing literature, although several methods for evaluating a multivariate normal probability have been discussed in Hajivasiliou, McFadden and Ruud (1996), Evans and Swartz (1995). Fortunately, at least two of these methods can be generalized to compute a multivariate Student-t probability in a rectangle. One is the GHK (Geweke-Hajivasiliou-Keane) method (see Hajivasiliou, McFadden and Ruud (1996)). Another is the Genz’s (1992) method as mentioned in Evans and Swartz (1995). For a detailed derivation of these two methods, see these references.

For any two models \mathcal{M}_i and \mathcal{M}_j , the Bayes factor is defined as

$$B_{ij} = \frac{m(Y_T|\mathcal{M}_i)}{m(Y_T|\mathcal{M}_j)}$$

The computation results of Bayes factors in logarithm terms are reported in Table 1. The Bayes factors strongly support Model 2 over all other models. Given that Models 3 and 4 are more flexible than Models 1 and 2, the difficulty introduced by the dynamic structures of Models 3 and 4 are penalized in the Bayes factors because of lackness of both dynamic effects after taking account of the effect of covariates. This means that Bayes factors favor simpler models than complicated ones if the simple model is adequate. This is known as one of great advantages of using Bayes factors to make model comparisons.

8 Prediction

Predictions for open market operations are important. To the Desk, predictions help conduct the open market operations. For economic agents, such as Fed watchers, predictions provide a useful tool to forecast the Fed's policy. In addition, predictions can be used to further assess the performances of alternative models. In this section, I first develop a general prediction procedure and then report the prediction results.

8.1 Method

Within a Bayesian framework, prediction is made based on a predictive likelihood, which links the observed quantity and model structure (parameters) to future observations. It takes account of the uncertainty in both unobserved future quantity and the estimated parameters. Although prediction is conceptually easy, its computation is often the major obstacle in many applications. This comes up with the third layer of the difficulties generated from the non-tractable likelihood discussed in Section 5. In this subsection, I will propose a general method

for making one-period-ahead prediction of a dynamic friction model (see Model 4). Multi-period predictions add no further difficulties. As for other models presented in this Chapter, predictions are only special cases of this general exposition.

Suppose I have observations $\{y_1, y_2, \dots, y_T\}$ and parameter (vector) θ . Let $p(\theta|Y_T)$ ($Y_T = \{y_1, y_2, \dots, y_T\}$) is the (estimated) posterior distribution of θ . The predictive likelihood for future observations y_{T+1} is

$$p(y_{T+1}|Y_T) = \int p(\theta|Y_T)L(y_{T+1}|Y_T, \theta)d\theta \quad (10)$$

The equation (10) reflects the Bayesian treatment of nuisance parameters, in this case, θ . Given a loss function, it is conceptually easy to make a point or an interval prediction for future observations.

Now turn to Model 4. For the sake of exposition, without loss of generality ⁹, I assume the last observation y_T is observable, i.e., $y_T \neq 0$. Let the posterior output be

$$\{\theta^{(i)}, z^{(i)}\} = \{\gamma^{(i)}, (\sigma^2)^{(i)}, \omega^{(i)}, \nu^{(i)}, (\delta^+)^{(i)}, (\delta^-)^{(i)}, z^{(i)}\} \quad i = m + 1, m + 2, \dots, m + M.$$

The model structure (Model 4) suggests that in order to predict y_t , I can predict y_t^* instead. Thus the predictive likelihood for y_{T+1}^* is

$$\begin{aligned} p(y_{T+1}|Y_T) &= \int p(\theta|Y_T)p(y_{T+1}^*|Y_T, \theta)d\theta \\ &= \frac{1}{M} \sum_{i=m+1}^{m+M} p(y_{T+1}^*|\theta^{(i)}, Y_T) \\ &= \frac{1}{M} \sum_{i=m+1}^{m+M} p(y_{T+1}^*|\theta^{(i)}, Y_T) \end{aligned}$$

If the prior support of ν is greater than 2 or the estimate of ν is sufficiently larger than 2.

⁹If $y_T = 0$, then I can simply find another y_{T^*} which is the last observable in Y_T . Instead of y_T , simply use y_{T^*} in computing the mean and variance of the $\{y_{T+1}^*, y_{T+2}^*, \dots, y_{T+u}^*\}$ for the following discussion.

prediction is simple. For example, under the quadratic loss, the point prediction of y_{T+1}^* is

$$\begin{aligned}
E(y_{T+1}^*|Y_T) &= \int y_{T+1}^* p(y_{T+1}^*|Y_T) dy_{T+1}^* \\
&= \int y_{T+1}^* \frac{1}{M} \sum_{i=m+1}^{m+M} p(y_{T+1}^*|\theta^{(i)}, Y_T) dy_{T+1}^* \\
&= \frac{1}{M} \sum_{i=m+1}^{m+M} \int y_{T+1}^* p(y_{T+1}^*|\theta^{(i)}, Y_T) dy_{T+1}^* \\
&= \frac{1}{M} \sum_{i=m+1}^{m+M} (Q_{T+1})^{(i)}
\end{aligned}$$

Where $(Q_{T+1})^{(i)}$ could be worked out similarly to (4). In another words, the predictive mean of y_{T+1}^* is the average of the mean of the distribution $p(y_{T+1}|\theta^{(i)}, Y_T)$ over i .

If further one-period-ahead prediction y_{T+2}^* is required, an important question is how to update the posterior $p(\theta|Y_T)$ by the new observation y_{T+1} . First, consider $y_{T+1} \neq 0$, then I have $y_{T+1}^* = y_{T+1} + \hat{\delta}^+$ if $y_{T+1} > 0$, and $y_{T+1}^* = y_{T+1} + \hat{\delta}^-$ if $y_{T+1} < 0$. Then the density function $p(y_{T+1}^*|\theta, Y_T)$, which is bounded up, can serve as an importance function to update the posterior distribution of $p(\theta|Y_T)$ (see Geweke (1995)). Then the predictive density of the y_{T+2}^* will be

$$\begin{aligned}
p(y_{T+2}^*|Y_{T+1}) &= \int p(\theta|Y_{T+1}) p(y_{T+2}^*|Y_{T+1}, \theta) d\theta \\
&= \int p(\theta|Y_T) p(y_{T+1}^*|Y_T, \theta) p(y_{T+2}^*|Y_{T+1}, \theta) d\theta \\
&= \frac{\sum_{i=m+1}^{m+M} w^*(\theta^{(i)}) p(y_{T+2}^*|\theta^{(i)}, Y_T)}{\sum_{i=m+1}^{m+M} w^*(\theta^{(i)})}
\end{aligned} \tag{11}$$

where

$$w^*(\theta) = p(y_{T+1}|\theta, Y_T)$$

If $y_{T+1} = 0$, then (11) still holds with the new weight

$$w^*(\theta) = \int_{\delta^-}^{\delta^+} p(y_{T+1}^*|\theta, Y_T) dy_{T+1}^*$$

Then the further step prediction can be made similarly according to the updated posterior output $\{\theta^{(i)}, w^*(\theta)\}$.

However, in my case, the estimated degree of freedom parameter is less than 2. The existence of the predictive mean and variance are questionable. I can thus still take the

symmetric linear loss structure. Then the point predictor becomes the predictive median, which always exists. Because the median lacks additive property, it can be difficult to compute the median from the predictive likelihood (11). In this case, to predict y_{T+1}^* . I can simply generate a variate $z_{T+1}^{(i)}$ from the distribution of $p(y_T^*|\theta^{(i)}, y_T)$. Then similar to the Markov Chain method, these variate can be thought as an i.i.d. sample approximately from the predictive distribution of $p(y_{T+1}^*|Y_T)$. The median of this sample is the approximation of the predictive median of y_{T+1}^* .

To compute the predicted value of y_{T+2}^* , I need again to update the posterior output $\theta^{(i)}$. However, the computation of the predictive median for y_t^* in this case is different from that in predicting y_{T+1} . This is because I have the updated posterior output $\{\theta^{(i)}, w^*(\theta)\}$. The way of efficiently utilizing this posterior information is the following. Generate a variate $z_{T+2}^{(i)}$ from the distribution $p(y_{T+2}|\theta^{(i)}, Y_{T+1})$, and then sort the generated sample in increasing order, with the corresponding sort of the weight $w^*(\theta)$. This weight is just like a distribution on the sorted sample $z_{T+2}^{(i)}$. Thus, the median of the sample with the “weight” distribution, which can be easily implemented in computer, is the approximation of the predictive median. Further step predictions are similar. Again, the HPD intervals are computed to measure the variability of the predictions.

The importance sampling Markov Chain updating method provides a fast computation method. This procedure requires no reestimation of the parameters for each prediction step. The density function $p(y_{T+1}^*|\theta, Y_{T+1})$ is the exact weight function (or importance function) to transform the simulated Monte Carlo sample from $p(\theta|Y_T)$ to a simulated Monte Carlo sample from $p(\theta|Y_{T+1})$. Thus this builds up an easy recursion for the computation of the one-period-ahead predictive density function $p(y_{T+\tau}^*|y_{T+\tau-1})$. As long as the importance sampling weights

$$\prod_{\tau=1}^r p(y_{T+\tau}^*|\theta^{(i)}, y_{T+\tau-1}) \quad (i = m + 1, m + 2, \dots, m + M)$$

remained well behaved, the recursion remains practical. It is generally more efficient to rees-

timate the parameters after certain prediction steps (see Geweke (1994)).

8.2 Results

To make predictions, first note that the estimation results reported in Section 6 show substantially different patterns during the two periods before and after October 1987. Thus, the sample is again separated into two periods accordingly for predictions. For comparison purpose, 30 day one-period-ahead predictions were made according to the method developed above. First, we estimate the Desk's reaction pattern for the period spanning from Jan 2, 1986 to July 31, 1987, and then predicting for the period from August 3, 1987 to September 14, 1987. Second, I estimate the Desk's reaction pattern for the period spanning from Feb. 2, 1987 to May 24, 1989, and then predicting for the next 30 days. The real observed values of y_t for the same prediction periods are held-out.. The predicted median and 95% HPD intervals are computed for each of four models. These results are associated with the informative prior (see Section 4) with the hyperparameters I (see Table 2). The predictions and the logarithms of predictive marginal likelihoods are graphed in Figures 7 and 8.

The major conclusions are summarized as follows:

1. The point predictions made by the four models have a roughly similar pattern. For the first period, it seems that the Desk's reactions are more volatile and thus the predictions are less accurate for all four models. This is probably because certain unpredictable shocks hit the market before the stock market crash of October 1987.

2. The estimated HPD predictive intervals appear strongly asymmetric around the predictive medians for Models 2, 3 and 4. If $y_t > 0$, the predicted y_t is largely skewed to the right. As $y_t < 0$, the predicted y_t tends to be skewed to the left. This effect is possibly due to the choice of a Student-t error structure, which attempts to capture the tailed observations in the data distribution. However, Model 1 does not possess this property.

3. Among 30 day predictions, for the first period, 16 day predictions for Model 1 lie on the outside of their 95% HPD intervals, and only 4 or 5 day predictions for Models 2 - 4 lie on the outside of their 95% HPD intervals; there is a similar picture for the second period. This implies that Model 1 systematically underestimates the uncertainty of the predictions

4. The average of sum squares between actual and predicted for Models 1 - 4 are 510.68, 533.89, 484.16 and 485.40, respectively, for the first period, and are 470.21, 474.02, 454.43, and 468.95, respectively, for the second period. The average of absolute difference between actual and predicted for Models 1 - 4 are 226.42, 221.66, 236.71, and 236.42, respectively, for the first period, and are 356.05, 347.09, 328.34, and 348.32, respectively, for the second period. Clearly, for the first period, Models 3 and 4 perform better than Models 1 and 2 in terms of both loss structures. For the second period, the results are mixed.

9 Conclusion

This study has modeled the open market Desk's intervention process in the Fed's reserves market within a Bayesian econometric context. The characteristics of the Desk's reactions are assessed through developing a dynamic friction model with a fat-tailed distribution. This model integrates three possible features of the data, namely censoring, unobserved heterogeneity and dynamics, into one specification. The development of major inferential procedures to the general model consists of the primary statistical contribution of this Chapter.

The major difficulty arising from a dynamic friction model has been discussed in three layers of making inferences: estimation, model comparison and prediction. It is originated from high dimensional integrals, induced by correlations among censored observations, in a likelihood function. The estimation procedure is proposed based on an idea developed in Wei (1995). The computation of a marginal likelihood, which is the key to Bayes factors, is discussed, relying on the recent papers by Chib (1995), Tsionas (1996).

Our examination concludes that a normal version of the friction model as in Feinman (1993) no longer properly describes the Desk's daily behavior. The data overwhelmingly support a strong degree of unobserved heterogeneity involved in the Desk's reactions. More precisely, a Cauchy distribution (since $\nu \approx 1$) has been identified for the data distribution. From an economic point of view, unobserved heterogeneity reflects: (1) the discretionary responses to shocks transmitted from various aspects of the economy; (2) an incomplete knowledge of the Fed's understanding of the real economy.

The inclusion of (both observed and latent) dynamics seems to have not improved the performances of our friction model with a fat-tailed data distribution after taking account of the covariates. This is because the dynamic effects involved in the Desk's operations have been possibly well captured in the projections made by the Fed's staff members. Although the Fed's projections are made based on different models ¹⁰ from the ones developed here, dynamic effects may be quite similar. The Bayes factors computed in Section 7 punish substantially the complicatedness of Models 3 and 4.

The estimated models are used to detect the possible shift of the implementation of the monetary policy after the stock market crash in Oct, 1987. There are substantial differences between the two periods. After the market crash, the Desk pay much more attention on the Fed's funds market (larger value of β_5) conditions, and consequently the operational cost may increase and has thus widen the threshold of the Desk's reactions.

My predictive results show that it is indeed hard to make a good prediction since the Desk often reacts discretionarily. The mixing parameter ω_t is the source of generating large predicted errors for future observations. One important feature for our in-sample one-step-ahead predictions is that the predictive likelihoods are strongly skewedly distributed. It is likely that when a future observation is positive, its predictive likelihood is skewed strongly

¹⁰Though there are some discussions on the Fed's projection models, no exact specification have been reported from either inside or outside the Fed.

to the right, and when a future observation is negative, its predictive likelihood is skewed strongly to the left.

It might be interesting to indicate that the method developed in this Chapter can be applied to a broad class of models which have likelihood functions involved in high dimensional integrals. For example, dynamic Tobit models, multivariate Tobit models, multivariate probit models, etc. A general treatment of these models deserves further research.

Table 1. Sample Statistics (Sample Size = 879)

Variable	Mean	Std. Dev.	Min.	Max.	Ske.	Kurt.
y	233.1	630.5	-1696.0	6512.0	3.6	24.74
x_1	135.4	321.8	-915.0	2941.0		
x_2	653.1	2246.0	-8847.0	14770.0		
x_3	261.2	1427.7	-6801.0	10130.0		
x_4	-102.5	1112.5	-8421.0	4887.0		
x_5	.088	.503	-1.25	10.7		

Table 2. The Designed Hyperparameters I

Hyperpara.	Model 1	Model 2'	Model 3'	Model 4'
$\underline{\beta}_1$	2.0	2.0	2.0	2.0
$\underline{\beta}_2$.2	.2	.2	.2
$\underline{\beta}_3$.2	.2	.2	.2
$\underline{\beta}_4$.2	.2	.2	.2
$\underline{\beta}_5$	500.0	500.0	500.0	500.0
$\underline{\lambda}$	—	—	.2	.2
$\underline{\nu}$	—	10.0	10.0	10.0
$\underline{s^2}$	—	.05	.05	.05
\underline{diagQ}	—	(10,10,10,10,400)	(10,10,10,10,400,10)	(10,10,10,10,400,10)
α	—	1.0	1.0	1.0

Table 3. The Designed Hyperparameters II

Hyperpara.	Model 1	Model 2'	Model 3'	Model 4'
$\underline{\beta}_1$.2	.2	.2	.2
$\underline{\beta}_2$.01	.01	.01	.01
$\underline{\beta}_3$.01	.01	.01	.01
$\underline{\beta}_4$.01	.01	.01	.01
$\underline{\beta}_5$	150.0	150.0	150.0	150.0
$\underline{\lambda}$	—	—	.02	.02
$\underline{\nu}$	—	100.0	100.0	100.0
$\underline{s^2}$	—	.005	.005	.005
\underline{diagQ}	—	(20,20,20,20,600)	(20,20,20,20,600,20)	(20,20,20,20,600,20)
α	—	.1	.1	.1

Table 4. Estimation Results with Prior I

Para. ↓	Model 1		Model 2		Model 3		Model 4	
	Med.	95% HPDI	Med.	95% HPDI	Med.	95% HPDI	Med.	95% HPDI
β_1	1.07	(.95, 1.18)	.71	(.63, .78)	.70	(.62, .78)	.67	(.59, .75)
β_2	.165	(.150, .180)	.052	(.043, .065)	.055	(.042, .069)	.052	(.040, .066)
β_3	.060	(.037, .082)	.028	(.016, .041)	.031	(.017, .044)	.029	(.016, .044)
β_4	.080	(.050, .109)	.063	(.044, .081)	.063	(.044, .085)	.058	(.040, .078)
β_5	188	(123, 252)	532	(444, 628)	536	(446, 627)	501	(415, 607)
λ	—	—	—	—	.027	(-.009, .073)	.041	(.006, .082)
σ	475	(450, 504)	122	(108, 143)	125	(108, 148)	122	(105, 142)
ν	—	—	1.2	(1.1, 1.4)	1.3	(1.1, 1.5)	1.2	(1.1, 1.5)
δ^+	184	(138, 231)	57	(42, 76)	64	(41, 89)	56	(32, 84)
δ^-	-506	(-570, -442)	-219	(-260, -197)	-229	(-274, -191)	-219	(-254, -186)

Note: Sample size = 879: January 2, 1986 to July 7, 1989.

Table 5. Estimation Results with Prior I

Para. ↓	Model 1		Model 2		Model 3		Model 4	
	Med.	95% HPDI	Med.	95% HPDI	Med.	95% HPDI	Med.	95% HPDI
β_1	1.13	(.97, 1.30)	.62	(.54, .73)	.59	(.49, .69)	.61	(.51, .75)
β_2	.15	(.13, .18)	.040	(.027, .058)	.040	(.027, .058)	.043	(.029, .071)
β_3	.053	(.022, .085)	.022	(.011, .035)	.022	(.012, .033)	.023	(.012, .037)
β_4	.093	(.046, .135)	.065	(.042, .087)	.062	(.041, .083)	.062	(.040, .088)
β_5	140	(83, 193)	342	(271, 439)	323	(255, 409)	328	(262, 429)
λ	—	—	—	—	.028	(-.004, .071)	.043	(.010, .085)
σ	404	(376, 435)	81	(66, 104)	78	(63, 97)	79	(63, 105)
ν	—	—	1.0	(.8, 1.3)	1.0	(.8, 1.2)	1.0	(.8, 1.2)
δ^+	146	(88, 222)	13	(0, 36)	13	(0, 37)	22	(0, 55)
δ^-	-418	(-508, -332)	-166	(-203, -136)	-159	(-197, -123)	-157	(-198, -123)

Note: Sample Size = 417: January 2, 1986 to September 1, 1987.

Table 6. Estimation Results with Prior I

Para. ↓	Model 1		Model 2		Model 3		Model 4	
	Med.	95% HPDI	Med.	95% HPDI	Med.	95% HPDI	Med.	95% HPDI
β_1	1.00	(.81, 1.17)	.75	(.66, .84)	.74	(.65, .83)	.71	(.60, .81)
β_2	.19	(.17, .21)	.047	(.032, .064)	.045	(.031, .059)	.046	(.032, .062)
β_3	.067	(.028, .107)	.018	(.002, .037)	.019	(.002, .035)	.018	(.002, .037)
β_4	.068	(.028, .112)	.055	(.032, .081)	.054	(.032, .079)	.054	(.033, .079)
β_5	1402	(1059, 1757)	836	(722, 964)	818	(697, 942)	816	(699, 965)
λ	—	—	—	—	.015	(-.018, .057)	.015	(-.016, .062)
σ	482	(443, 524)	97	(76, 123)	93	(75, 117)	95	(75, 122)
ν	—	—	1.1	(.9, 1.3)	1.1	(.9, 1.4)	1.1	(.8, 1.3)
δ^+	367	(295, 453)	108	(76, 133)	98	(68, 134)	100	(73, 146)
δ^-	-494	(-599, -393)	-226	(-271, -185)	-216	(-268, -173)	-219	(-262, -177)

Note: Sample size is 345 and sample period is from Feb. 22, 1988 to July 7, 1987.

Table 7. Estimation Results with Prior II (Sensitivity Analysis)

Hyperpara. Para. ↓	Model 1		Model 2		Model 3		Model 4	
	Med.	95% HPDI	Med.	95% HPDI	Med.	95% HPDI	Med.	95% HPDI
β_1	1.05	(.95, 1.15)	.70	(.69, .88)	.69	(.61, .77)	.68	(.60, .77)
β_2	.165	(.150, .181)	.051	(.065, .097)	.052	(.041, .066)	.051	(.040, .065)
β_3	.060	(.036, .080)	.027	(.032, .065)	.029	(.017, .041)	.028	(.016, .041)
β_4	.080	(.050, .109)	.063	(.050, .095)	.062	(.043, .084)	.058	(.039, .080)
β_5	187	(125, 253)	534	(444, 628)	526	(455, 647)	502	(413, 597)
λ	—	—	—	—	.024	(-.007, .061)	.037	(.007, .078)
σ	480	(450, 509)	125	(107, 146)	124	(106, 148)	129	(112, 149)
ν	—	—	1.2	(1.0, 1.4)	1.2	(1.0, 1.4)	1.2	(1.0, 1.4)
δ^+	183	(146, 242)	59	(40, 87)	58	(40, 79)	59	(36, 80)
δ^-	-520	(-587, -454)	-228	(-264, -196)	-224	(-264, -192)	-220	(-261, -186)

Note: Sample size is 879 and sample period is from January 2, 1986 to July 7, 1987.

Table 8. Estimation Results with Partly Diffused Prior

Para. ↓	Model 1		Model 2		Model 3		Model 4	
	Med.	95% HPDI	Med.	95% HPDI	Med.	95% HPDI	Med.	95% HPDI
β_1	1.06	(.94, 1.18)	.72	(.65, .81)	.69	(.59, .77)	.69	(.60, .77)
β_2	.165	(.150, .182)	.056	(.044, .070)	.054	(.042, .068)	.058	(.045, .074)
β_3	.060	(.039, .083)	.031	(.019, .044)	.031	(.017, .042)	.033	(.020, .047)
β_4	.079	(.052, .110)	.065	(.046, .085)	.065	(.050, .080)	.063	(.043, .083)
β_5	190	(122, 252)	547	(472, 649)	542	(422, 612)	503	(443, 627)
λ	—	—	—	—	.024	(-.007, .070)	.050	(.012, .089)
σ	482	(456, 509)	131	(117, 154)	131	(113, 160)	132	(119, 165)
ν	—	—	1.3	(1.1, 1.6)	1.3	(1.1, 1.5)	1.3	(1.1, 1.6)
δ^+	186	(140, 232)	64	(46, 86)	60	(37, 71)	51	(42, 88)
δ^-	-513	(-591, -442)	-228	(-273, -198)	-235	(-278, -209)	-228	(-278, -202)

Note: Sample size is 879 and sample period is from January 2, 1986 to July 7, 1989.

Table 9. Logarithms of Bayes Factors

Hypotheses	Whole Period	Period I	Period II
B_{21}	28.78	167.39	12.73
B_{32}	-42.94	-143.68	-20.25
B_{43}	0.5	36.36	10.90

Table 10. Prediction For First Period

<i>t</i>	<i>y_t</i>	Model 1		Model 2		Model 3		Model 4	
		Med.	95% HPDI	Med.	95% HPDI	Med.	95% HPDI	Med.	95% HPDI
1	185	179	(80,222)	247	[0, 1089)	265	(11, 1378)	256	[0, 1283)
2	133	277	(234,327)	284	[0, 1185)	283	(15, 1376)	293	(9, 1487)
3	122	180	(148,216)	98	[0, 730)	155	[0, 868)	147	[0, 969)
4	370	409	(368,451)	352	[0, 1481)	361	(39, 2027)	344	(17, 1675)
5	0	0	[0]	0	[0]	20	[0, 109)	0	[0, 119)
6	89	9	[0, 58)	0	[0, 165)	55	[0, 313)	73	[0, 482)
7	0	0	(27, 94)	0	[0, 165)	0	(-58, 0]	0	[0, 32)
8	0	0	[0]	0	[0,5)	18	[0, 177)	11	[0, 183)
9	25	0	[0, 15)	0	[0, 412)	121	[0, 657)	101	[0, 629)
10	532	418	(370, 468)	524	[0, 1738)	456	(66, 2414)	451	(49, 2165)
11	100	0	[0]	0	[0]	0	[0, 9)	0	[0, 59)
12	106	147	(124, 170)	99	[0, 640)	171	[0, 832)	166	[0, 846)
13	89	277	(241,313)	172	[0, 909)	231	[0, 1198)	244	[0, 1171)
14	0	307	(271, 341)	182	[0, 971)	272	[21, 1246)	285	(13, 1502)
15	115	124	(80,170)	165	[0, 806)	238	(28, 1187)	241	(46, 1302)
16	151	98	(65, 130)	162	[0, 887)	242	(13, 1258)	244	(15, 1308)
17	325	373	(314, 433)	613	(10, 2278)	582	(71, 3000)	627	(83, 3098)
18	230	368	(329, 409)	364	[0, 1384)	316	(16, 1887)	316	(20, 1605)
19	286	165	(102,234)	134	[0, 782)	184	(7, 1199)	195	[0, 972)
20	128	193	(145,234)	172	[0, 1137)	213	(15, 1130)	217	(1, 1243)
21	451	517	(453, 587)	392	0, 1739)	304	(30, 1579)	335	(28, 1818)
22	1330	472	(427, 520)	312	[0, 1528)	356	(29, 1731)	400	(27, 1878)
23	60	205	(178, 234)	88	[0, 661)	251	(2, 1253)	290	[0, 1470)
24	13	252	(210, 295)	0	[0, 175)	121	[0, 684)	110	[0, 790)
25	16	0	[0]	0	(-164, 0]	0	(-223, 0]	0	(-172, 0]
26	130	0	[0]	205	[0, 1033)	245	(18, 1370)	239	[0, 1252)
27	77	125	(53, 200)	32	[0, 636)	243	(64, 2983)	248	(0, 1136)
28	91	1418	(1166, 1661)	1401	(165, 4328)	1136	(204, 5126)	1078	(167, 5332)
29	2536	125	[0, 239)	18	[0, 689)	248	[0, 1415)	206	[0, 1277)
30	147	351	(269, 433)	460	[0, 1560)	549	(55, 2678)	569	(67, 2761)

Note: Predictive period is from August 3, 1987 to September 14, 1987

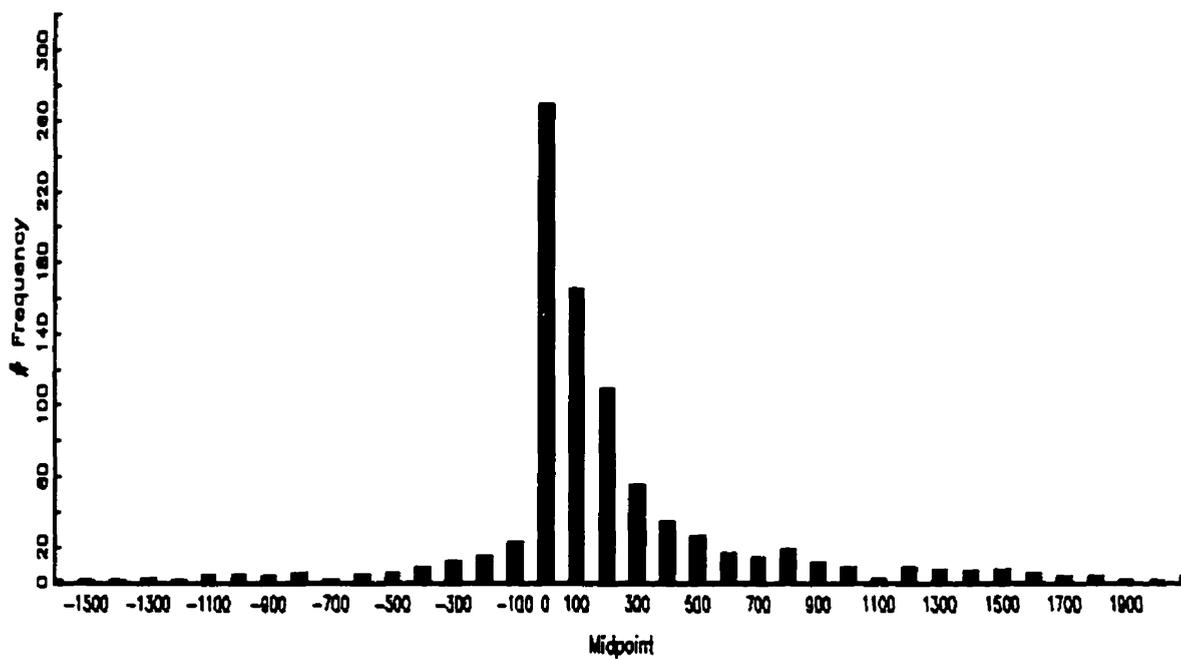
Table 11. Prediction For Second Period

t	y_t	Model 1		Model 2		Model 3		Model 4	
		Med.	95% HPDI	Med.	95% HPDI	Med.	95% HPDI	Med.	95% HPDI
1	-1326	-91	(-148, -33)	-192	(-2588, 0]	-249	(-2317, 0]	-264	(-2812, 0]
2	-9	-78	(-163, 0]	-413	(-4138, 0]	-453	(-3651, 0]	-424	(-3740, 0]
3	-1054	-376	(-565, -182)	-566	(-3919, 0]	-563	(-4359, 0]	-568	(-3818, 0]
4	145	246	(105, 383)	363	[0, 2492)	380	[0, 2283)	351	(0, 2324)
5	-127	-828	(-949, -695)	-278	(-3084, 0]	-343	(2760, 0]	-336	(-2448, 0]
6	-1077	-127	(-1513, -999)	-1357	(-9653, 0]	-1228	(-8168, -9)	-1466	(-9403, -65)
7	-882	-256	(-484, -144)	-428	(-3750, 0]	-425	(-2852, 0]	-464	(-3193, 0]
8	0	0	[0]	-16	(-1637, 0]	-36	(-1192, 0]	-37	(1438, 0]
9	-580	0	[0]	-38	(-1629, 0]	-66	(-1252, 0]	-65	(1353, 0]
10	-754	-132	(-188, -72)	-165	(-2564, 0]	-221	(-2182, 0]	-226	(-2248, 0]
11	-298	0	[0]	-383	(3428, 0]	-379	(-2759, 0]	-350	(-3093, 0]
12	-553	-74	(-212, 0]	-216	(-2317, 0]	-211	(-1985, 0]	-308	(2697, 0]
13	-177	0	(-22, 0]	-147	(-2455, 0]	-119	(-1717, 0]	-141	(-1905, 0]
14	0	0	[0]	0	[0, 24)	0	[0, 15)	0	[0, 7)
15	-1678	-1311	(-1519, -1100)	-234	(-2323, 0]	-284	(-2371, 0]	-274	(-2877, 0]
16	-716	0	[0]	114	[0, 1323)	127	(-387, 1313)	133	[0, 1020)
17	-214	0	[0]	0	[0, 183)	0	(-24, 284)	0	[0, 235)
18	-239	0	(-34, 0]	0	(-790, 0]	-12	(829, 0]	-33	(-800, 0]
19	0	743	(626, 858)	304	[0, 2011)	384	(4, 2256)	296	(13, 1301)
20	1205	754	(681, 831)	568	(26, 2883)	579	(43, 3353)	452	(19, 3611)
21	368	161	(54, 258)	562	(46, 3192)	582	(51, 2916)	715	(43, 3746)
22	0	0	[0, 6)	0	[0, 128)	0	[0, 153)	0	[0, 195)
23	0	8	[0, 149)	91	[0, 982)	135	[0, 891)	95	[0, 822)
24	124	101	[0, 238)	120	[0, 1141)	193	[0, 1269)	233	[0, 1597)
25	261	897	(752, 1064)	149	(-83, 696)	151	[0, 1050)	120	(-461, 991)
26	320	0	[0]	0	(-618, 66)	0	(-651, 259)	0	(-686, 50)
27	-906	-917	(-1028, -807)	-656	(-4468, 0]	-644	(-5074, 0]	-699	(-5332, 0]
28	-998	-230	(-296, -159)	-172	(-2739, 0]	-556	(-2617, 0]	-396	(-2460, 748)
29	-38	0	[0]	0	(-1325, 0]	-30	(-1056, 0]	-317	(-1132, 0]
30	-479	-296	(-388, -200)	-625	(-3915, 0]	-511	(-2715, 0]	-382	(-4124, -44)

Note: Predictive period is from May 25, 1989 to July 7, 1989

Figure 1

Histogram of y (with foreign account)



Time Series of y (with foreign account)

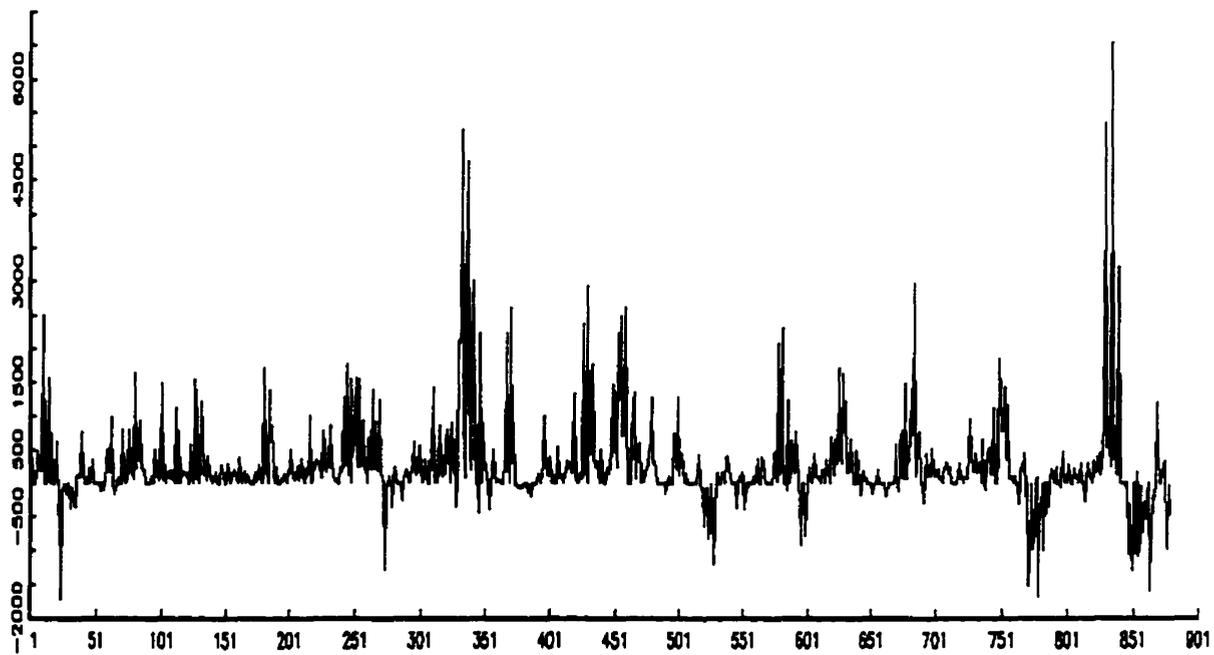


Figure 2: CUMSUM Plots

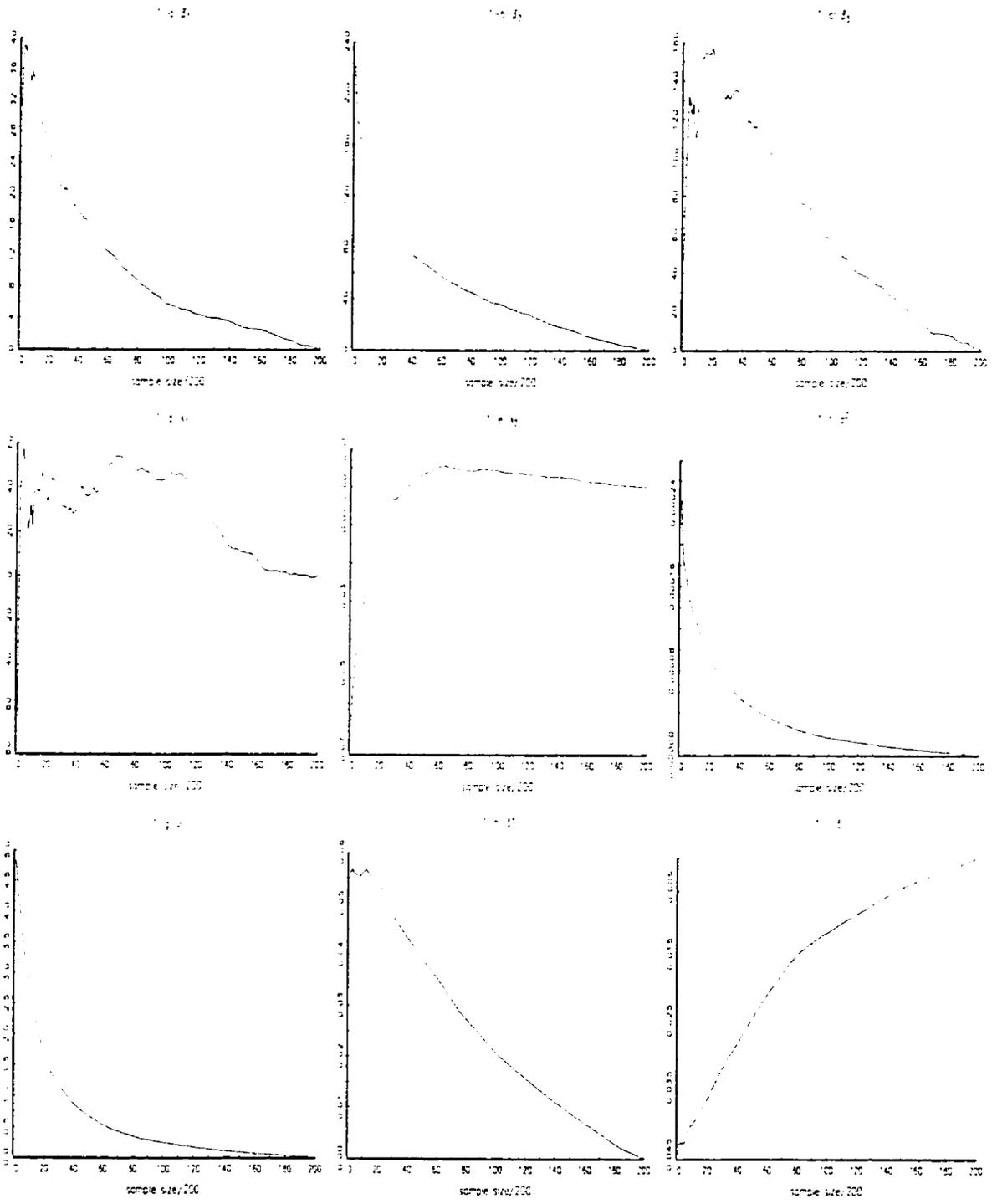
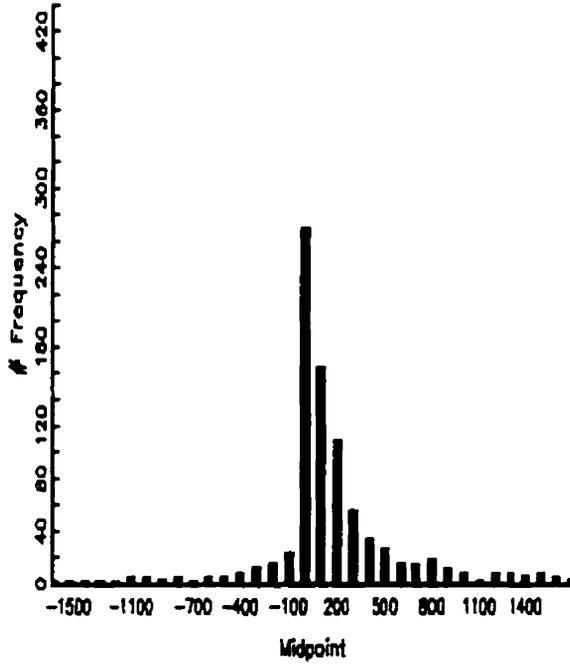
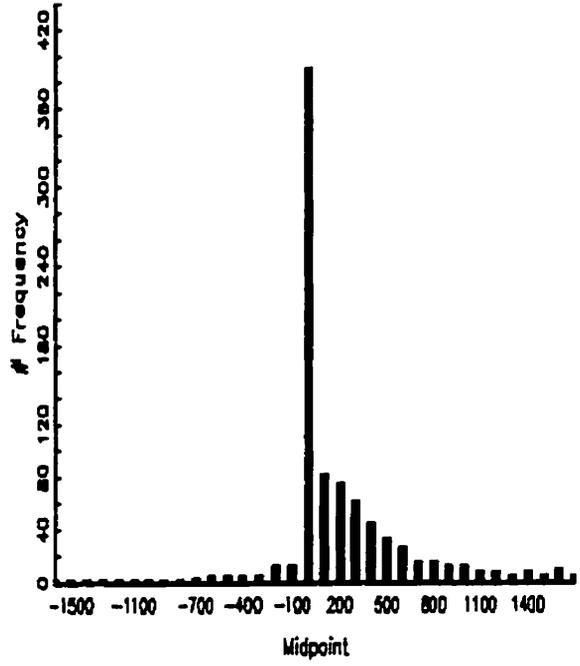


Figure 3: Implied Histogram of Model 1

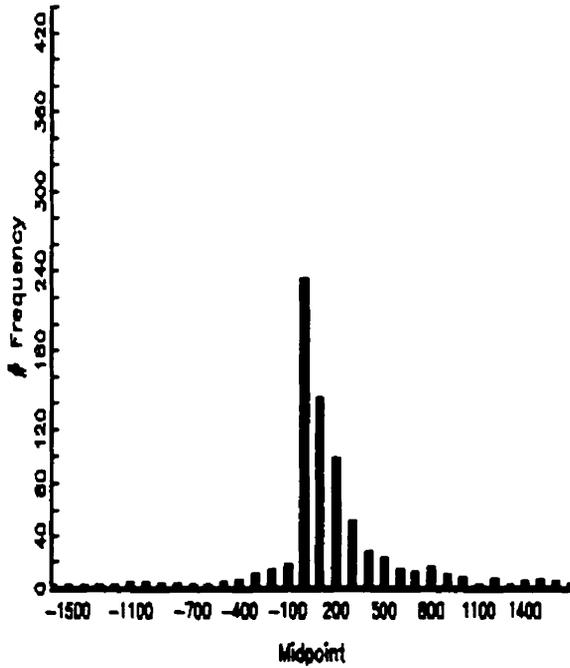
Histogram of Y (Real Data + One Piece)



Implied Histogram of Y (One Piece)



Histogram of Y (Real Data + Two Piece)



Implied Histogram of Y (Two Piece)

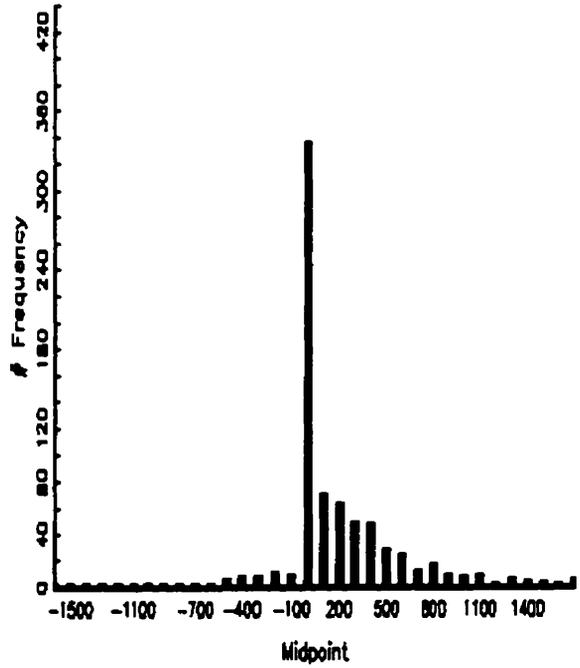
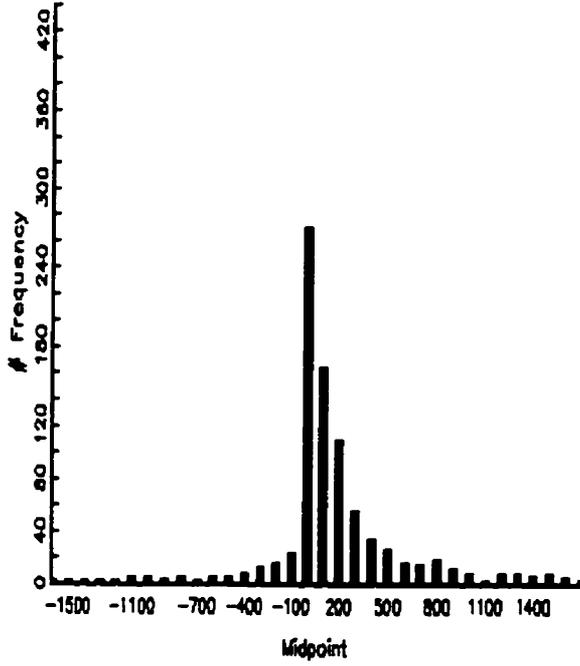
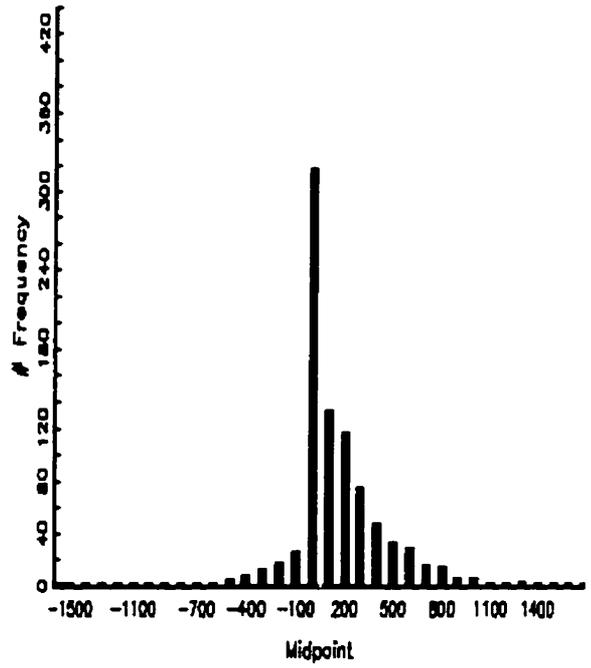


Figure 4: Implied Histogram of Model 2

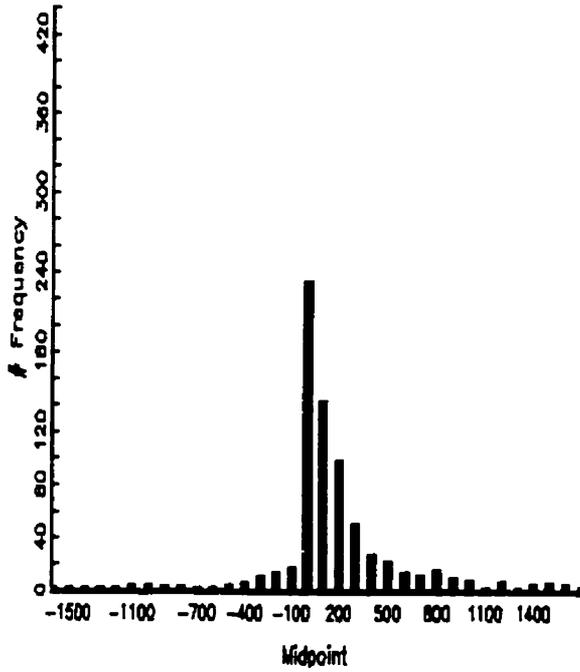
Histogram of Y (Real Data + One Piece)



Implied Histogram of Y (One Piece)



Histogram of Y (Real Data + Two Piece)



Implied Histogram of Y (Two Piece)

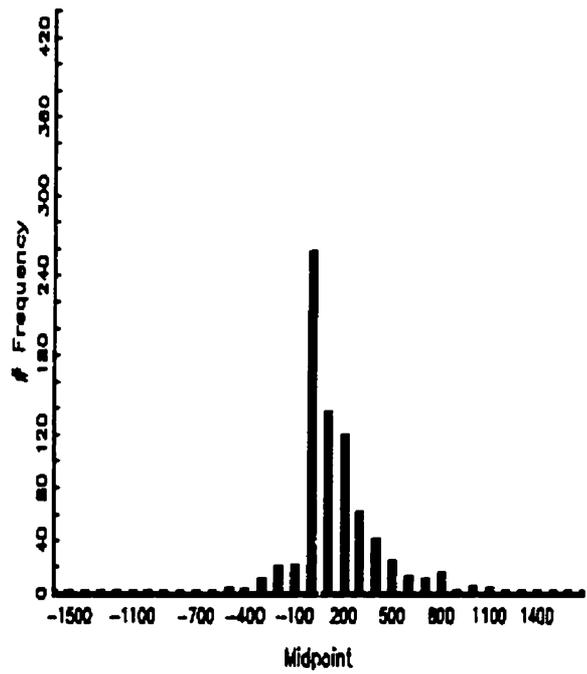
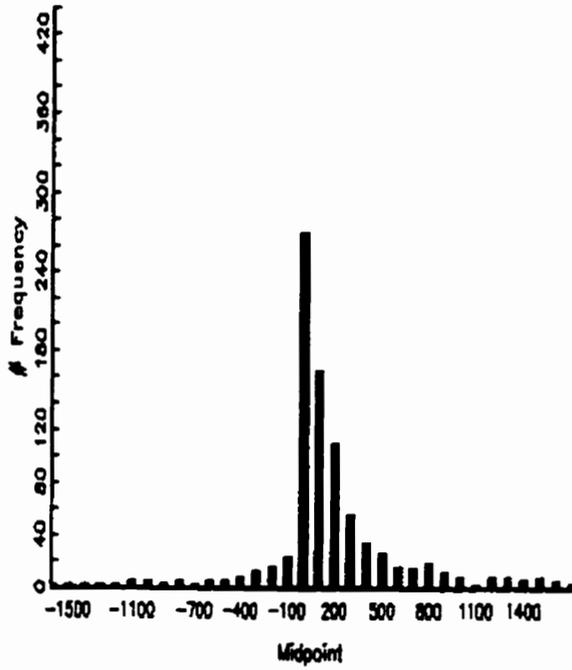
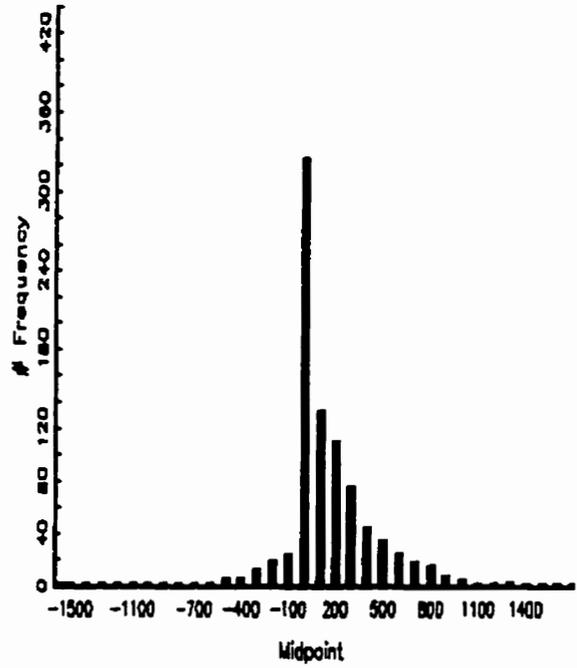


Figure 5: Implied Histogram of Model 3

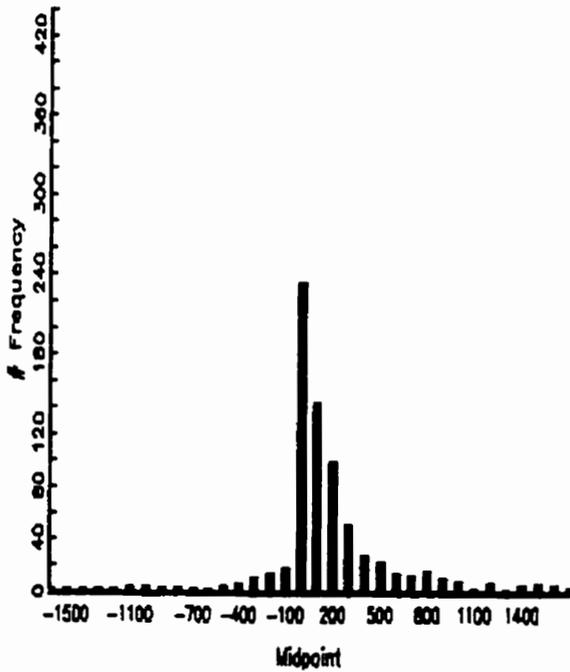
Histogram of Y (Real Data + One Piece)



Implied Histogram of Y (One Piece)



Histogram of Y (Real Data + Two Piece)



Implied Histogram of Y (Two Piece)

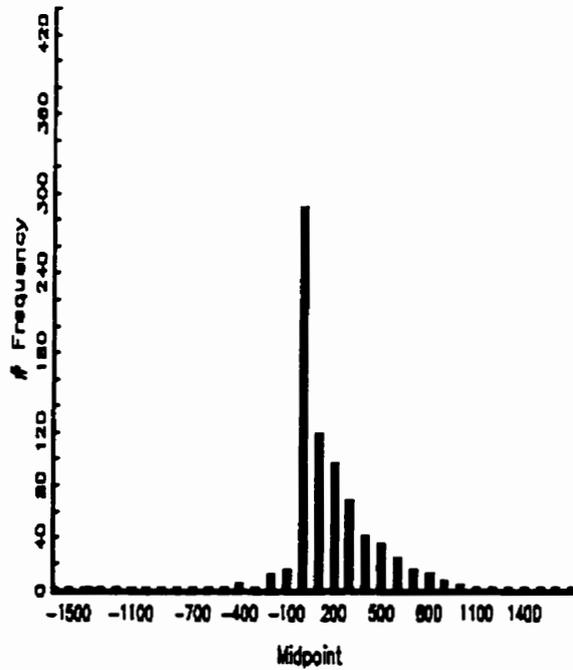
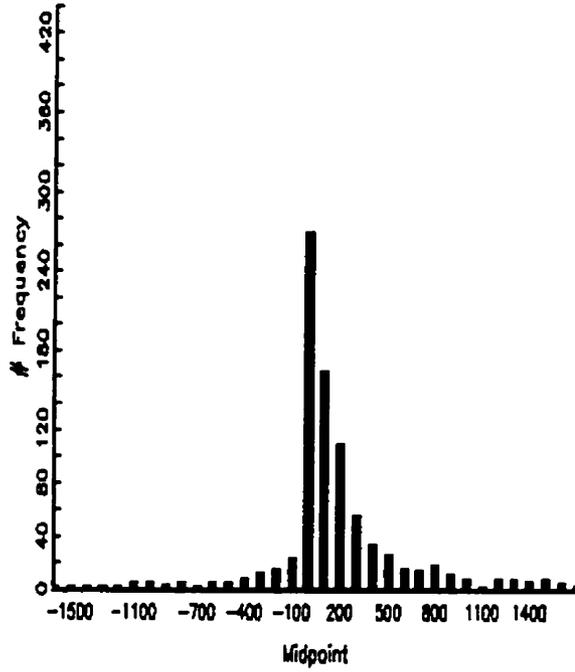
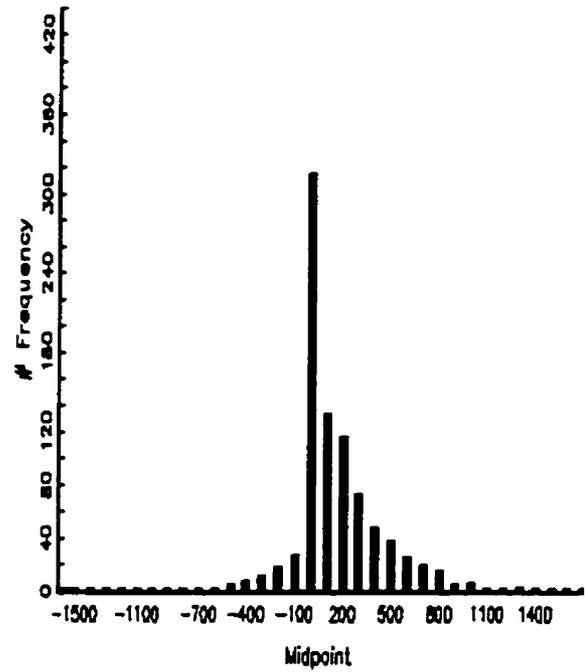


Figure 6: Implied Histogram of Model 4

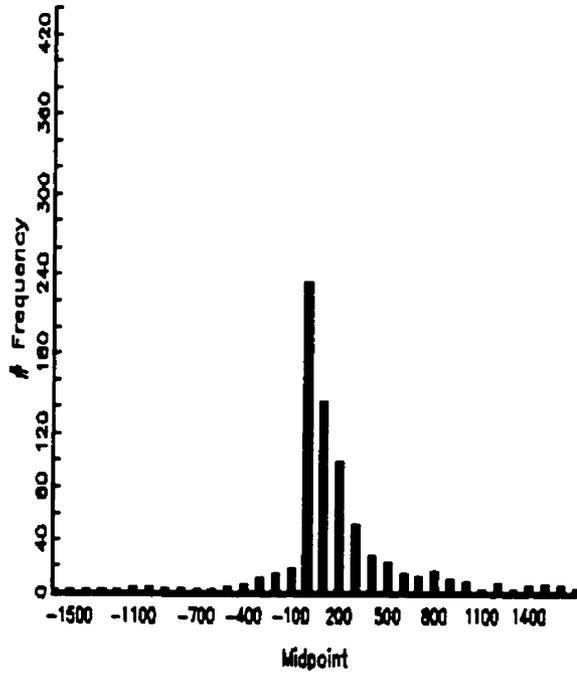
Histogram of Y (Real Data + One Piece)



Implied Histogram of Y (One Piece)



Histogram of Y (Real Data + Two Piece)



Implied Histogram of Y (Two Piece)

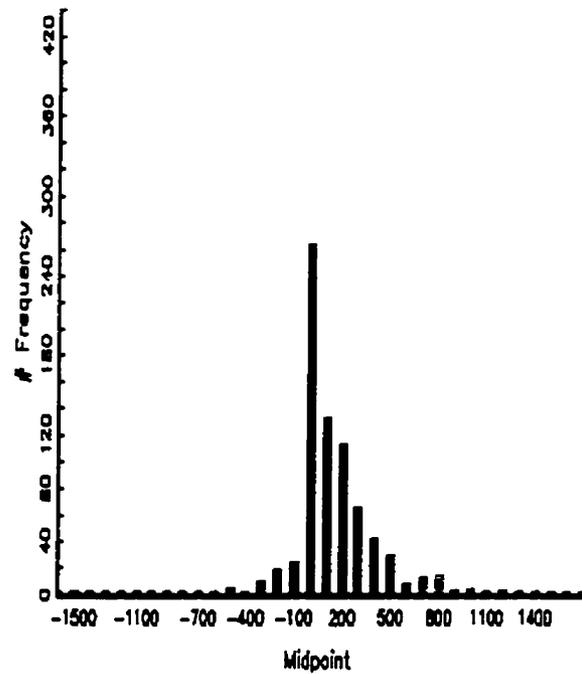
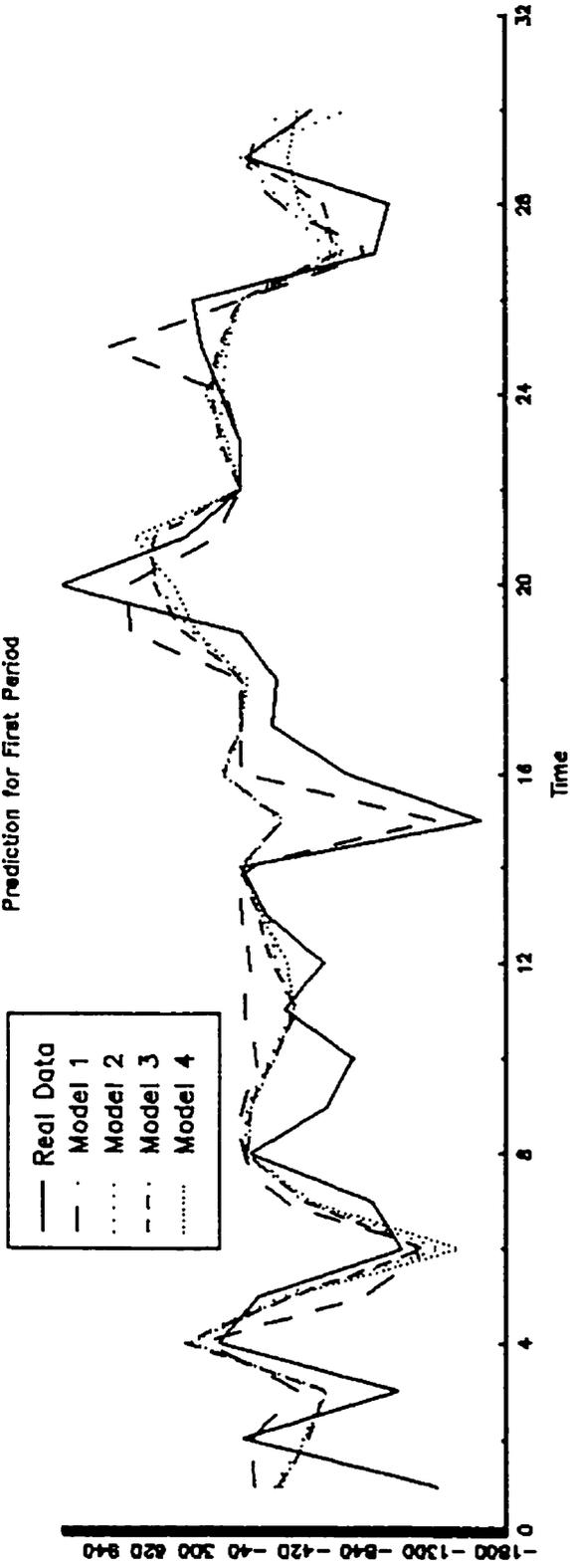


Figure 7

Prediction for First Period



Prediction for Second Period

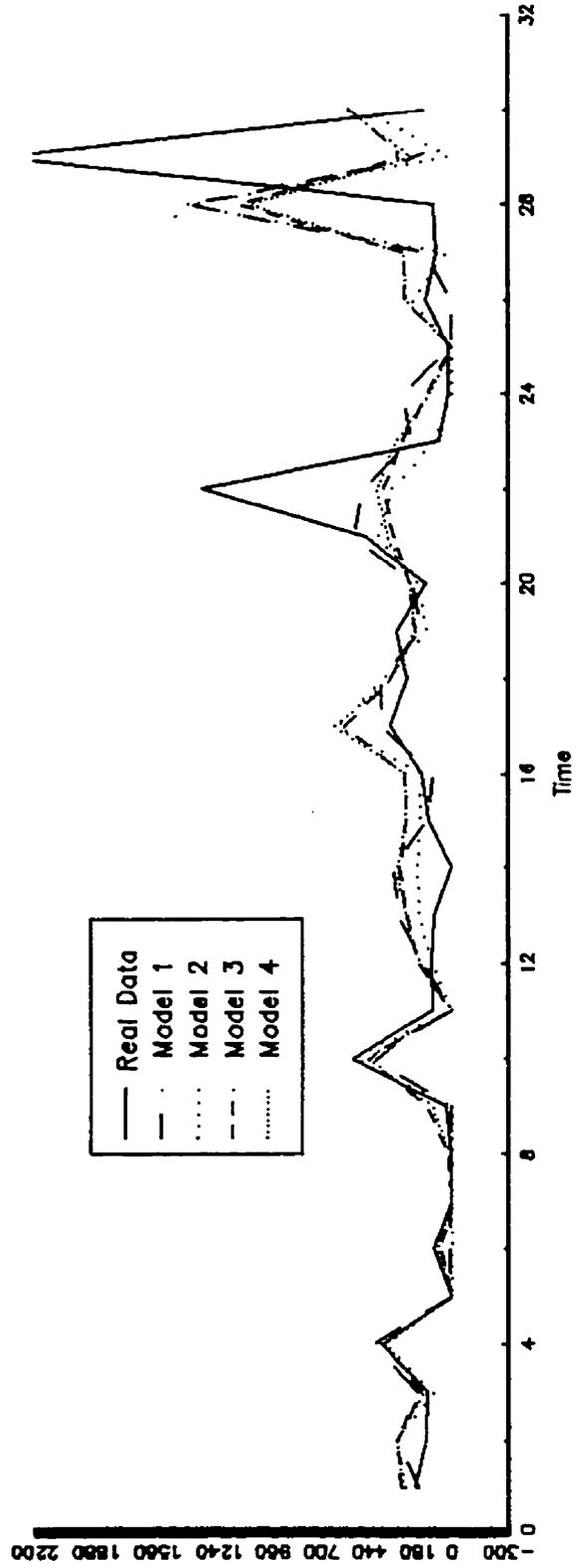
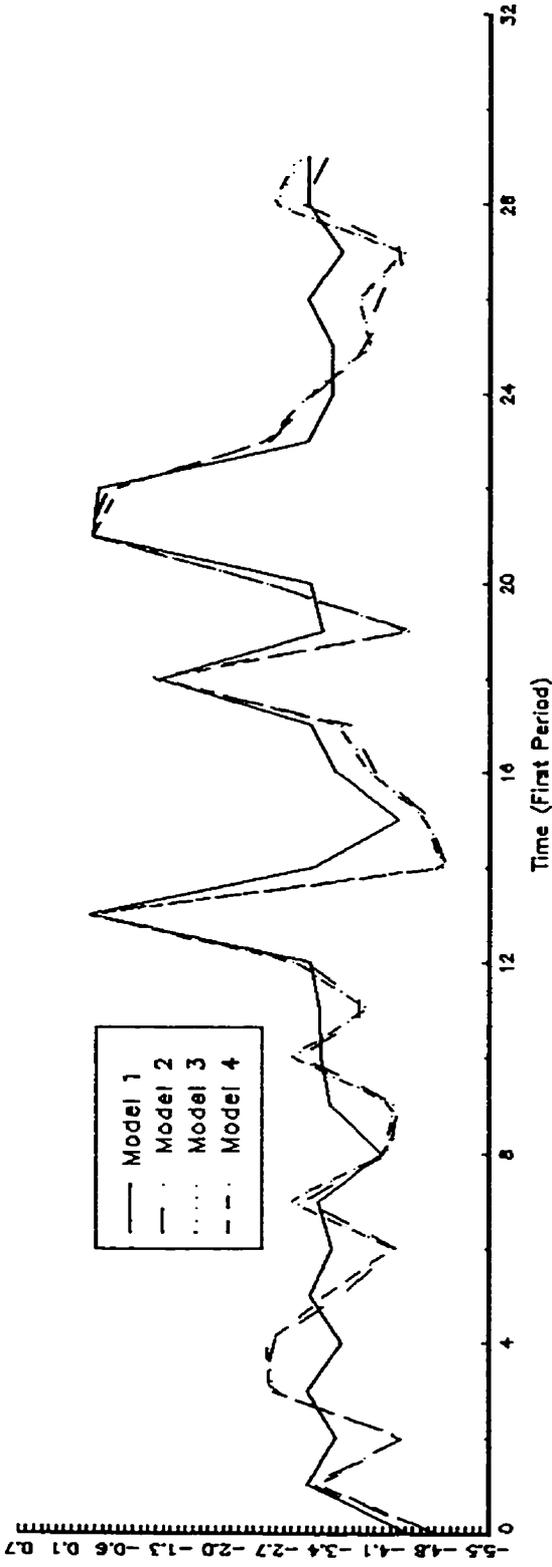
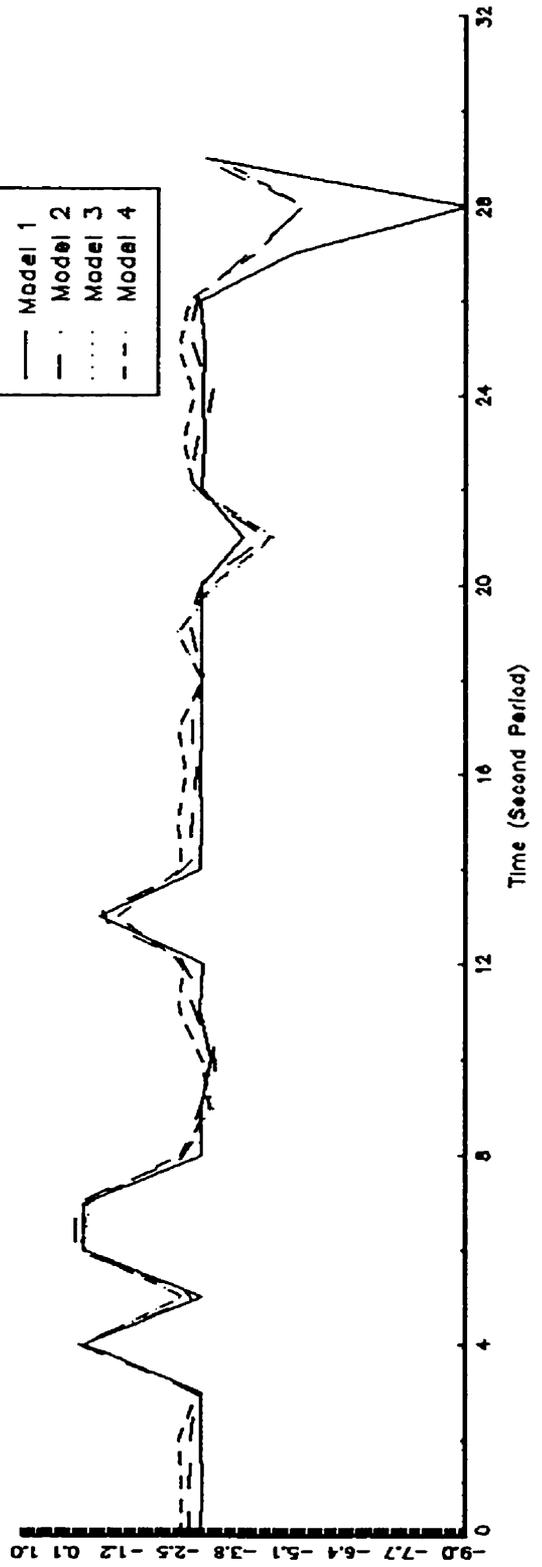


Figure 8

Log Predictive Marginal Likelihood



Log Predictive Marginal Likelihood



A Appendix: Proof of Proposition 6.1

An apparent way of avoiding non-tractable likelihood function in the derivation is to use the “augmented likelihood function” . Note that

$$\begin{aligned}
L(Z|y_t = 0, t \in I_2, \theta, \sigma^2, \omega, \delta^+, \delta^-) \\
&= \prod_{t \in I_1} \frac{1}{\sigma \sqrt{\omega_t}} \phi \left(\frac{y_t - X_t' \beta - \lambda y_{t-1}^* + \delta^+}{\sigma \sqrt{\omega_t}} \right) \prod_{t \in I_2} \frac{1}{\sigma \sqrt{\omega_t}} \phi \left(\frac{y_t^* - X_t' \beta - \lambda y_{t-1}^*}{\sigma \sqrt{\omega_t}} \right) \\
&\quad \prod_{t \in I_3} \frac{1}{\sigma \sqrt{\omega_t}} \phi \left(\frac{y_t - X_t' \beta - \lambda y_{t-1}^* + \delta^-}{\sigma \sqrt{\omega_t}} \right)
\end{aligned}$$

The middle term of the product implies that $\delta^- \leq y_t^* \leq \delta^+$. Hence, for any prior of δ^+ , say, $\delta^+ = \pi(\delta^+)$, the posterior distribution of δ^+ is

$$\begin{aligned}
\delta^+ | \{Z, \theta, \sigma^2, \omega, \delta^-\} &\sim \pi(\delta^+) L(Z, |y_t = 0, t \in I_2, \theta, \sigma^2, \omega, \delta^-) \\
&\sim \pi(\delta^+) \prod_{t \in I_1} \frac{1}{\sigma \omega_t} \phi \left(\frac{y_t - X_t' \beta - \lambda y_{t-1}^* + \delta^+}{\sigma \sqrt{\omega_t}} \right) \prod_{t \in I_2} \phi(y_t^* - X_t' \beta - \lambda y_{t-1}^*) \\
&\sim \pi(\delta^+) \exp \left[-\frac{1}{2} \sum_{t \in I_1} \left(\frac{y_t - X_t' \beta - \lambda y_{t-1}^* + \delta^+}{\sigma \sqrt{\omega_t}} \right)^2 \right] \prod_{t \in I_2} \phi \left(\frac{y_t^* - X_t' \beta - \lambda y_{t-1}^*}{\sigma \sqrt{\omega_t}} \right) \\
&\sim \pi(\delta^+) \exp \left[-\frac{1}{2} \sum_{t \in I_1} \left(\frac{\delta^+ - u_t}{\sigma \sqrt{\omega_t}} \right)^2 \right] I_{\{y_t^* \leq \delta^+, t \in I_2\}} \\
&\sim \pi(\delta^+) \exp \left[-\frac{1}{2} \left(\frac{\delta^+ - \sum_{t \in I_1} u_t \omega_t^{-1} / \sum_{t \in I_1} \omega_t^{-1}}{\sigma (\sum_{t \in I_1} \omega_t^{-1})^{-\frac{1}{2}}} \right)^2 \right] I_{\{\delta^+ \geq \{y_t^*, t \in I_2\}\}}
\end{aligned}$$

where $u_t = -(y_t - X_t' \beta - \lambda y_{t-1}^*)$, $t \in I_1$.

In this derivation, it is useful to note the following. As $t \in I_1 \cup I_3$, I have an exact linear relationship among the variables y_t , y_t^* and the parameters δ^+ or δ^- . This means that if the information of y_t , δ^+ or δ^- (and also others) is used to update the latent data y_t^* , then the information of y_t and y_t^* (and also others) cannot be used again to update δ^+ and δ^- . If $t \in I_2$, then the information of y_t and y_t^* updates the threshold parameters δ^+ and δ^- . The conditional distribution of δ^- can be worked out similarly.

References

- [1] Abrams, R.K., Froyen, R. and Waud, R.N. (1980): "Monetary Policy Reaction Functions. Consistent Expectations and the Burns Era," *Journal of Money, Credit and Banking* **12**, 30-42.
- [2] Bradley, M. D., and Dennis, W. J. (1986): "Federal Reserve Operating Procedure in the Eighties," *Journal of Money, Credit, and Banking* **18**, 325-335.
- [3] Brunner, K. and Meltzer, A. (1990): "Money Supply," in *Handbook of Monetary Economics* **1**, eds by B.M. Friedman and F.H. Hahn. 357-398.
- [4] Chib, S. (1995): "Marginal Likelihood from the Gibbs Output." *Journal of the American Statistical Association* **90**, 1313-1321.
- [5] Chib, S. and Greenberg, E. (1994): "Understanding the Metropolis-Hastings Algorithm." Forthcoming.
- [6] Dagenais, M.G. (1969): "A Threshold Regression Model," *Econometrica* **37**, 193-203.
- [7] Dagenais, M.G. (1973): "Application of a Threshold Regression Model to Household Purchases of Automobiles," *The Review of Economics and Statistics*. 275-285.
- [8] Derosa, P. and Gary, H.S. (1977): "Monetary Control and the Federal Funds Rate." *Journal of Monetary Economics* **3**, 217-230.
- [9] Desarbo, W. S., Rao, V. R., Steckel, J. H., Wind J. and Colombo R. (1987): "A Friction Model for Describing and Forecasting Price Changes," *Marketing Science* **6**, 299-319.
- [10] Dotsey, M. (1989): "Monetary Control under Alternative Operating Procedures," *Journal of Money, Credit, and Banking* Vol. **21**, No. 3 (August), 273-90.

- [11] Evans, M. and Swartz, T. (1995): "Methods for Approximating Integral in Statistics with Special Emphasis on Bayesian Integration Problems," *Statistical Science* Vol. 10, No. 3, 254-272.
- [12] Feinman, J. (1993): "Estimating the Open Market Desk's Daily Reaction Function," *Journal of Money, Credit, and Banking* 25, 231-147.
- [13] Friedman, B.M. (1988): "Lessons on Monetary Policy from the 1980s," *Journal of Economic Perspective* V., No. 3, 51-72.
- [14] Friedman, B.M. (1990): "Targets and Instruments of Monetary Policy," *Handbook on Monetary Economics*, edited by Benjamin Friedman and Frank Hahn. Amsterdam: North-Holland.
- [15] Genz, A. (1992): "Numerical Computation of Multivariate Normal Probabilities," *Journal of Computational and Graphical Statistics* 1, 141-150.
- [16] Geweke, J. (1995): "Bayesian Comparison of Econometric Models," University of Minnesota and Federal Reserve Bank of Minneapolis *Working Paper* 532.
- [17] Geweke, J. (1993): "Bayesian Treatment of the Independent Student-t Linear Model," *Journal of Applied Econometrics* 8, 19-40.
- [18] Geweke, J. (1992): "Priors for Macroeconomic Time Series and Their Application," *Institute for Empirical Macroeconomics Discussion Paper No. 64*, Federal Reserve Bank of Minneapolis.
- [19] Goodhart, C.A.E. (1989): "Money, Information and Uncertainty," 2nd edition, New York: Macmillian.
- [20] Hajivassiliou V., McFadden D. and Ruud P. (1996): "Simulation of Multivariate Normal Rectangle Probabilities and Their Derivatives: Theoretical and Computational Results," *Journal of Econometrics* 72, 85-134.

- [21] Havrilesky, T. M. (1975): "A Test of Monetary Policy," *Journal of Political Economy*, 299 - 304.
- [22] Heller, H.R. (1988): "Implementing Monetary Policy," *Federal Reserve Bulletin (July)*, 419 - 429.
- [23] Kohn, D. (1989a): "Policy Targes and Operating Procedures in the 1990s," *Monetary Policy Issues in the 1990s*, Federal Reserve Bank of Kansas City, 129 - 41.
- [24] Kohn, D. (1989b): "Monetary Policy in an Era of Change," *Federal Reserve Bulletin (February)*, 53-57.
- [25] Lang, K. L., Little, R. J. A. and Taylor, J. M. G. (1989): "Robust statistical modelling using the t-distribution," *Journal of the American Statistical Association* 84, 881-896.
- [26] Lindsey, D. (1986): "The Monetary Regime of the Federal Reserve System." Chapter 5 in C. Campbell and W. Dougan (eds), *Alternative Monetary Regimes*. Baltimore: Johns Hopkins University Press.
- [27] Lombra, R.E. (1993): "The Conduct of U.S. Monetary Policy," in *Monetary Policy in Developed Economics*, eds by Fratianni M.U. and Salvatore D., Greenwood Press
- [28] Maddala, G.S. (1983): *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.
- [29] Mauskopt, E. (1990): "The Transmission Channels of Monetary Policy: How Have They Changed?," *Federal Reserve Bulletin (December)*, 985-1008.
- [30] McNeese, S. K. (1986): "Modelling the Fed: A forward-looking Monetary Policy Reaction Function," *New England Economic Review* .
- [31] Melton, W. (1985): "Inside the Fed: Making Monetary Policy," Homewood, IL: Dow Jones-Irwin.

- [32] Meltzer A. (1991): *Monetary Policy on the 75th Anniversary of the Federal Reserve System*. Norwell, Mass. : Kluwe Academic Publishers, 3-65.
- [33] Meulendyke, A. (1988): "A Review of Federal Reserve Policy Targets and Operating Guides in Recent Decades," *Quarterly Review, Federal Reserve Bank of New York (Autumn)*, 6 - 17.
- [34] Meulendyke, A. (1989): *U.S. Monetary Policy and Financial Markets*. New York: Federal Reserve Bank of New York.
- [35] Partlan, J., Hamdani, K. and Camilli K. (1986): "Reserves Forecasting for Open Market Operations," *Quarterly Review, Federal Reserve Bank of New York (Spring)*, 19 - 33.
- [36] Poirier, D. J. (1995): *Intermediate Statistics and Econometrics: A Comparative Approach*, MIT Press.
- [37] Poirier, D. J. (1988): "Bayesian Diagnostic Testing in the General Linear Normal Regression Model," in *Bayesian Statistics 3*, New York: Clarendon Press, 725 - 732.
- [38] Roosa, R. V. (1956): "Federal Reserve Operations in the Money and Government Securities Markets," *Federal Reserve Bank of New York*.
- [39] Rosett, R.N. (1959): "A Statistical Model of Friction in Economics," *Econometrica* **26**, 263 - 267.
- [40] Rouse, R.G. (1961): "Implementation of the Policies of the Federal Open Market Committee," *Federal Reserve Bank of New York, Monthly Review, July*, 126 - 129.
- [41] Simon, D. P. (1991): "Secrecy, Signalling and the Accuracy of Expectations during the Borrowed Reserves Operating Regime," *Journal of Banking and Finance* **15**, 329-341.
- [42] Smith, A. F. M. and Roberts, G. O. (1993): "Bayesian Computation via the Gibbs sampler and Related Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society Series B* **55**, No. 1, 3-23.

- [43] Sternligh, P. et al. (1990): "Monetary Policy and Open Market Operations in 1989." *Federal Reserve Bank of New York Quarterly Review* **15**.
- [44] Tanner, M.A. and Wong, W.-H. (1987): "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association* **82**, 528-550.
- [45] Terrell, G.R. (1990): "The Maximal Smoothing Principle in Density Estimation." *Journal of the American Statistical Association* **85**, 470-477.
- [46] Terrell, G.R. and Scott, D. W. (1985): "Oversmoothed Nonparametric Density Estimates," *Journal of the American Statistical Association* **80**, 209-214.
- [47] Tierney, L. (1994): "Markov Chains for Exploring Posterior Distributions (with discussion)," *Annals of Statistics* **22**, 1701-1762.
- [48] Thornton, D. (1988): "The Borrowed Reserves Operating Procedure: Theory and Evidence," *Review, Federal Reserve Bank of St. Louis (January/February)*, 30-54.
- [49] Tobin, J. (1958): "Estimation of Relationship for Limited Dependent Variables." *Econometrica* **26**, 24-36.
- [50] Tsionas, E. G. (1996): "Bayes Factors for General Likelihoods via Posterior Simulation." *Mimo*, University of Toronto
- [51] Wei, S. X. (1995): "A Bayesian Approach to Dynamic Tobit Models." *Manuscript*, Department of Economics, University of Toronto.
- [52] Wood, J. H. (1967): "A Model of Federal Reserve behavior," In *Monetary Process and Policy: A Symposium*, edited by George Horwich, Homewodd, IL: Richard D. Irwin, 135 - 166.
- [53] William, P. (1988): "Monetary Policy Lessons of Recent Inflation and Disinflation." *Journal of Economic Perspectives* Vol. **2**, No. 3, 73 - 100.

**Do Limit Rules of Futures Prices Matter
in GARCH Processes?**

1 Introduction

Futures markets appear to be the only markets in which a daily **price limit** rule exists. This peculiar institutional feature has been analyzed in a theoretical framework by Brennan (1986) who argued that the imposition of price limits on futures prices is optimal with respect to reducing (costly) margin requirements. From a practical point of view, futures exchange officials contend that limiting the magnitude of daily price changes prevents large price movements caused by speculative excesses and thus enables the markets to move in a more orderly fashion. In addition, minimizing the incentive to renege on contract commitments (*i.e.*, lowering potential default risk in futures contracts) adds another justification to imposing price limits in futures markets. Based on these arguments, it seems reasonable to believe that the behaviors of economic agents who engage in futures trading might be affected by price limits. Truly, the observability of futures market equilibrium prices is limited in a daily varying band. Reasoning along this line, the current Chapter raises and solves, by drawing empirical evidence, a general and fundamental issue: whether the price limits significantly affect the processes of futures returns.

In the existing empirical work (for example, Hodrick and Srivastava (1987), McCurdy and Morgan (1987, 1988)), price limits in futures markets have been largely ignored. In the event of a limit move, these papers treated the observed settlement price as if it was the market equilibrium price. The advantage of so doing is clear: modeling is much simpler. This “simplified” approach is called the *Ostrich algorithm* (see Davidson and MacKinnon (1993) and Morgan and Trevor (1996)). In fact, ignoring price limits amounts to assuming that the price censoring² affects neither the *behavior* of market equilibrium prices, nor the *ways* (*i.e.*, models) to learn about this behavior. In terms of an empirical hypothesis, this states that limit rules in futures markets do not matter in the processes of futures

²In the event of a limit move, the settlement price that equals the reached limit serves as an indicator that the market equilibrium price is beyond the limit. In this sense, price limits are a well-defined *censored* problem.

prices³. Two implications might be drawn from this hypothesis: (1) the existing empirical analyses that ignore the price limits are thus justified; and (2) the imposition of price limits in futures markets seems not necessary because the market equilibrium price of a futures contract would not significantly deviate from its observed counterpart in the event of a limit move.

However, there are two major arguments against the “ignorance” hypothesis. *Firstly*, price limits may change the behavior of equilibrium (futures) prices because market participants may act based on observed market prices. Price limits can thus enter the process of the equilibrium prices through the observed market prices. This might be called a *real* effect of price limits. Under the assumption of competitive futures markets, this real effect is *distortionary* (to the economy) simply because the competitive market equilibrium prices without the limit regulation allocate resources efficiently⁴. *Secondly*, price limits change the information transformation mechanism through which to view futures markets. Brennan (1986), among others, argued that investors in futures markets may effectively learn in the event of a limit move the market equilibrium price of a futures contract from related markets. If investors’ decisions are made based on the inferred equilibrium prices, then limit rules may indeed not alter the equilibrium price process. If so, price limits serve only as sampling thresholds since the market equilibrium prices are unobservable under the limit-induced market-breaks. In this circumstance, the effect of price limits is nothing more than a *veil*, or it is *nominal*. In sharp contrast with the above real perspective, the nominal view implies a non-distortionary effect of price limits, because the fundamental equilibrium price process is kept non-distorted.. The favoring evidence of this nominal hypothesis would thus justify and support the imposition of price limits.

From a statistical point of view, limit rules can affect both first and second moments

³For expository convenience, the two terms — the process of futures prices and the process of futures returns — are used interchangeably throughout this Chapter because they determine each other.

⁴This efficiency is understood as Pareto optimality.

of futures returns. The current literature in empirical finance relies heavily on ARCH-type specifications (see Bollerslev, Chou and Kroner (1992) for a survey, and references therein). The key characteristic of these models is to successfully capture time-varying second-moments. After the pioneering work of Engle (1982) and its various extensions, the GARCH(1.1) model introduced by Bollerslev (1986) has emerged as a reliable workhorse that is able to approximate the main characteristics of financial data in most cases. Limit rules, if imposed, can play roles in both the conditional means and the conditional variances in a GARCH process. This consideration motivates us to develop censored GARCH models for futures data. The recursive structure of the second moments for a GARCH process may carry the censoring effect (from the price limits) over to the estimates of all future conditional second moments, consequently to the estimates of all parameters and to making other inferences. Needless to say, this censoring problem, when combining with GARCH errors, deserves a serious investigation. This is the focus of the current Chapter.

Two recent papers by Kodres (1993) and Morgan and Trevor (1996) involve the issue of price limits in futures markets. Kodres (1993) engages in investigating whether the price limits would affect a hypothesis test: is the futures rate an unbiased predictor for the future spot rate? She concludes that researchers using foreign exchange futures data no longer need to be concerned about the effects of price limits when performing the tests of unbiasedness. The major contribution of Kodres' approach is to combine an explicit model for price limits with a model of conditional variance. Criticizing the complexity of Kodres' approach, Morgan and Trevor (1996) explain how to estimate the GARCH model with price limits by an modified EM algorithm. Realizing the very low degree of limit moves (at most 2%) in Kodres' samples, they employ U.S. Treasury bill futures data over the most volatile period (October 1979 – October 1982). The conclusion is that the price limits have a significant effect on the estimated parameters for a GARCH (1,1) model.

There are three shortcomings with the two approaches. *Firstly*, their conclusions are all

based on asymptotic theory which is notoriously difficult for ARCH type models (Bollerslev, Chou, and Kroner (1992)). Although Lumsdain (1991) shows that standard asymptotically based inference procedures are generally valid, Hong (1988) already presents Monte Carlo evidence to demonstrate that sample sizes must be quite large for the asymptotic distributions to provide good approximations. The sample sizes are 778 for Morgan and Trevor's data, and about 3500 for Kodres five data sets. It seems no widely accepted standard for how large a sample size must be in order to use asymptotic theory. In particular, the sample size for Morgan and Trevor may not be easy to justify being quite large. *Secondly*, all the tests they adopted are based on the criteria of in-sample fit. This may not be always convincing because one model may fit data nearly perfectly, but possess little predictive power (see, for example, Dumas, Fleming and Whaley (1995)). Finally, detailed examinations of why the price limits matter or do not matter were not presented in their papers.

Removing all the above shortcomings, this Chapter addresses the significance of price limits in futures markets by conducting a Bayesian analysis. One additional advantage of this prior-posterior analysis is that the non-nested model specifications in this Chapter can be easily dealt with. Two statistical innovations are obtained: (1) a general technique of computing Bayes factors for censored models, especially for dynamic censored models; and (2) a modified concept of predictive Bayes factor.

The remainder of this Chapter is organized as follows. GARCH and censored GARCH models are motivated and discussed in Section 2. A simple and practical posterior simulation technique (the Griddy Gibbs sampler with data augmentation) is proposed in Section 3. Section 4 discusses the methods of assessing the significance of the price limits and the developed methods are applied to U.S. Treasury bill futures data in Section 5. Section 6 concludes the Chapter.

2 GARCH and Censored-GARCH Models

The widespread existence of ARCH-type effects for most financial data has been well-known and extensively researched since the seminal work of Engle (1982) (See Figure 1 for the ARCH-type effect, *i.e.* volatility clustering, of the returns of U.S. Treasury bill futures). The generalized ARCH (*i.e.*, GARCH) models introduced by Bollerslev (1986) have proven attractive for the returns of most financial assets. In terms of Akaike or Schwarz information criteria (Bollerslev, Chou and Kroner (1992)), a GARCH (1,1) model typically dominates other GARCH (or ARCH) specifications. Ignoring price limits, GARCH models can be directly applied to observed futures returns or prices (see, for example, McCurdy and Morgan (1988)). However, there are major reasons not to ignore price limits.

(1) In the event of a limit price move, the market equilibrium price is unobservable. The market-break may result in a significant information loss for investors. The increased uncertainty in the investors' decisions may potentially affect the behavior of futures returns. Thus, the adequacy of a pure GARCH process that ignores the effect of price limits becomes questionable.

(2) There are two possible conjectures about investors' behaviors. One is that they act based on the observed market prices. This implies that limit prices can enter the process of underlying equilibrium prices through the observed market prices, and impose a *real* effect on the economy. Under the assumption of competitive futures markets, this real effect is distortionary simply because the equilibrium prices without the limit regulation allocate resources efficiently.

(3) Another conjecture is that investors make their decisions as if they could "see" the equilibrium prices in the events of limit moves. This view is justified by assuming investors can effectively learn the equilibrium prices from related markets in the events of limit moves

(see Brennan (1986)). Investor's decisions are based on the (inferred) market equilibrium prices. This means that price limits would not alter the market equilibrium prices, rather serving only as thresholds for sampling. Intuitively, the effect of price limits is nothing more than a *veil*, or it is *nominal*. In sharp contrast with the above "real" perspective, this "nominal" view asserts a non-distortionary effect of price limits.

(4) The observed futures returns with the limited price moves create artificial autocorrelations of the futures returns (see Kodres (1993)). For example, the zero risk premium hypothesis of futures returns is more likely violated. This implies that the limit rules in futures markets may have a significant effect on the estimation of the conditional mean of the futures returns.

(5) The limit rule can also affect the conditional variance process of the futures returns. An immediate observation is that the price limits will reduce the conditional variances of the observed returns. The recursive structure of the conditional variance specification in a GARCH model implies that today's reduced conditional variance will be carried over to all future estimates of conditional variances. This may be viewed as the effect of the price limits on the second moment of the futures return process.

Thus, it is reasonable to believe that with the possibility of daily price limits, the view of futures returns through a pure GARCH "window" (*i.e.*, likelihood, see Poirier (1995) who used "window" as a metaphor to capture the essential role of likelihood) is inappropriate. A simple review of censored model literature (see for example, Maddala (1987) and Ameiyia (1985)) suggests that a pure GARCH specification is incorrect because it does not take account of the censoring effect. To represent the conditional volatility more adequately, two censored-GARCH models, one of which is a modified version of Kodres (1993) and Morgan and Trevor (1996), are formulated in this section⁵. The first model captures this

⁵ Also see Calzolari and Fiorentini (1996) who developed a classical estimation method for Tobit-GARCH models. Morgan and Trevor (1996) explored the applicability of the EM estimation method to this kind of

nominal effect that the institutional feature merely affects the observability of the underlying equilibrium return process. This implies that the underlying equilibrium returns follow a GARCH process which is not governed by price limits. The second one, however, assumes that price limits affect not only the mean process, but also the volatility process, of the underlying equilibrium returns. Let

F_t^* be the equilibrium futures price at time t ,

F_t be the observed futures price at time t ,

y_t^* be the rate of change of the equilibrium futures price from $t-1$ to t ,

y_t be the rate of change of the observed futures price from $t-1$ to t ,

h_t^* be the scaled variance of y_t^* conditional on information I_{t-1} available at $t-1$.

h_t be the scaled variance of y_t conditional on information I_{t-1} available at $t-1$.

The conventional approach to futures returns is to ignore the daily price limits. A GARCH(1, 1) model is given by

Model 1.

$$\begin{cases} y_t = \frac{F_t}{F_{t-1}} - 1 = \mu + \varepsilon_t \\ \varepsilon_t | I_{t-1} \sim t(0, h_t, \nu) \\ h_t = \alpha + \gamma \varepsilon_{t-1}^2 + \delta h_{t-1} \end{cases} \quad (1)$$

where F_t and y_t are observed futures prices and returns and treated as if they were the equilibrium counterparts; the innovation ε_t is orthogonal to all the available information at time $t-1$, I_{t-1} ; the predictive distribution of $\varepsilon_t | I_{t-1}$ is assumed to follow a Student- t distribution with mean zero, variance $h_t \frac{\nu}{\nu-2}$ if $\nu > 2$, and ν is the degree of freedom parameter; the parameters of the volatility function are restricted to be positive. The parameter μ represents the mean return, or (constant) risk premium of futures contracts.

models.

This model is widely used in current practice partly because introduction of price limits makes modeling much more complicated.

The first censored-GARCH model, based on Morgan and Trevor (1996) and Kodres (1993), assumes that the underlying equilibrium returns follow a GARCH(1.1) process.

Model 2.

$$y_t = \begin{cases} \bar{c}_t & \text{if } y_t^* \geq \bar{c}_t \\ y_t^* & \text{if } \underline{c}_t \leq y_t^* \leq \bar{c}_t \\ \underline{c}_t & \text{if } y_t^* \leq \underline{c}_t \end{cases} \quad (2)$$

$$\begin{cases} y_t^* = \frac{F_t^*}{F_{t-1}^*} - 1 = \mu + \varepsilon_t^* \\ \varepsilon_t^* | I_{t-1} \sim t(0, h_t^*, \nu) \\ h_t^* = \alpha + \gamma(\varepsilon_{t-1}^*)^2 + \delta h_{t-1}^* \end{cases}$$

where \bar{c}_t and \underline{c}_t are the lower and upper limits of the returns⁶ (missing observations can be treated to be censored within an interval $(\underline{c}_t, \bar{c}_t)$ where $\underline{c}_t = -\infty$ and $\bar{c}_t = +\infty$) and are assumed to be known *ex ante*; and other notations are the same as those in Model 1. This model says that the price limits merely affect the way to learn about the behavior of the underlying equilibrium returns, but have no effect on the equilibrium return process. The observed limit moves are a pure sampling problem.

The second censored-GARCH model is given by

Model 3.

$$y_t = \begin{cases} \bar{c}_t & \text{if } y_t^* \geq \bar{c}_t \\ y_t^* & \text{if } \underline{c}_t \leq y_t^* \leq \bar{c}_t \\ \underline{c}_t & \text{if } y_t^* \leq \underline{c}_t \end{cases} \quad (3)$$

⁶These limited returns are based on the corresponding limited prices.

$$\begin{cases} y_t^* = \frac{F_t^*}{F_{t-1}^*} - 1 = \mu + \varepsilon_t^* \\ \varepsilon_t^* | I_{t-1} \sim t(0, h_t, \nu) \\ h_t = \alpha + \gamma \varepsilon_{t-1}^2 + \delta h_{t-1} \\ \varepsilon_{t-1} = y_{t-1} - \mu \end{cases}$$

where the notations here are the same as those in Model 2. In sharp contrast with Model 2, this specification emphasizes the governing effect of limit rules on the underlying equilibrium process of returns. It is the observed lagged error ε_{t-1} , rather than the unobserved lagged error ε_{t-1}^* , that dictates the volatility process of futures returns.

The likelihood function of Model 1 (*i.e.*, a GARCH(1,1) model) is given by

$$L(\mu, \alpha, \gamma, \delta; Y) = \prod_{t=1}^T \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{\nu}{2})} (\nu h_t)^{-1/2} [1 + (y_t - \mu)^2 / \nu h_t]^{-\frac{\nu+1}{2}}. \quad (4)$$

and the likelihood functions of Model 2 and Model 3 (*i.e.*, censored-GARCH(1,1) models) are derived in Appendix A. It is noted, however, that the “augmented likelihoods” of the two censored GARCH models are the same as (4). The “augmented likelihood” is the imaginary “likelihood” as if the latent data are “observable”. Following the GARCH literature, h_0 is estimated from the unconditional sample variance of return data. The prior distribution of each individual parameter is given by

$$\begin{aligned} \mu &\sim U(-\xi_1, \xi_2), \\ \alpha &\sim U(0, a), \\ \gamma &\sim U(0, 1), \\ \delta &\sim U(0, 1), \\ \nu &\sim EXP(\lambda^{-1}) \quad \nu > 0, \end{aligned} \quad (5)$$

where $U(c, d)$ denotes a uniform distribution in the interval (c, d) ; $EXP(\lambda^{-1})$ is an exponential distribution with parameter λ^{-1} ; and $\{\xi_1, \xi_2, a, \lambda\}$ are hyperparameters indexing the

prior distributions. All parameters are assumed to be *independently* distributed, *ex ante*. The stationary condition of the innovation process is $\gamma \frac{\nu}{\nu-2} + \delta < 1$ if $\nu > 2$ (see Bauwens and Lubrano (1996)). These prior specifications, though proper, are extremely flat, except for the degree of freedom parameter. Particular specifications of the prior distributions do not matter with respect to the posterior computations of subsequent sections because integrations are carried out using a “black-box” numerical method. If the supports of μ and α are large enough, and λ is sufficiently small, these priors are (almost) non-informative, though still proper. Thus the posterior distribution will depend only on data information. It might be interesting to use non-uniform priors to capture more precise opinions on the parameters. This Chapter does not report the results with more informative priors. Choices of the hyperparameters will be discussed in Section 5.

3 Posterior Simulations

This section considers simulations from posterior distributions. I adopt a special version of the Gibbs sampler with data augmentation technique: the Griddy Gibbs sampler (Ritter and Tanner (1992)) with data augmentation. The Griddy Gibbs has been recently used by Bauwens and Lubrano (1996) to solve a GARCH model. The reason for choosing this method, particularly in this Chapter, is that the full conditional distributions of both parameters and latent data are in non-standard forms (*i.e.*, the distributions or distribution kernels cannot be directly or easily sampled) and that I need to sample from truncated (unknown) distributions. A brief review of the sampler is given next.

3.1 The Griddy Gibbs Sampler with Data Augmentation

The idea of the Gibbs sampler with data augmentation is the following. Suppose the parameters of interest θ can be (often naturally) split into two (or more) blocks, $\theta = (\theta_1, \theta_2)$.

Let y be observations, z latent data (they might be also split into blocks when necessary), if any. If the conditional distributions

$$\begin{aligned}\theta_1 &| \theta_2, z, y \\ \theta_2 &| \theta_1, z, y \\ z &| \theta_1, \theta_2, y\end{aligned}\tag{6}$$

are standard (for example, normal, gamma, or Student-t distributions), then the Gibbs sampler with data augmentation algorithm is to sample iteratively from the above conditional distributions. However, the conditional distributions (6) are not tractable for both the parameters and latent data (to be discussed later) of GARCH and censored GARCH models. Consequently, the Griddy method is taken to draw variates from each of the full conditional distributions. The algorithm (the Griddy Gibbs sampler with data augmentation) works as follows for M draws (see Bauwens and Lubrano (1996)):

- *step 1*: initialize the chain at any value $\theta^{(0)}$ of the support of the parameter θ .
- *step 2*: start the loop at $n = 1$.
- *step 3*: compute $p(\theta_1|\theta_2, z, y)$ over the grid $(\zeta_1, \zeta_2, \dots, \zeta_G)$ to obtain the vector $G_\kappa = (\kappa_1, \kappa_2, \dots, \kappa_G)$.

- *step 4*: take integration to get the values $G_p = (0, \Phi_2, \dots, \Phi_G)$ where

$$\Phi_i = \int_{\zeta_1}^{\zeta_i} p(\theta_1|\theta_2, z, y) d\theta_1 \quad i = 2, 3, \dots, G$$

compute the normalized pdf values $G_\varphi = G_\kappa/\Phi_G$ of $p(\theta_1|\theta_2, z, y)$.

- *step 5*: generate $u \sim U(0, 1)$ and invert $\Phi(\theta_1|\theta_2, z, y)$ to get a draw $\theta_1^{(n)}$.
- *step 6*: redo step 3-6 for θ_2 and each of latent data y_i^* , if any.
- *step 7*: increment n by 1 and go to step 3 unless $n > M$.
- *step 8*: discard the initial m draws, and return all other draws.

The posterior output⁷ from the above algorithm can be written as

$$\left\{ \theta_1^{(n)}, \theta_2^{(n)}, z^{(n)} \right\}_{n=m+1}^M.$$

The convergencies of the MCMC chains are crucial for making correct inferences. Yu and Mykland (1994) developed a simple and convenient convergence criterion called the CUM-CUM method⁸. This method has been advocated by Robert (1995) and Bauwens (1996). Let N be the draws of a Monte Carlo Markov chain (MCMC). Then the statistic

$$CS_t = \left(\frac{1}{t} \sum_{n=1}^t \theta_1^n - \bar{\theta}_1 \right) / \sigma_{\theta_1}$$

is used to visually detect the convergence of the chain. The $\bar{\theta}_1$ and σ_{θ_1} are the empirical mean and standard deviation of the N draws. If the MCMC chain converges, the graph of CS_t against t should be converge smoothly to zero. On the other hand, a long and regular excursion away from zero are the indication of an absence of convergence.

There are three advantages of the Griddy Gibbs sampler. *Firstly*, it can readily deal with a general density function (or kernel). *Secondly*, it works with the density kernel without the need of the normalizing constant of the density. *Finally*, it is quite convenient to sample from truncated distributions. This final point is especially important to this Chapter because I need to sample latent data from truncated kernels. However, the disadvantages

⁷The parameter estimation can be done by simply using the posterior output and the conditional structures of the parameters. For example, the posterior mean and variance of the parameter θ_1 can be written as a Monte Carlo integration:

$$\begin{aligned} \hat{E}(\theta_1|y) &= \frac{1}{M} \sum_{n=m+1}^{m+M} E(\theta_1|\theta_2^{(n)}, z^{(n)}, y) \\ \widehat{Var}(\theta_1|y) &= E(\theta_1^2|y) - \{\hat{E}(\theta_1|y)\}^2 \end{aligned}$$

The estimations of parameter θ_2 and latent data z have similar expressions. Based on a Rao-Blackwell argument (see Gelfand and Smith (1990)), these estimators are more efficient than the sample mean and variance of θ_1 . Of course, if the mean and variance for $\theta_1|\theta_2, z, y$ are not in closed or tractable forms, the sample mean and variance of the Gibbs output for θ_1 (θ_2 and z) are readily used as the estimate of θ_1 . This may require more iterations of the cycles (3.1) to obtain the same accuracy as the ones mentioned just above.

⁸Many have been working on the convergence criteria of MCMC chains, but only a few results have been delivered. For example, Zellner and Min (1995) introduced three simple criteria which are valid only for the case of two blocks as in (3.2).

of the Griddy method are (1) the computational speed is low and (2) the dimensions of (dependent) parameters and latent data should be relatively small. Fortunately, these are not a problem for this Chapter.

3.2 Conditional Distributions

The derivation of full conditional distributions (see (6)) is the key to the Griddy Gibbs sampler with data augmentation. Attention will be focused on Model 2 which represents the most general case among the three models discussed in Section 2 from a statistical point of view. First of all, consider that given the latent data, Model 2 becomes a GARCH (1.1) process. The augmented posterior distribution of the parameters which can be derived from its “augmented likelihood” function is given by,

$$p(\mu, \alpha, \gamma, \delta, \nu | Z) = \frac{\lambda \exp(-\lambda\nu)}{a(\xi_2 - \xi_1)} \prod_{t=1}^T \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\Gamma(\frac{1}{2})} (\nu h_t)^{-1/2} [1 + (z_t - \mu)^2 / \nu h_t]^{-\frac{\nu+1}{2}} \quad (7)$$

where $Z = \{z_t\}_{t=1}^T$, $z_t = y_t$ if y_t^* is observable, and z_t is simulated otherwise. The conditional distributions of the parameters thus have the following density kernels.

$$\begin{aligned} \mu | \alpha, \gamma, \delta, \nu, Z &\sim \prod_{t=\bar{T}+1}^T h_t^{-1/2} [1 + (z_t - \mu)^2 / \nu h_t]^{-\frac{\nu+1}{2}} & \xi_1 < \mu < \xi_2. \\ \alpha | \mu, \gamma, \delta, \nu, Z &\sim \prod_{t=\bar{T}+1}^T h_t^{-1/2} [1 + (z_t - \mu)^2 / \nu h_t]^{-\frac{\nu+1}{2}} & 0 < \alpha < a. \\ \gamma | \mu, \alpha, \delta, \nu, Z &\sim \prod_{t=\bar{T}+1}^T h_t^{-1/2} [1 + (z_t - \mu)^2 / \nu h_t]^{-\frac{\nu+1}{2}} & 0 \leq \gamma < 1. \\ \delta | \mu, \alpha, \delta, \nu, Z &\sim \prod_{t=1}^{\bar{T}} h_t^{-1/2} [1 + (z_t - \mu)^2 / \nu h_t]^{-\frac{\nu+1}{2}} & 0 \leq \delta < 1 - \gamma. \\ \nu | \mu, \alpha, \gamma, \delta, Z &\sim \exp(-\lambda\nu) \prod_{t=1}^T \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} (\nu h_t)^{-1/2} [1 + (z_t - \mu)^2 / \nu h_t]^{-\frac{\nu+1}{2}} & \nu > 0. \end{aligned} \quad (8)$$

All the distributions are non-standard. This is where the Griddy Gibbs sampler intervenes.

Now consider deriving the conditional distributions of the latent data. Wei (1995) gave a

general method in this aspect. It is noted that although there is no correlation between two consecutive (equilibrium) returns, the observations are *not* independent. For the current models, this implies that today's return enters the distribution of tomorrow's return. This makes the matter a bit complicated. For convenience, define a *latent string* as a subset of consecutive observations which begins with an uncensored observation following immediately a censored observation, and ends immediately after the next uncensored observation. A latent string consists of an information unit to learn about the censored data in this string. Some examples and detailed discussion of the concept in a dynamic setting can be found in Wei (1995). The conditional distribution of the latent data in a latent string is given in the following proposition.

Proposition 3.1 The conditional distribution of the latent data $\{y_{t+1}^*, \dots, y_{t+n_t}^*\}$ in a latent string $\{y_t, y_{t+1}^*, \dots, y_{t+n_t}^*, y_{t+n_t+1}\}$ is

$$z_{t+1}, \dots, z_{t+n_t} | \Omega \sim h_{t+n_t+1}^{-1/2} \prod_{j=1}^{n_t} h_{t+j}^{-1/2} [1 + (z_{t+j} - \mu)^2 / \nu h_{t+j}]^{-\frac{\nu+1}{2}} I_{\{z_{t+j} \in A_{t+j}\}} \quad (9)$$

where $A_t = \{z_t | z_t \leq \underline{c}_t \text{ if } y_t = \underline{c}_t, z_t \geq \bar{c}_t \text{ if } y_t = \bar{c}_t\}$; $\Omega = \{y_{t-1}, y_{t+n_t+1}, \theta\}$; and $I(\cdot)$ is an indicator function.

The proof of the proposition is provided in Appendix B. Apparently, this is a non-standard distribution because the latent variables enter the recursive expressions of h_{t+j} . Finding an algorithm to draw the latent data from the joint conditional distribution might be difficult. This is because I need to draw from a truncated non-standard joint distribution. However, it is possible to partition the latent data in a latent string into several one-dimensional latent variables, just like the treatment of parameters. From (9), the conditional density kernel of one dimensional latent observation can be simply derived as

$$z_{t+j} | \Omega_{j|-j} \sim h_{t+j+1}^{-1/2} h_{t+j}^{-1/2} [1 + (z_{t+j} - \mu)^2 / \nu h_{t+j}]^{-\frac{\nu+1}{2}} I_{\{z_{t+j} \in A_{t+j}\}} \quad (10)$$

where $\Omega_{j|-j} = \{y_{t-1}, y_{t+n_t+1}, z_{t+\tau}, \tau \neq j, \theta\}$.

Now all the kernels of full conditional distributions of the parameters and latent data have been derived. The posterior output can be obtained by applying the Griddy Gibbs sampler with data augmentation as discussed in Subsection 3.1.

4 The Assessment of Price Limits

The informational content of the significance of price limits is reflected in the sharp contrast among the GARCH and censored GARCH specifications (see (1) – (3)). The assessment of price limits, in fact, involves primarily comparing the performances of these models. Given that the three models are non-nested, classical approaches (to model choices) are unsatisfactory (see, for example, a discussion by Gelfand (1996)) if they exist. Bayesian methods provide at least a better alternative, but some conceptual modification and further development of a computational technique are required for the current situation. This section is intended to explain and develop these ideas.

Model comparisons are addressed through features of predictive distributions. They involve comparing the relative performances of predictions from different models. Basically, two types of predictive distributions are widely used: (1) prior-predictive distributions; and (2) posterior-predictive distributions. Let me explain each of them and their roles in comparing models.

A complete Bayesian model is defined by the specification of a joint distribution⁹ for observables, say Y_T (*i.e.*, data), unobservables, say θ (*i.e.*, parameters). Denote $f(\cdot)$ as a generic notation for the joint distribution, which is the product of the likelihood function and the prior distribution. Then the prior-predictive distribution is defined as

⁹I use distribution rather than density because in latent data models, the joint distribution is not always a density function. The same consideration is applied to my subsequent context.

$$f(Y_T|\mathcal{M}) = \int f(Y_T, \theta|\mathcal{M})d\theta = \int L(Y_T|\theta, \mathcal{M})\pi(\theta|\mathcal{M})d\theta \quad (11)$$

where \mathcal{M} indicates a model. This distribution is also called the predictive distribution of data by statisticians, and the marginal or marginalized likelihood by Bayesian econometricians. Based on this distribution, a widely-used model comparison measure — *Bayes factor* — can be readily defined. Given any two models, say \mathcal{M}_2 and \mathcal{M}_3 , Bayes factor is given by¹⁰

$$B_{23} = \frac{f(Y_T|\mathcal{M}_2)}{f(Y_T|\mathcal{M}_3)}. \quad (12)$$

The Bayes factor provides the relative weight of evidence for model \mathcal{M}_2 compared with model \mathcal{M}_3 . The calibration of the factor has been discussed in Kass and Raftery (1995), among others. Its computation has to rely on sampling-based methods, due to the complicated nature of the current models. One major task of this section is to develop a general technique for computing Bayes factor over any two dynamic censored models. I will also show that it is misleading to compare Model 1 with Model 2 and Model 3 in this fashion because Model 1 treats the censored observations as uncensored (see the detailed discussion in the next subsection).

A posterior-predictive distribution is defined similarly. Divide the sample (or subsample, in case if the prediction for some particular period of data is interested) into two parts, say, $\{y_1, y_2, \dots, y_{T_1}\}$ and $\{y_{T_1+1}, y_{T_1+2}, \dots, y_{T_1+T_2}\}$ ($T_1 + T_2 \leq T$). The first part of observations is used to estimate the model or to derive the posterior distribution of parameters, *i.e.* $f(\theta|Y_{T_1})$. Actually, I need only a sample from the posterior distribution in this Chapter. Then, the posterior-predictive distribution for the second part of observations can be written

¹⁰Here I tried to avoid the comparison of Model 1 with Model 2 and Model 3 because Bayes factors are misleading measures for this special case. See my detailed discussion later.

as

$$\begin{aligned}
& f(y_{T_1+1}, y_{T_1+2}, \dots, y_{T_1+T_2} | Y_{T_1}, \mathcal{M}) \\
&= \int f(y_{T_1+1}, y_{T_1+2}, \dots, y_{T_1+T_2}, \theta | Y_{T_1}, \mathcal{M}) d\theta \tag{13} \\
&= \int L(y_{T_1+1}, y_{T_1+2}, \dots, y_{T_1+T_2} | Y_{T_1}, \theta, \mathcal{M}) f(\theta | Y_{T_1}, \mathcal{M}) d\theta
\end{aligned}$$

In this distribution, the likelihood of the second part of the sample (or subsample) is averaged (or marginalized) by the posterior distribution $f(\theta | Y_{T_1}, \mathcal{M})$. The differences between two types of predictive distributions are clear: (1) a prior-predictive distribution examines the global performance of a model, and a posterior-predictive distribution checks the local performance of a model; (2) because a posterior is often less sensitive to prior changes, it seems that a posterior-predictive distribution is less influenced by prior than a prior-predictive distribution; and (3) while a prior-predictive distribution reflects the in-sample features of a model, a posterior-predictive distribution shows the out-of-sample properties of a model. These characteristics are also shown in the model comparison measures built on the two different kinds of predictive distributions.

Similarly to the concept of Bayes factor, predictive Bayes factor can be readily defined in terms of the posterior-predictive distribution (13). It is discussed by Geweke (1995) in a linked fashion and favored by many Bayesians (for example, Gelfand (1996), Geweke (1995) and Evans and Swartz (1995), among others). Given the difficulty of comparing Model 1 with Model 2 and Model 3, it makes sense to compare the out-of-sample predictability of models. This idea will be explained and modified in this section for my current models. Certainly, other predictive distributions can also be used for model comparisons (for example, cross-validation distributions (Gelfand, Dey and Chang (1992)), and intrinsic predictive distributions (Berger and Pericchi (1994))). They are not further pursued in this Chapter because they are computationally demanding.

4.1 In-Sample Fit Model Comparison

The advantages of using Bayes factors over classical methods in hypothesis testing have been widely discussed in the econometric and statistical literature (see Kass and Raftery (1995), Poirier (1995), Geweke (1995), Koop and Potter (1996), etc.). Three of the advantages are worth mentioning with respect to the current study. *Firstly*, Bayes factors can easily handle both nested and non-nested hypotheses, which is in sharp contrast with the classical procedures which are often restricted to nested hypotheses. *Secondly*, there exists an automatic penalty built to Bayes factors for more complex models, and classical methods have no built-in protection against over-parameterization (Occam's Razor). *Finally*, the uncertainty of a likelihood function is taken into account by averaging it with the prior distribution over the whole parameter space. In most cases, classical methods evaluate the likelihood function at a fixed point (ML estimate).

In practice, Bayes factors are often criticized to be associated with the computational difficulty involved in performing high orders of integrations. This issue has been largely solved in recent empirical applications (Kass and Raftery (1995), Chib (1995), Verdinelli and Wasserman (1993)). As a further contribution to this body of research, this subsection provides a simulation-based approach to a fairly general class of models, especially to dynamic censored models. For an illustration, note that the Bayes factor of Model 2 (\mathcal{M}_2) vs Model 3 (\mathcal{M}_3) can be written as

$$B_{23} = \frac{f(Y_T|\mathcal{M}_2)}{f(Y_T|\mathcal{M}_3)} = \left[\frac{L(Y_T|\theta_{(2)}, \mathcal{M}_2)}{L(Y_T|\theta_2, \mathcal{M}_3)} \right] \left[\frac{\pi(\theta_{(2)}|\mathcal{M}_2)}{f(\theta_{(2)}|Y_T, \mathcal{M}_2)} \frac{\pi(\theta_{(3)}|\mathcal{M}_3)}{f(\theta_{(3)}|Y_T, \mathcal{M}_3)} \right] \quad (14)$$

where $f(Y_T|\mathcal{M})$ (the subscripts are suppressed) is the marginal likelihood of the model \mathcal{M} and the second equality is a rearrangement in terms of an alternative expression of the Bayes rule (see Chib (1995)); $\theta_{(i)}$ $i = 1, 2$ is the parameter vector associated with Model i . This expression says that the Bayes factor of Model 2 over Model 3 equals their likelihood ratio

adjusted by a factor determined by their prior and posterior distributions, all evaluated at fixed parameter vectors, say $\theta_{(i)}$. Theoretically, choices of $\theta_{(i)}$ do not matter because of the identity of marginal likelihood (see Chib (1995)). In practice, it is more convenient to take values of $\theta_{(i)}$ close to the mode of the posterior distribution.

On the computationally convenient log scale,

$$\ln B_{23} = \ln \left[\frac{L(Y_T|\theta_{(2)}, \mathcal{M}_2)}{L(Y_T|\theta_{(3)}, \mathcal{M}_3)} \right] + \ln \left[\frac{\pi(\theta_{(2)}|\mathcal{M}_2)}{\pi(\theta_{(3)}|\mathcal{M}_3)} \right] - \ln \left[\frac{p(\theta_{(2)}|Y_T, \mathcal{M}_2)}{p(\theta_{(3)}|Y_T, \mathcal{M}_3)} \right] \quad (15)$$

That is, Bayes factor is jointly determined by three ratios: likelihood, prior, and posterior. Again, this equality holds at any fixed $\theta_{(2)}$ and $\theta_{(3)}$. The strategy of so doing is intended to avoid multiple integrals involved in the marginal likelihoods, because the parameter θ (usually high dimension) has to be integrated out. In fact, this line of approach has only partly success in avoiding multiple integrals, since they are also involved in the likelihoods of the two models. The additional integrals can be avoided in the following estimates for the three terms in (15).

1. Evaluation of $\ln \left[\frac{L(Y_T|\theta_{(2)}, \mathcal{M}_2)}{L(Y_T|\theta_{(3)}, \mathcal{M}_3)} \right]$

It is easy to prove that this term can be consistently estimated by

$$\frac{1}{M} \sum_{i=m}^{m+M} \frac{L(Y_T, z^{(j)}|\theta_{(2)}, \mathcal{M}_2)}{L(Y_T, z^{(j)}|\theta_{(3)}, \mathcal{M}_3)} \quad (16)$$

where $L(Y_T, z|\theta, \mathcal{M})$ is the ‘‘augmented likelihood’’ of the model \mathcal{M} and $z^{(i)}$ are the simulated latent data from the conditional distribution of $z|Y_T, \theta_{(3)}$ in Model 3. This expression is understood as a Monte Carlo integration from

$$\int \frac{L(Y_T, z|\theta_{(2)}, \mathcal{M}_2)}{L(Y_T, z|\theta_{(3)}, \mathcal{M}_3)} p(z|Y_T, \theta_{(3)}, \mathcal{M}_3) dz$$

where it is easy to show that this is just an alternative expression of the likelihood ratio of these two models. The sample $z^{(j)}$ is drawn by cycling the data augmentation procedure with the fixed $\theta_{(i)}$, $i = 1, 2$. Taking the logarithm to (16) yields a consistent estimate of the first term of (15).

2. Evaluation of $\ln \left[\frac{p(\theta_{(2)}|Y_T, \mathcal{M}_2)}{p(\theta_{(3)}|Y_T, \mathcal{M}_3)} \right]$

Suggested by Tsionas (1996), the posterior density can be generally estimated by using a multivariate kernel density estimator (\mathcal{M} is suppressed):

$$\hat{p}(\theta|Y_T) = (M \det(A))^{-1} \sum_{n=m+1}^{m+M} K(A^{-1}(\theta^{(n)} - \theta))$$

where $AA' = cS$, c is a constant, S is the sample covariance of the posterior output, and K is a density kernel, for example, $K(u) = \exp(-uu'/2)I_{\Theta}(\theta)$, where $I_{\Theta}(\theta)$ is an indicator function, i.e., $I_{\Theta}(\theta) = 1$ if $\theta \in \Theta$ and 0 otherwise. The constant c can be optimally determined by the maximal smoothing principle (see Tsionas (1996) and references therein):

$$c = \left[\frac{(k+8)^{(k+6)/2} \pi^{k/2}}{16M\Gamma((k+8)/2)(k+2)} \int_{\Theta} K(u)^2 du \right]^{2/(k+4)}$$

where the integral is available analytically.

The advantage of this procedure is its generality. The only input required is a sample from the posterior distribution of θ . Thus, it is convenient to combine it with simulation-based methods, for example, the MCMC methods. The accuracy of the procedure can be improved by thinning the sample, say returning every tenth iterate (suggested by Tsionas) for dependent draws. This is designed to obtain an approximately random sample from the posterior distribution. More efficiently, we can compute or plot the autocorrelations of the posterior draws. The choice of thinning step can be essentially based upon the computed or plotted autocorrelation structure.

Now it is important to point out the inappropriateness of comparing Model 1 with Model 2 or Model 3 in this fashion.. To understand, recall the definitions of prior-predictive distribution and Bayes factor as defined early in this section. In the event of a limit move, the marginal likelihood contribution of this observation is computed as probabilities for Model 2 and Model 3, but as a density value for Model 1 (because Model 1 deliberately and “wrongly” treats this observation as uncensored!). The contribution of the observation to Bayes factor B_{12} (or B_{13}) can be simply written as a Density/Probability ratio. This ratio is misleading to measure the relative performances of models. This can be understood by noting that with a small variance of the marginal likelihood density, this ratio can be extremely large no matter which model is better. On the opposite situations, this ratio can be extremely small. Therefore, in-sample fit model comparison is inappropriate in this fashion for Model 1 versus Model 2 and Model 3. It seems difficult to modify the concept of Bayes factor to adapt to this special situation.

If one is still interested in computing the Bayes factor of Model 1 versus Model 2 or Model 3, the above method is no longer applicable. The harmonic mean of the likelihood values proposed by Newton and Raftery (1994)

$$f(Y_T) = \left\{ \frac{1}{M} \sum_{j=m+1}^{m+M} L(Y_T, z^{(j)} | \theta_{(2)}^{(i)})^{-1} \right\}^{-1}$$

can be instead used for the general purposes. This estimation is, however, often criticized to be “unstable” because the inverses of the likelihood values evaluated at some tailed draws of $\theta_{(2)}$ may overwhelmingly dominate the sum. A simple modification of the estimation is to calculate

$$f(Y_T) = \frac{1}{M} \sum_{j=m+1}^{m+M} \frac{g(\theta_{(2)})}{L(Y_T, z^{(j)} | \theta_{(2)}) \pi(\theta_{(2)})}$$

where $g(\cdot)$ is any d-dimensional probability density (d = the number of total parameters in the evaluated model) . This was suggested by Gelfand and Dey (1994).

It is worth commenting on the developed computational technique for Bayes factor. Comparing with those appeared in the literature, this technique is quite general in the sense that it is appropriate and powerful for various applications, especially involving latent observations. The generalized Savage-Dickey density ratio method (see Verdinelli and Wasserman (1995)) is simple, but is restricted only to nested hypotheses. The method proposed above is clearly without this restriction. The method proposed by Chib (1995) is also restrictive in the sense that the conditional posterior distributions of the parameters and latent data need to return their normalizing constants. This might be extremely difficult when the conditionals are non-standard (in most cases, only density kernels are known). The method proposed here is more flexible than Chib's (1995) approach. Comparing with the harmonic mean of the likelihood values (see Kass and Raftery (1995) and references therein), this method is simple and stable for dynamic censored models.

4.2 Out-of-Sample Predictive Model Comparison

Whether data favor one model rather than another should not be only determined in terms of their in-sample fits, but also based on their out-of-sample predictions. It is often the case that a better in-sample fit does not necessarily imply a better out-of-sample prediction. Given the difficulty of comparing Model 1 with Model 2 and Model 3 in terms of their in-sample fits, it makes sense to compare the out-of-sample predictability of these models. One such measure, discussed and favored by Geweke (1995), is *linked predictive Bayes factors*. It is different from the Bayes factor as discussed before, since they are based on posterior-predictive distributions. This section illustrates and modifies this idea for comparing the current models.

Rather than looking at the predictions for the second part of the data directly (recall the discussion and notation in the beginning of this section), this method is based on a

decomposition of the predictive distribution (13) (see Geweke (1995)):

$$f(y_{T_1+1}, y_{T_1+2}, \dots, y_{T_1+T_2} | Y_{T_1}, \mathcal{M}) = \prod_{\tau=1}^{T_2} f(y_{T_1+\tau} | Y_{T_1+\tau-1}, \mathcal{M}) \quad (17)$$

where $f(y_{T_1+\tau} | Y_{T_1+\tau-1}, \mathcal{M})$ is just one period-ahead prediction¹¹. Then the corresponding decompositions of the predictive Bayes factor is

$$B_{j|i, T_2} = \prod_{u=1}^{T_2} B_{j|i, u} \quad (18)$$

where $B_{j|i, u} = \frac{f(y_{T_1+u} | Y_{T_1+u-1}, \mathcal{M}_j)}{f(y_{T_1+u} | Y_{T_1+u-1}, \mathcal{M}_i)}$. This decomposition essentially examines the individual observations that are more probable under one model or the other. A few things here need to be clarified. *Firstly*, to compute one-period ahead prediction, say y_{T_1+u+1} given Y_{T_1+u} , the posterior sample $\{\theta^{(n)}, \omega^{(n)}\}_{n=m+1}^{m+M}$ ($\omega^{(n)} = 1$ if $u = 1$) needs to be updated in each further step. This can be done in terms of Geweke's (1995) importance MCMC method and its modified procedure for censored models (see Wei (1996)). Then, an updated posterior sample $\{\theta^{(n)}, \omega_i^{(n)}\}_{n=m+1}^{m+M}$ is obtained, from $\theta | Y_{T_1}$, and $\omega_i^{(n)}$ is the weight derived from the importance function for Model i ($i = 1, 2$, and 3). Based on the updated posterior samples, one-period ahead predictive distribution of y_{T_1+u+1} is given by

$$\begin{aligned} f(y_{T_1+u+1} | Y_{T_1+u}, \mathcal{M}_1) &= \int L(y_{T_1+u+1} | Y_{T_1+u}, \theta, \mathcal{M}_1) f(\theta | Y_{T_1+u}, \mathcal{M}_1) d\theta \\ &= \sum_{n=m+1}^{m+M} \omega_1^{(n)} L(y_{T_1+u+1} | Y_{T_1+u}, \theta^{(n)}, \mathcal{M}_1) / \sum_{n=m+1}^{m+M} \omega_1^{(n)} \end{aligned} \quad (19)$$

for Model 1, and

$$\begin{aligned} f(y_{T_1+u+1}^* | Y_{T_1+u}, \mathcal{M}) &= \int L(y_{T_1+u+1}^* | Y_{T_1+u}, \theta, \mathcal{M}) f(\theta | Y_{T_1+u}, \mathcal{M}) d\theta \\ &= \sum_{n=m+1}^{m+M} \omega_1^{(n)} L(y_{T_1+u+1}^* | Y_{T_1+u}, \theta^{(n)}, \mathcal{M}) / \sum_{n=m+1}^{m+M} \omega_1^{(n)} \end{aligned} \quad (20)$$

¹¹Multi-period ahead prediction is a trivial extension of one-period ahead prediction.

for Model 2 and Model 3.

Secondly, analogously to the discussion in the last subsection, if y_{T_1+u+1} is a limit move, then it is again misleading to compare the value $f(y_{T_1+u+1}|Y_{T_1+u}, \mathcal{M}_1)$ with the value of $f(y_{T_1+u+1}^* \in A_{T_1+u+1}|Y_{T_1+u}, \mathcal{M}_i)$ ($i = 2, 3$)! Recall the problem of the Density/Probability ratio in Bayes factor as discussed in the last subsection. To solve this problem, now think about comparing Model 2 with Model 3. If y_{T_1+u+1} is a limit move, the predictive Bayes factor involves comparing two probability values of $f(y_{T_1+u+1}|Y_{T_1+u}, \mathcal{M}_i)$ ($i = 2, 3$), which are defined as

$$\begin{aligned}
& f(y_{T_1+u+1}|Y_{T_1+u}, \mathcal{M}_i) \\
&= P(y_{T_1+u+1}^* \in A_{T_1+u+1}|Y_{T_1+u}, \mathcal{M}_i) \\
&= \int \left\{ \int_{A_{T_1+u+1}} L(y_{T_1+u+1}^* | Y_{T_1+u}, \theta, \mathcal{M}_i) dy_{T_1+u+1}^* \right\} f(\theta|Y_{T_1+u}, \mathcal{M}_i) d\theta \\
&= \sum_{n=m+1}^{m+M} \omega_i^{(n)} \int_{A_{T_1+u+1}} L(y_{T_1+u+1}^* | Y_{T_1+u}, \theta^{(n)}, \mathcal{M}_i) dy_{T_1+u+1}^* / \sum_{n=m+1}^{m+M} \omega_i^{(n)}.
\end{aligned} \tag{21}$$

Intuitively, because the true market equilibrium return is unobservable in the event of a limit move, the predictive probabilities that the market equilibrium return reaching the limit for the two models are compared. In order to compare Model 1 with Model 2 and Model 3 in this manner, it makes sense to compute the predictive probability as in (21) for Model 1 as well, and then compare it with those obtained from Model 2 and Model 3. Given a limit move observation Y_{T_1+u+1} , the probability of reaching the limit for Model 1 can be calculated by

$$\begin{aligned}
& P(y_{T_1+u+1} \in A_{T_1+u+1}|Y_{T_1+u}, \mathcal{M}_1) \\
&= \int \left\{ \int_{A_{T_1+u+1}} L(y_{T_1+u+1} | Y_{T_1+u}, \theta, \mathcal{M}_1) dy_{T_1+u+1} \right\} f(\theta|Y_{T_1+u}, \mathcal{M}_1) d\theta \\
&= \sum_{n=m+1}^{m+M} \omega_1^{(n)} \int_{A_{T_1+u+1}} L(y_{T_1+u+1} | Y_{T_1+u}, \theta^{(n)}, \mathcal{M}_1) dy_{T_1+u+1} / \sum_{n=m+1}^{m+M} \omega_1^{(n)}.
\end{aligned} \tag{22}$$

This above modification makes it possible and reasonable to compare the relative performances of Model 1 with Model i ($i = 2, 3$).

These (linked) predictive Bayes factors can be further understood by examining the predictive distributions of (especially censored) observations. This turns out to be quite easy because the predictive densities have been derived in the above discussion. It may be also interesting and important to see the effect of price limits on the second moments of the models. I will thus show the implications of the price limits on the predictive conditional variances of all three models because these quantities are the major concern for GARCH type models.

5 Application to Treasury Bill Futures

5.1 Background and Data

A Treasury Bill (T-bill) futures contract is an agreement to buy or sell Treasury bills in a specified future for a specified price (known as the futures price). U.S. Treasury bill futures was first introduced in January, 1976 at the International Monetary Market (IMM), a division of the Chicago Mercantile Exchange. The major role played by this financial instrument is to hedge short-run interest rate risks.

The data used in this study are the returns (or prices) of 3-month U.S. T-bill futures which is the most heavily traded interest rate futures¹². The sample covers the period from October 1979 to October 1982 inclusive. During this period, the Federal Reserve System (the Fed) adopted an operating procedure which focused on controlling money supply and allowed the interest rate to adjust more freely. The purpose of the implementing procedure is to combat non-tolerable inflation rate. The direct impact of the Fed's new strategy on the 3-month T-bill futures was the exceptional volatility in the market (see Figure 1).

Price limits of the 3-month T-bill futures are clearly shown in Table 2. Their converted return counterparts are displayed in Figure 2 (see the symbol \times). According to the market

¹²The data are provided by I.G. Morgan (see Morgan and Trevor (1996)).

regulation (IMM Yearbook (1983), p. 52), the price limits were set at ± 0.50 of the market index of the previous day (*i.e.*, 50 basis points above or below the preceding day's settlement price) before June 19, 1980 and then changed to ± 0.60 afterward.

There were limit moves in the next up contracts (*i.e.*, the contracts next up to the near by contracts) on 57 days out of 778 in the period. The ratio of the limit moves to total observations in the current sample is 7.3%, which is much higher than that for the foreign exchange futures used by Kodres (1993). For detailed sample statistics, see Table 1. The selection of this particular sample (in the special period) are desirable because the effect of the price limits, if significant, is hoped to be easily identified. Furthermore, this sample shows that 9 of the 57 price limit moves were occurred immediately following days with limit moves in the same direction and 1 was occurred in the opposite direction.

We need to be careful when converting price limits into return limits. For an isolated (price) limited day, this is easy. Suppose that F_t is the observed settlement price for a limit-up move. Then y_t is a limit-up return move and y_{t+1} is a limit-down return move. In consecutive limit moves, however, this simple conversion relationship will no longer hold. For an illustration, suppose F_t and F_{t+1} are both limit-up price moves. Then y_t (recalling that $y_t = (F_t - F_{t-1})/F_{t-1}$) is still a limit-up return move and y_{t+2} a limit-down return move. Unfortunately, y_{t+1} becomes indeterministic because no information can identify which of the two underlying prices F_t^* and F_{t+1}^* is larger. Consequently, y_{t+1}^* has to be treated as a missing observation (rather than a censored observation) for Model 2 and Model 3. In the current analysis, 9 of such observations are treated in such a way.

5.2 Empirical Results

This subsection reports the empirical results of implementing the tests discussed in the above subsection. Before proceeding, I point out that the volatility clustering, *i.e.* ARCH type

effect, is clearly shown in Figure 1. This provides an insight on why this analysis should be based on GARCH type models. Further, look at Figure 2 which displays the limits (including missing observations) for T-bill futures returns. At first glance, the presence of so many limited moves (see the symbol \times) suggests that we should *not* ignore these limits.

The designed hyperparameters of the three models are presented in Table 3.

Table 3. Hyperparameter Specification

Hyperparameter	Prior I			Prior II			Prior III		
	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3
ξ_1	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25	-0.25
ξ_2	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
a	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60
λ	0.01	0.01	0.01	0.05	0.05	0.05	0.02	0.10	0.08

Because of the complexity of GARCH and censored GARCH models, discriminating the prior beliefs on the three models becomes a difficult task at this point. This is possibly why Bauwens and Lubrano (1996) prefer to use flat priors for their GARCH model. In a similar vein, I selected flat priors for all parameters except the degree of freedom parameter in Section 2. In fact, the prior specification of the degree of freedom parameter, if its hyperparameter is small, is also flat. Therefore, once the support intervals determined by hyperparameters are specified to be large enough and the hyperparameter for the degree of freedom is small enough, the posterior distributions of the parameters for all three models will be dominated by the sample information. The interval (ξ_1, ξ_2) is specified to be the support of the distribution of the process mean of the futures returns. The parameter μ in each model captures the (daily) constant risk premium, if any, of the futures returns. A flat prior for it implies no discrimination among positive, negative and zero (constant) risk premiums. The efficient market hypothesis suggests a zero value for the process mean μ . The absolute values of ξ_1 and ξ_2 are allowed to be different so that any prior opinion on positive, or zero, or negative risk premiums can be imposed. The hyperparameter $a = .6$ can be understood similarly. The hyperparameter λ indexes the distribution of the degree-

of-freedom parameter. This specification follows Geweke (1993) who employed a power distribution to form the prior opinion on the degree of freedom parameter for a Student-t model. The three models have clearly different implications for the degree of freedom parameters. Model 2 has the fattest tails and Model 1 the thinnest tails, among the three models. This is simply because Model 2 allows for more flexible movements of the futures returns than Model 1 and Model 3, and Model 1 is most restrictive in taking extreme values. Priors II and III are designed based on this simple idea.

The in-sample model comparison is conducted through the Bayes factor of Model 2 over Model 3 (recalling the inappropriateness of comparing Model 1 with Model 2 and Model 3 in this fashion). The results are reported in Table 4, showing a strong support to Model 2. Based on this examination, current models show no evidence against non-distortionary effect of the price limits in U.S. T-bill futures markets. This empirical result might support the imposition of price limits in futures markets.

Table 4. Log Bayes Factors

\mathcal{M}_2	\mathcal{M}_3	$\ln B_{23}$
Prior I	Prior I	1.657
Prior II	Prior II	2.230
Prior III	Prior III	1.869

The out-of-sample predictive model comparisons are based on linked predictive Bayes factors over all the three models. Predictive marginalized likelihoods and predictive Bayes factors are computed in one-month increments. In each case the out-of-sample extends from the first business day of one month to the last business day of the month. In each model, predictive likelihoods are then formed for the next day, for the next two days, and so on until all business days through the whole month are included. The results are displayed in Figure 3. The upper panel indicates the log predictive likelihoods for the data in the month indicated, based on the posterior distribution for the sample extending through the first

working day of the month indicated.

Look at the first page of Figure 3. For August 1982, the log predictive Bayes factors show that Model 1 performs slightly better than Model 2 and Model 3 before the limit day (August 27, 1982). This pattern is true for September and October 1982 (see the next two pages of Figure 3). It is special for September 1982 because in this month, prices never reached their limits. Consequently, Model 1 performs slightly better than the two censored GARCH models for the whole month. However, the accumulated advantage of Model 1 vanishes immediately once the limit moves were reached (see the first and last pages of Figure 3). This shows that the GARCH model (Model 1) that ignores price limits is difficult to adapt to large shocks to the market. Statistically, this is due to the poor performance of Model 1 in tails. Given that on average more than 1 limit day occurs per month, based on Figure 3, intuitively it is appealing that Model 1 performs worse than Model 2 and Model 3. Figure 3 also shows that Model 2 performs insignificantly better than Model 3 for out-of-sample predictions. Though no contradiction is reached when comparing with the in-sample fit model comparison above, this again implies that in-sample fit and out-of-sample predictability are different measures. In empirical analysis, if possible, it is better to try both type of comparisons. Further, to see the behavior of limit observations, I plot the predictive densities for these observations (during the last month of the sample, *i.e.*, October 1982) and the observations themselves, in Figure 4. On October 7, 1982 the observed return 1.53779 (rescaled by 1000) reached the upper tail of its predictive distribution. The lower panel of the first page of Figure 4 illustrates why Model 2 and Model 3 perform much better than Model 1. See the next page of Figure 4 for a similar plot for October 11, 1982. However, on October 8 and 12, 1982, though they are converted return limit days, there are actually no shocks for the days, and consequently the observed returns lie in the high density regions of their predictive distributions. This explains why Bayes factors are declining sharply for the first day of a return limit, and have not much

change for the second day of a return limit. in Figure 3. The second days are actually not shocks.

Finally, I examine the predictive variances derived from all the three models. This is an important issue since ARCH-type models are mainly concerned about these quantities and they have a fundamental effect on option pricing and hedging. Figure 5 shows the predictions for September and October 1982. Given that Model 2 is the best among the three, the basic message behind this figure is that Model 1 systematically underpredicts the volatility of T-bill futures returns. It seems that when volatilities trend up, Model 3 tends to overpredict the volatilities, and when volatilities go down, Model 3 tends to underpredict the volatilities.. This pattern seems to be true for other months too (I predicted some other months, and results are similar but not reported because of space limitation).

6 Conclusions

This Chapter has shown a strong significance of price limits in futures markets. Two censored GARCH models are developed to address this issue. This is important because the model specifications differ substantially in terms of their economic implications. Because of the poor performance of Model 1 in tails, it is not surprising to obtain a significant result for price limits. Lack of accommodating large shocks (inducing tailed observations) is probably the reason why a pure GARCH model cannot compete with the censored counterparts.. In comparing Model 2 with Model 3, my results show that data support Model 2 better. One important economic or policy implication is that price limits seems non-distortionary. This may be viewed as the empirical supporting evidence for the imposition of price limits in futures markets. This Chapter also points out that a pure GARCH model systematically underpredicts the conditional variances of futures returns. This result suggests that people in both academy and the financial industry should take account of the price limits in futures

markets seriously, especially when the chances of hitting the price limits are high.

The technical contributions of this Chapter include: (1) an extension of the Griddy Gibbs sampler with data augmentation algorithm to censored GARCH models. This could be thought as a further extension of Bauwens and Lubrano (1996). It is also found that this technique is able to sample from truncated kernels easily. As emphasized within this Chapter, this posterior simulation algorithm is in fact a “black box” method which is quite general. The low sampling speed and the required low dimensionality are disadvantages of this method. Fortunately, these are not crucial problems for the models introduced in this Chapter; (2) point out a potential problem for model comparison in the Bayesian paradigm. In certain instances, Bayes factor and predictive Bayes factors can be misleading indicators because of the Density/Probability ratio. The concept of the predictive Bayes factor is modified, but it seems that there is no easy way to modify the Bayes factor; (3) develop a quite general method to compute Bayes factors for censored models. This method is applicable to a larger set of problems than the existing methods in the literature.

A Appendix: Likelihood Functions of Models 2 and 3

This appendix derives the likelihood functions of Models 2 and 3 (see Section 2). Notations here are the same as those in Sections 2 and 3. The idea is rather simple: integrate out the unobserved quantities in the “augmented likelihoods”. The augmented likelihood is the “imaginary likelihood” by treating the latent data as being observable. For instance, the “augmented likelihood” of Model 2 is

$$L(\theta|Y^*) = \prod_{t=1}^T \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{\nu}{2})} (\nu h_t)^{-1/2} [1 + (y_t^* - \mu)^2 / \nu h_t]^{-\frac{\nu+1}{2}} \quad (23)$$

where $h_t = \alpha + \gamma(\epsilon_{t-1}^*)^2 + \delta h_{t-1}$. This is however *not* the likelihood of the model, because some of y_t^* s are indeed unobservable. To integrate out the unobserved y_t^* , the dependences among the latent observations have to be considered. This would result in multiple integrals in the likelihood. The non-standard integrands prevent the use of any analytical integration methods, and this is where the data augmentation algorithm intervenes.. In fact, it is always possible to write out the likelihoods explicitly. For Model 2, it is given by,

$$L(\theta|Y) = \prod_{t \notin \varphi} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{\nu}{2})} (\nu h_t)^{-1/2} [1 + (y_t - \mu)^2 / \nu h_t]^{-\frac{\nu+1}{2}} \\ \prod_{t \in \varphi} \int_{A_t} \cdots \int_{A_t + n_t} \prod_{j=0}^{n_t+1} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{\nu}{2})} (\nu h_t)^{-1/2} [1 + (y_t^* - \mu)^2 / \nu h_t]^{-\frac{\nu+1}{2}} dy_t^* \cdots dy_{t+n_t}^*$$

where $A_t = \{y_t^*; y_t^* \leq \underline{c}_t, \text{ if } y_t = \underline{c}_t, y_t^* \geq \underline{c}_t, \text{ if } y_t = \bar{c}_t\}$. It is, however, rather difficult to explore the likelihood directly. Consequently, the “augmented likelihood” is instead used with the data augmentation technique.

Similarly, the likelihood of Model 3 is

$$L(\theta|Y) = \prod_{t \notin C} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{\nu}{2})} (\nu h_t)^{-1/2} [-(y_t - \mu)^2 / \nu h_t]^{-\frac{\nu+1}{2}}$$

$$\prod_{t \in C} \int_{A_t} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{\nu}{2})} (\nu h_t)^{-1/2} [-(y_t^* - \mu)^2 / \nu h_t]^{-\frac{\nu+1}{2}} dy_t^*$$

where C denotes the censored time points.

B Appendix: Proof of Proposition 1

This appendix provides a simple proof of Proposition 1, i.e., derives the distribution of latent data in each latent string conditioning on all observations and parameters. This involves a crucial step in the cycles of the Gibbs sampler with data augmentation as mentioned in Section 3.1.

Based on the augmented likelihood of Model 2 (see (19)), the latent data in a latent string, say,

$$y_t^*, \dots, y_{t+n_t}^*, y_{t+n_t}^*$$

conditioning on all parameters follow a distribution with the kernel

$$\prod_{j=1}^{n_t} (\nu h_{t+j})^{-1} \left[1 + \frac{(y_{t+j}^* - \mu)^2}{\nu h_{t+j}} \right]^{-\frac{\nu+1}{2}}. \quad (24)$$

However, the kernel contains h_{t+j} which is also a function of the latent data in the string. Thus this distribution is not easy to sample. In addition, given that the latent data $\{y_t^*, \dots, y_{t+n_t}^*\}$ fall in a special region, $\{y_t^*, \dots, y_{t+n_t}^*\}$ follows a truncated distribution with the truncation kernel

$$\prod_{j=1}^{n_t} (\nu h_{t+j})^{-1} \left[1 + \frac{(y_{t+j}^* - \mu)^2}{\nu h_{t+j}} \right]^{-\frac{\nu+1}{2}} I_{\{y_{t+j}^* \in A_{t+j}\}} \quad (25)$$

where $A_t = \{y_t^*; y_t^* \leq \underline{c}_t \text{ if } y_t = \underline{c}_t, y_t^* \geq \underline{c}_t \text{ if } y_t = \bar{c}_t\}$ and $I(\cdot)$ is an indicator function.

Q. E. D.

Table 1. Sample Statistics (Daily Return $\times 1000$) of T-Bill Futures

Sample Size	778	Minimum Value of Return	-1.56
Sample Mean	-0.0038	Number of Converted Return Limit	95
Sample Variance	0.75	Number of Converted Missing Values	9
Maximum Value of Return	2.26		

Table 2. List of Days with Limit Price Moves

Month	Day	Year	Indicator	Month	Day	Year	Indicator
10	8 ₍₁₎	1979	-	10	27	—	-
10	9 ₍₂₎	—	-	11	7	—	+
10	18 ₍₁₎	—	-	12	11	—	-
10	19 ₍₂₎	—	-	12	19 ₍₁₎	—	+
10	26	—	+	12	22 ₍₂₎	—	+
11	9 ₍₁₎	—	+	1	5	1981	+
11	12 ₍₂₎	—	+	1	21	—	-
1	11	1980	-	1	26	—	+
1	24 ₍₁₎	—	-	2	19	—	+
1	25 ₍₂₎	—	-	4	6	—	-
2	15 ₍₁₎	—	-	5	4 ₍₁₎	—	-
2	19 ₍₂₎	—	-	5	5 ₍₂₎	—	-
3	3	—	-	5	21	—	-
3	6	—	-	6	8	—	+
3	14 ₍₁₎	—	-	6	18	—	-
3	17 ₍₂₎	—	+	7	20	—	-
3	24	—	-	8	3	—	-
4	3 ₍₁₎	—	+	8	24	—	-
4	7 ₍₂₎	—	+	9	11	—	+
4	10 ₍₁₎	—	+	11	24	—	+
4	11 ₍₂₎	—	+	12	4	—	+
4	16	—	+	12	10	—	-
4	28	—	+	2	1	1982	-
5	2	—	+	2	22	—	+
5	15	—	-	6	1	—	-
5	23	—	+	8	27	—	-
8	18	—	-	10	7	—	+
9	22	—	-	10	11	—	+
10	3	—	+				

Note: 1. The positive sign + in the indicator columns represents limit-up price move, and the negative sign - limit-down price move.
 2. The subscripts (1) and (2) denote, respectively, the first day and the second day, of two-day consecutive price limit moves.

Figure 1. Return of U.S. T-Bill Futures
(A General Plot)

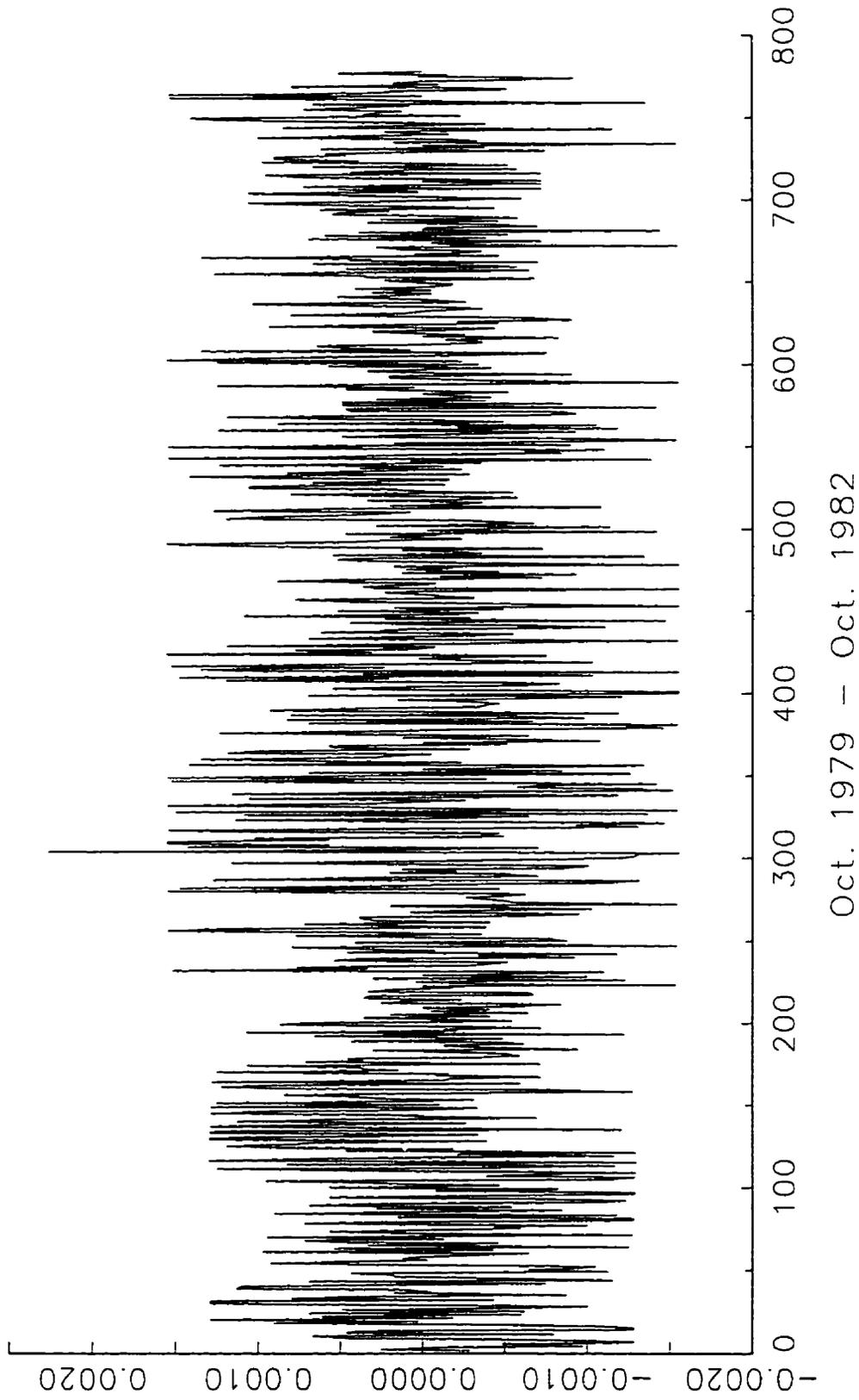


Figure 2. Return of U.S. T-Bill Futures
(A Detailed Plot)

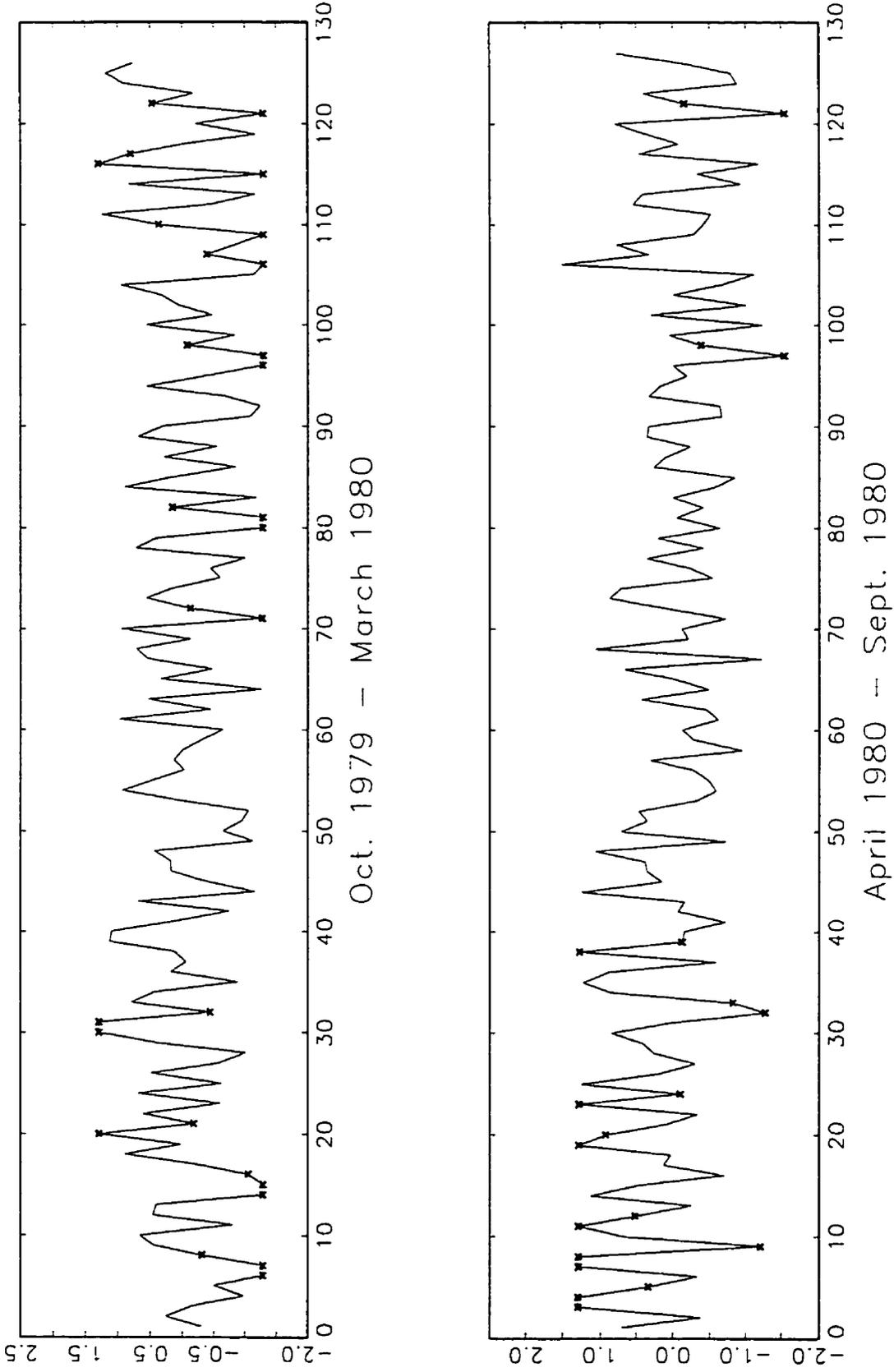


Figure 2. (Continued)

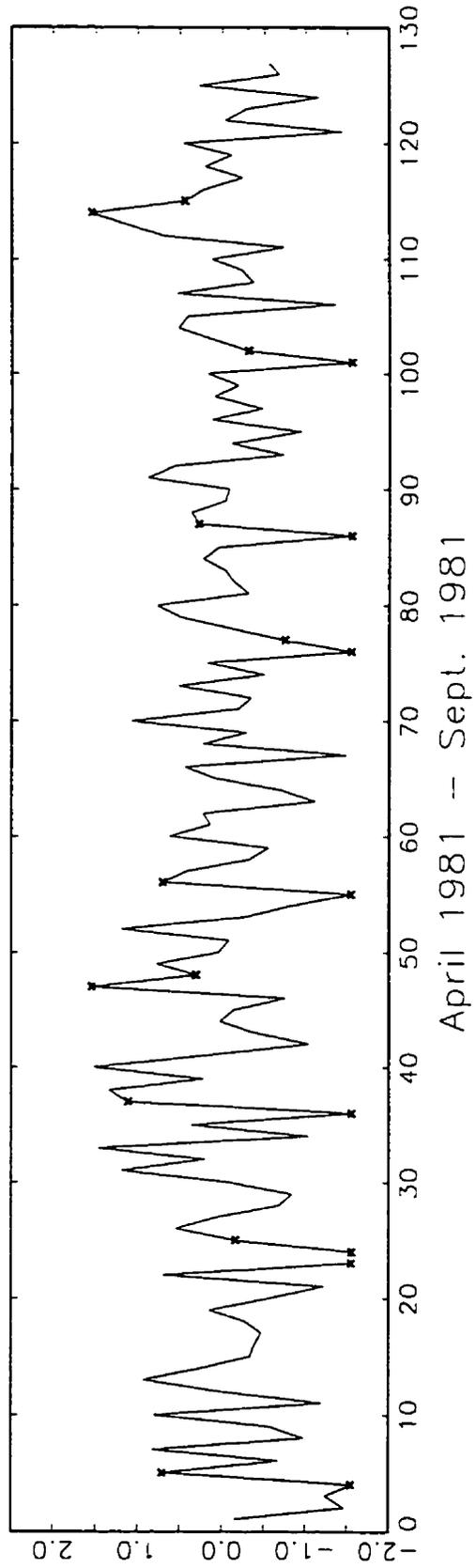
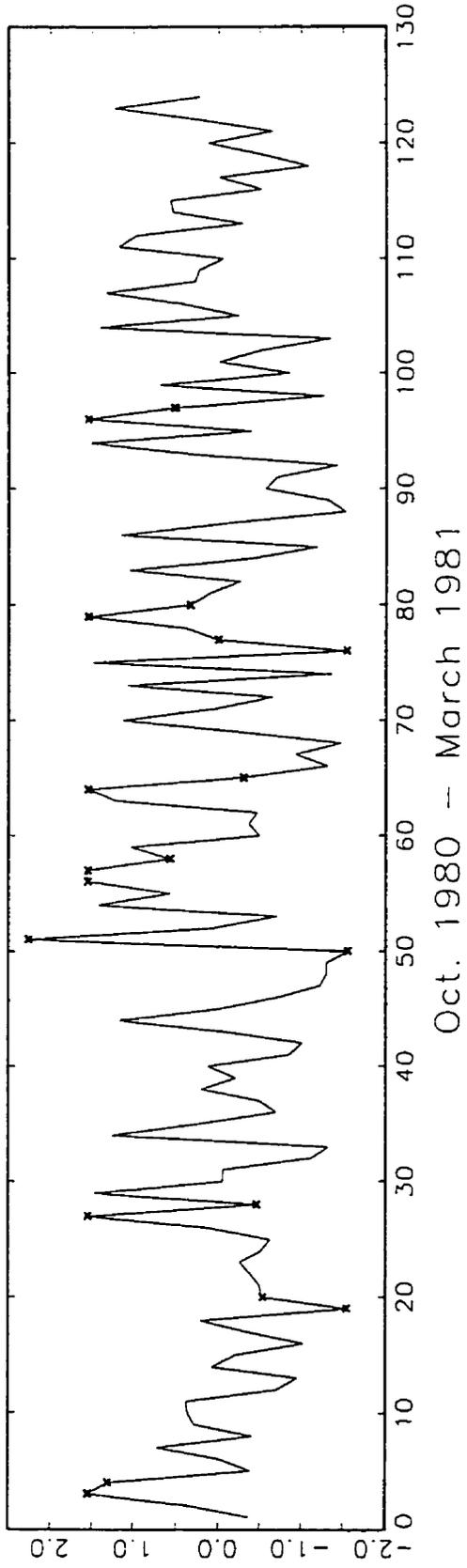


Figure 2. (Continued)

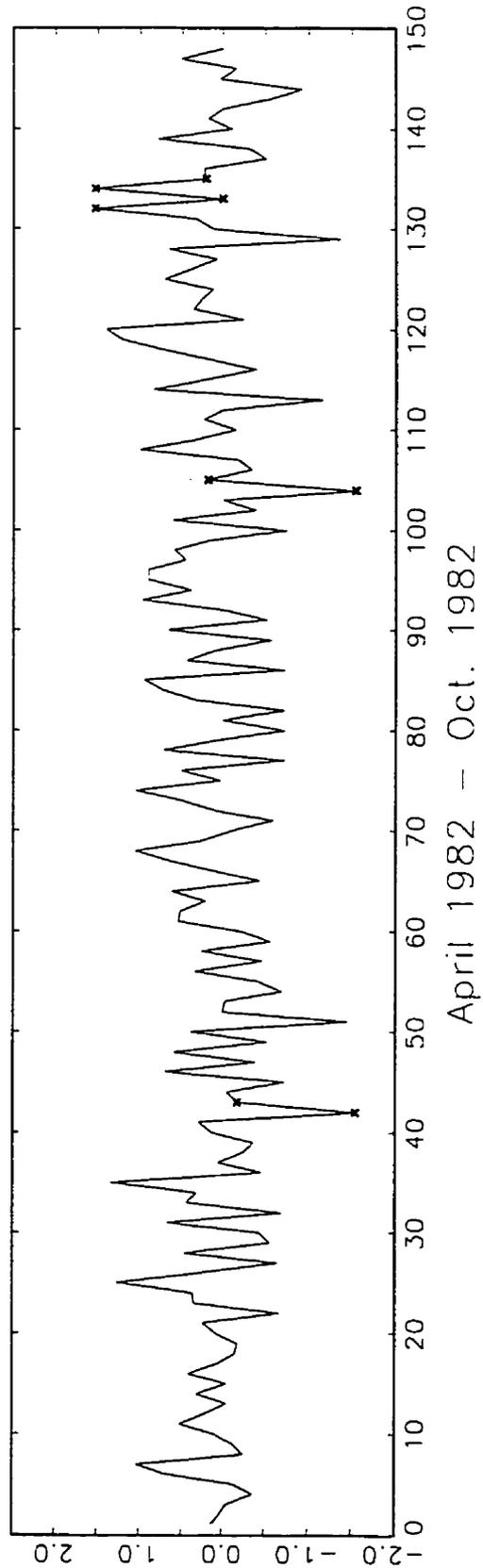
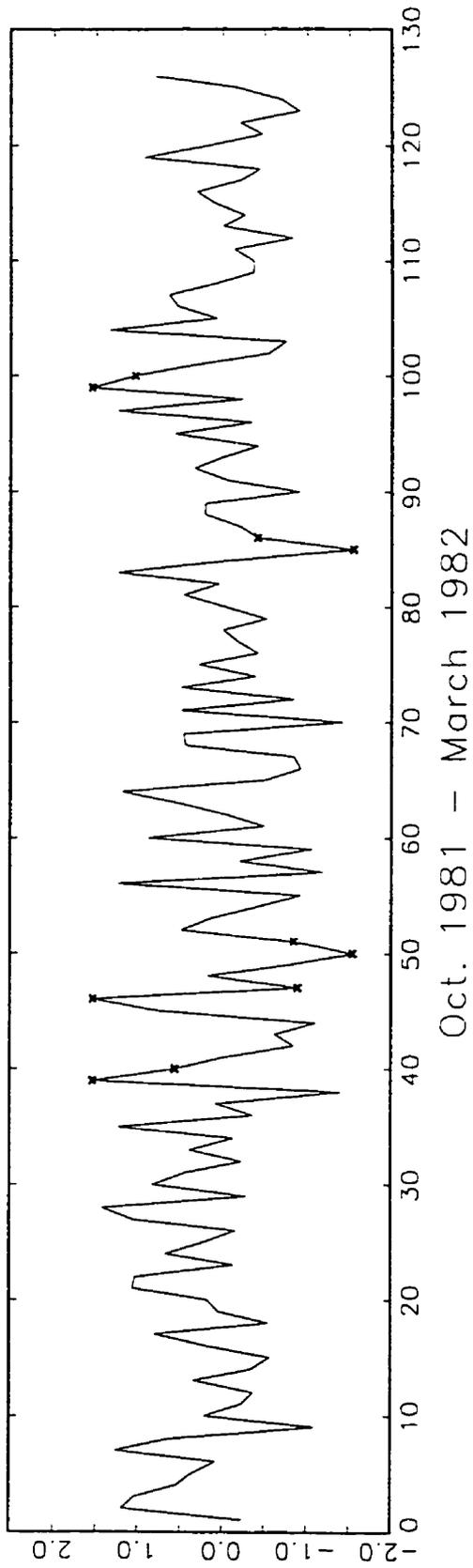
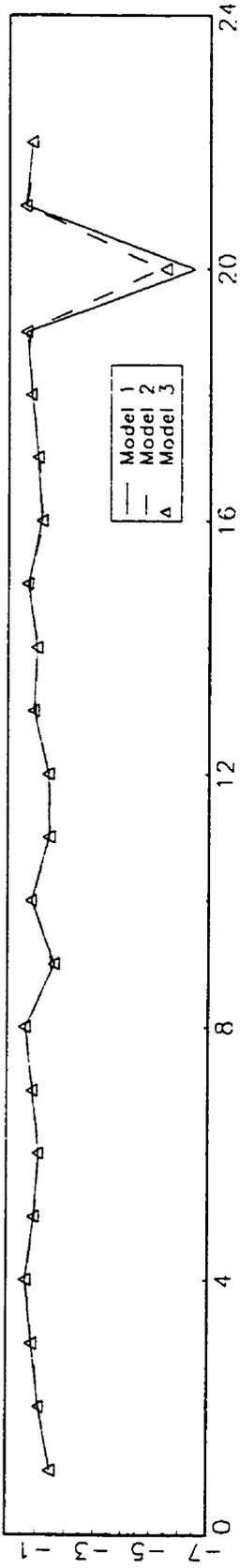
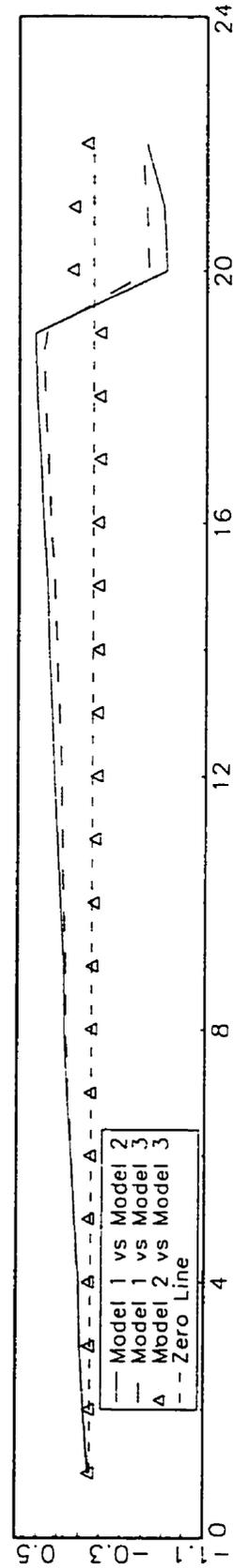


Figure 3.
August 1982
Log Marginalized Likelihoods



Log Predictive Bayes Factors



Return of T-Bill Futures

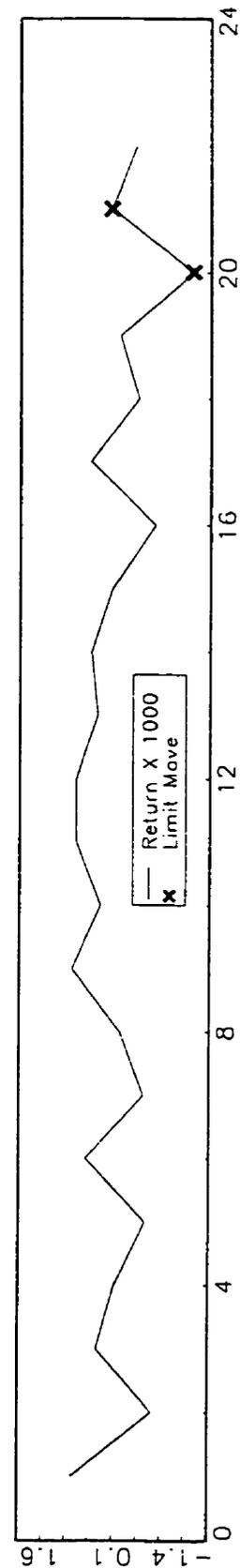
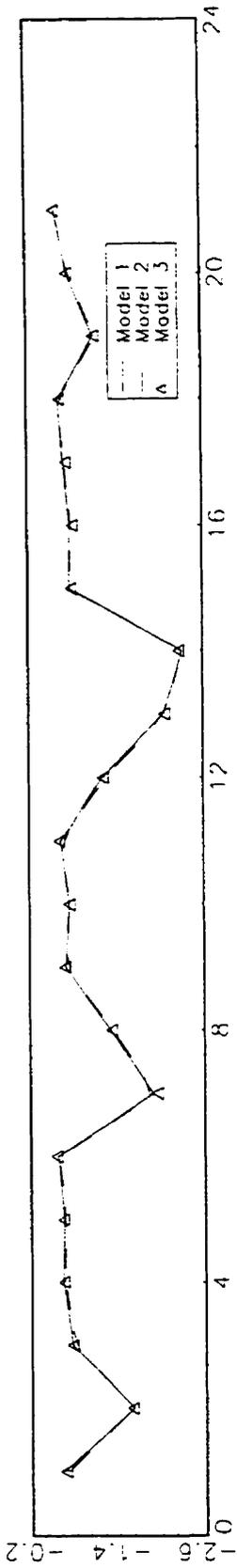
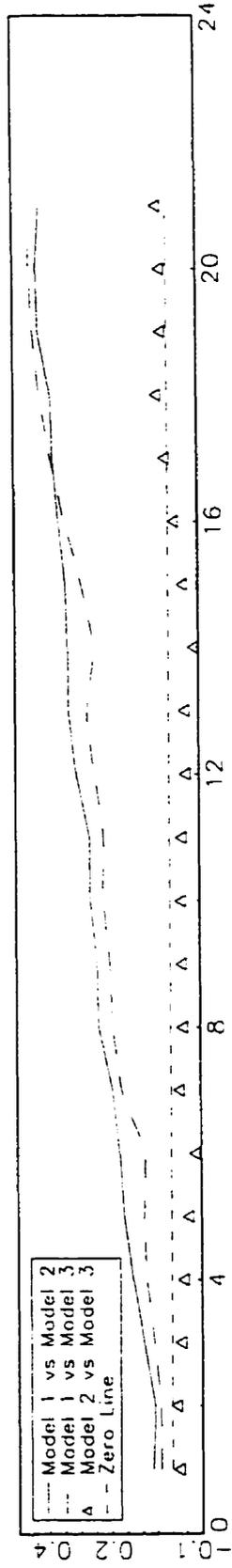


Figure 3.(Continued)
 September 1982
 Log Marginalized Likelihoods



Log Predictive Bayes factors



1--Bill futures Returns

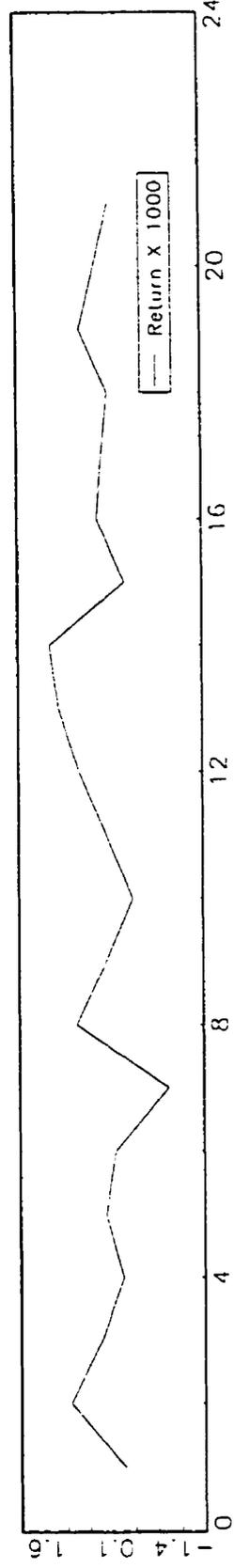
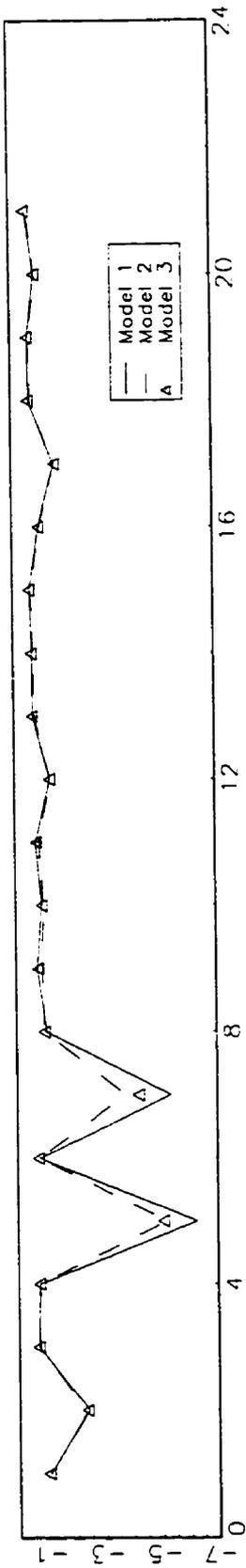
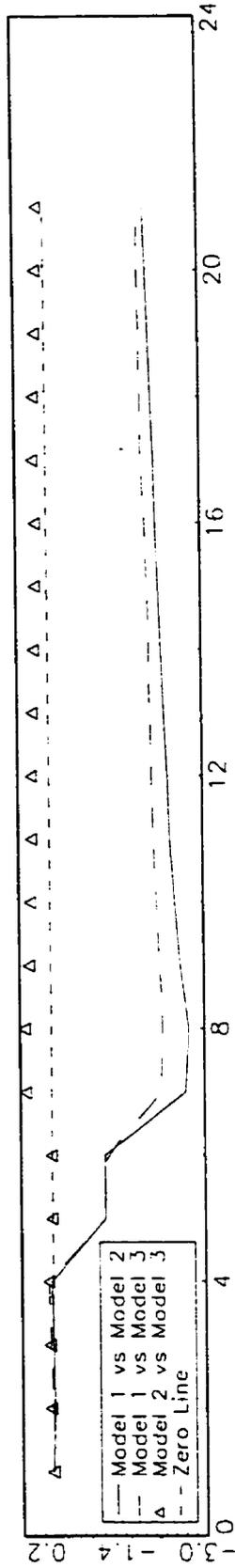


Figure 3. (Continued)
 October 1982
 Log Marginalized Likelihoods



Log Predictive Bayes Factors



Return of 1-Bill Futures

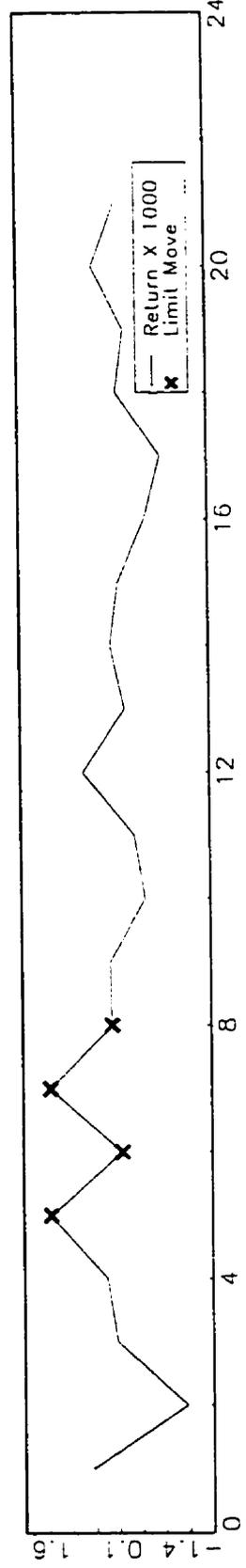
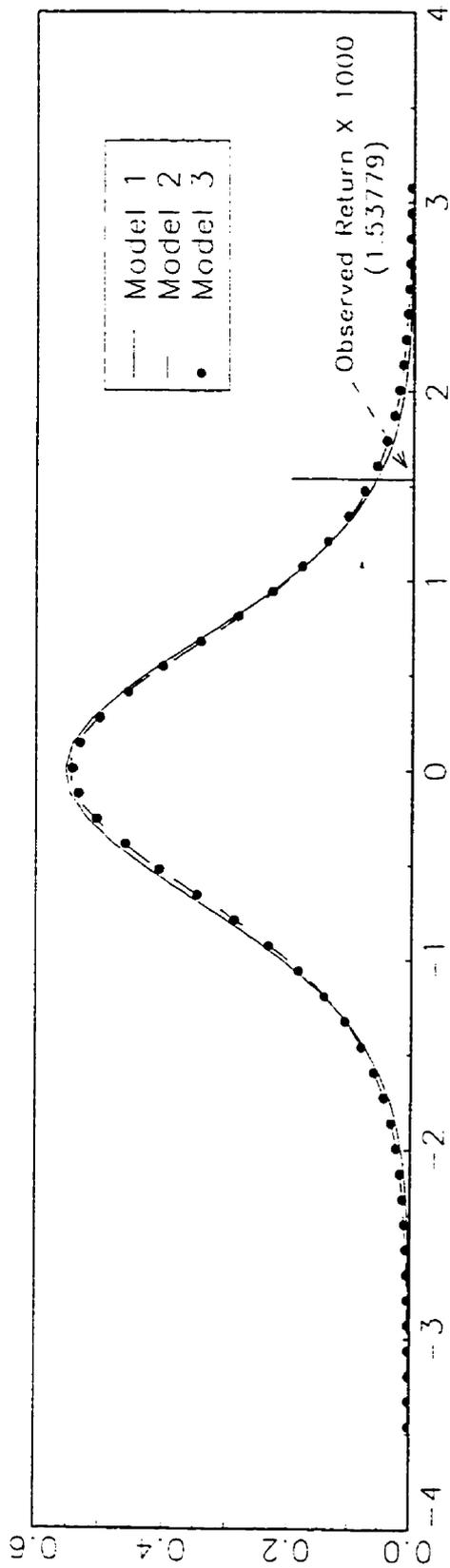


Figure 4. Predictive Density
October 7, 1982



October 7, 1982 (A Tail Plot)

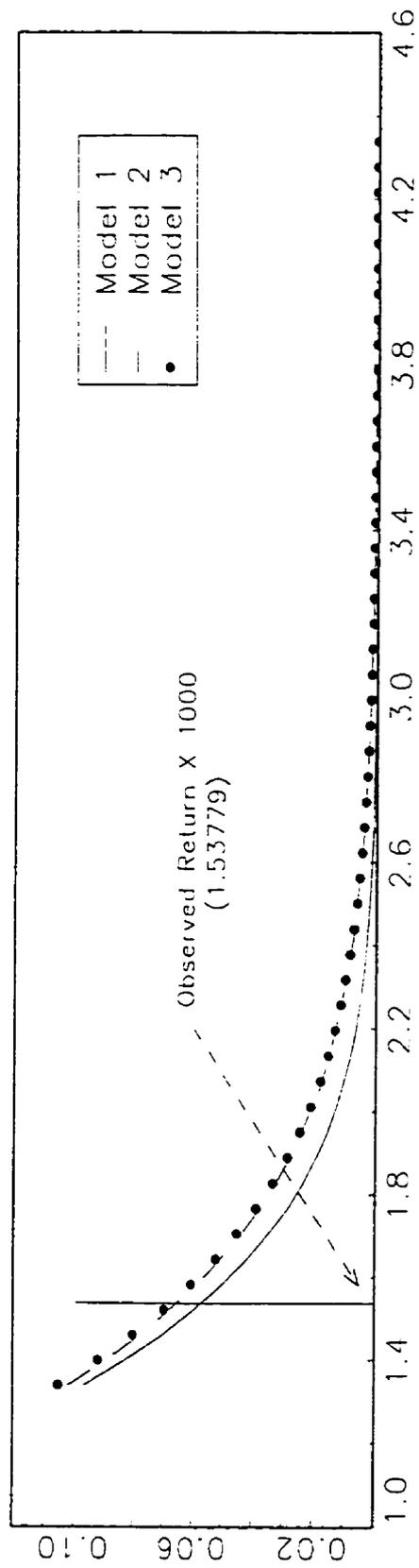
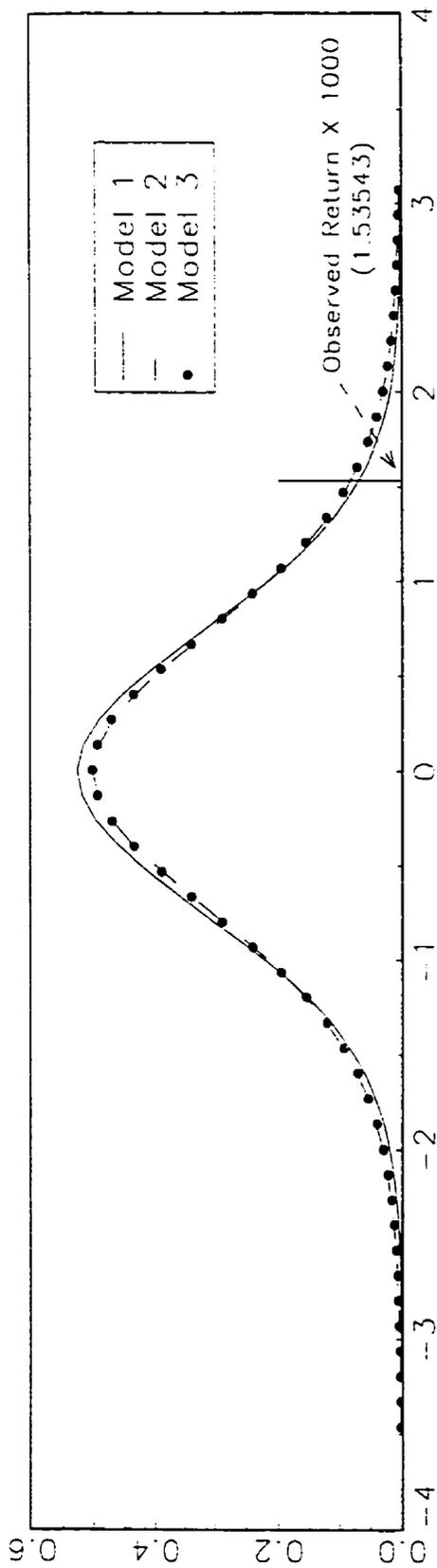


Figure 4. (Continued)
 October 11, 1982



October 11, 1982 (A Tail Plot)

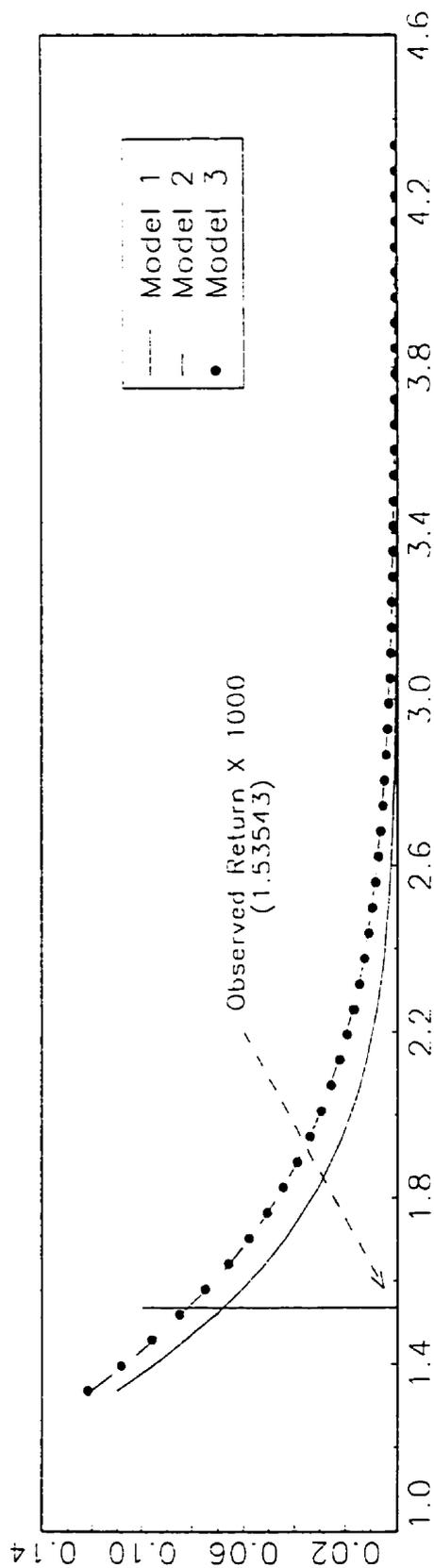
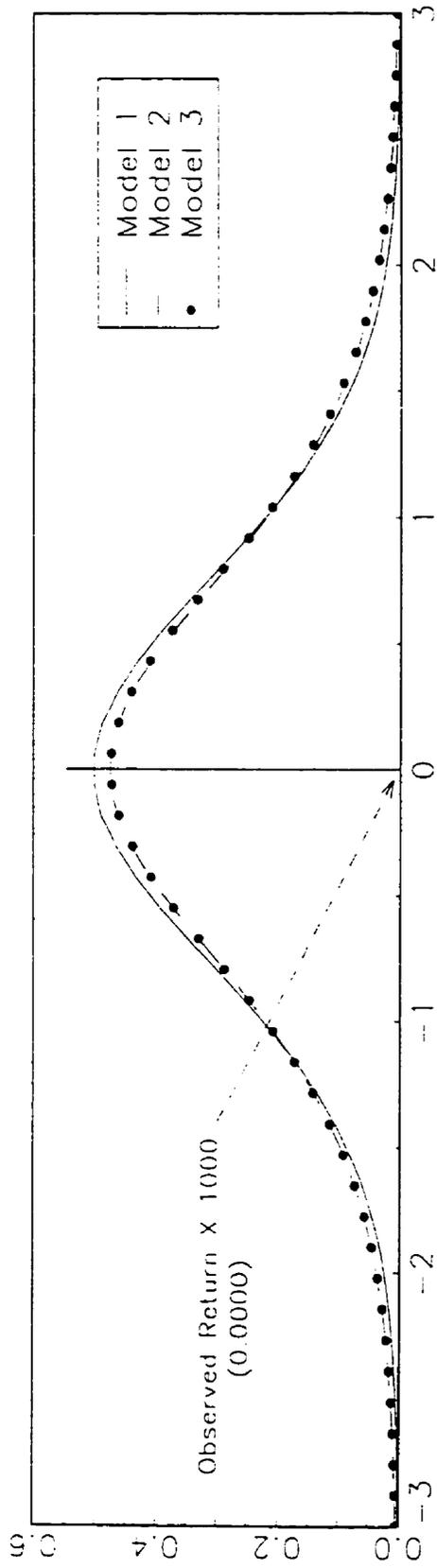


Figure 4. (Continued)
October 8, 1982



October 12, 1982

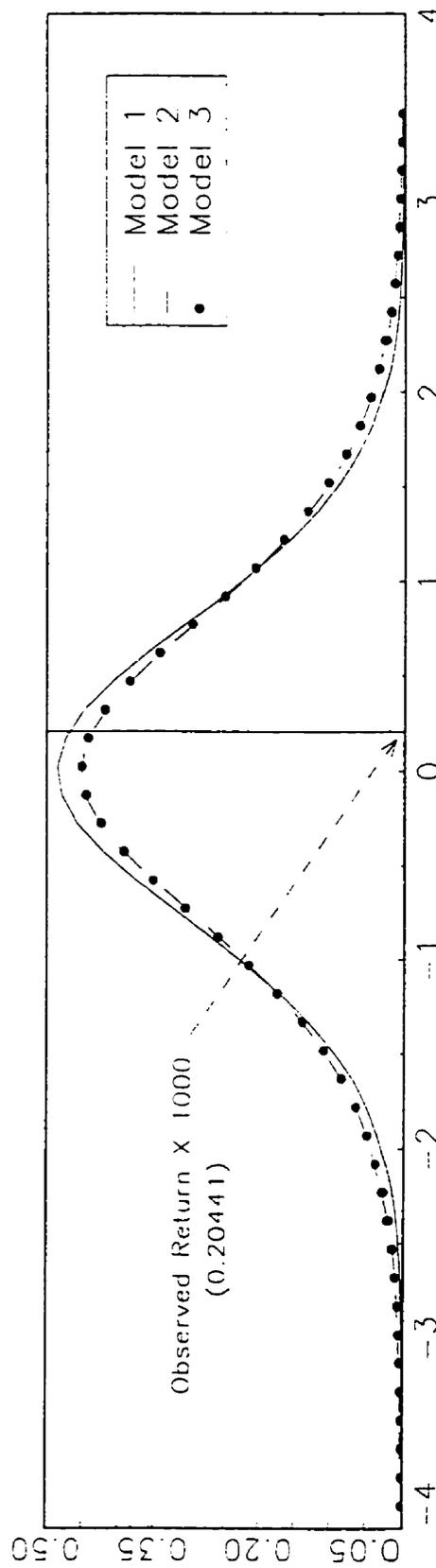
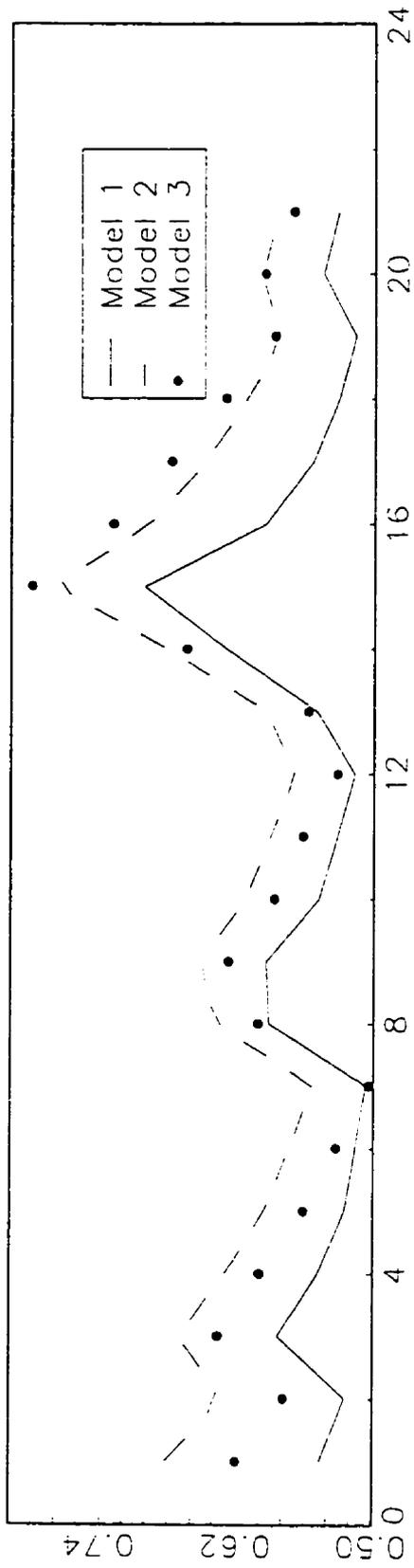
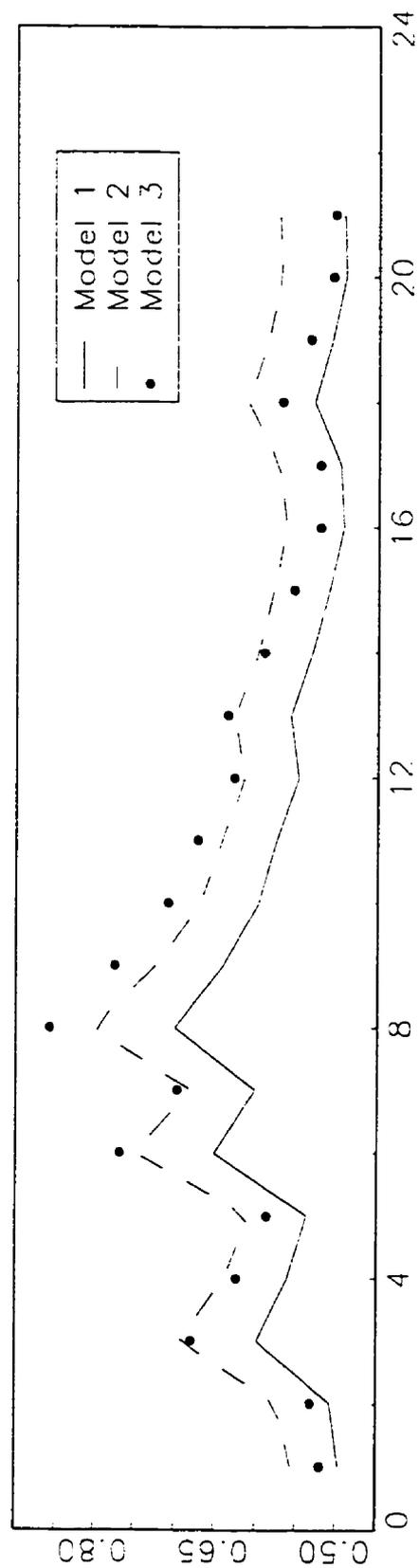


Figure 5. Predictive Variance
September 1982



October 1982



References

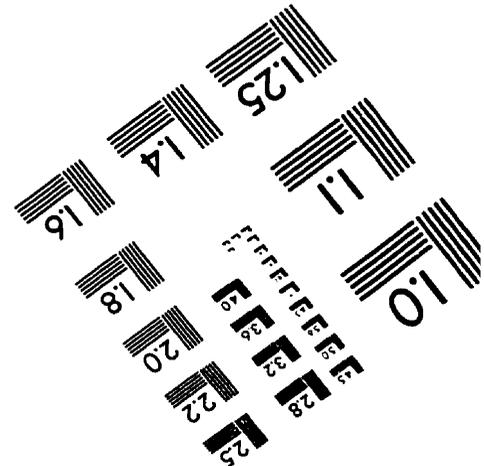
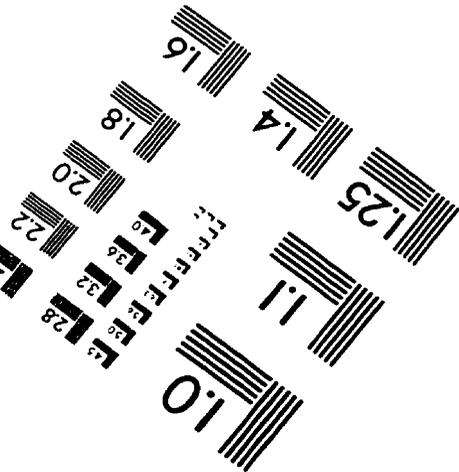
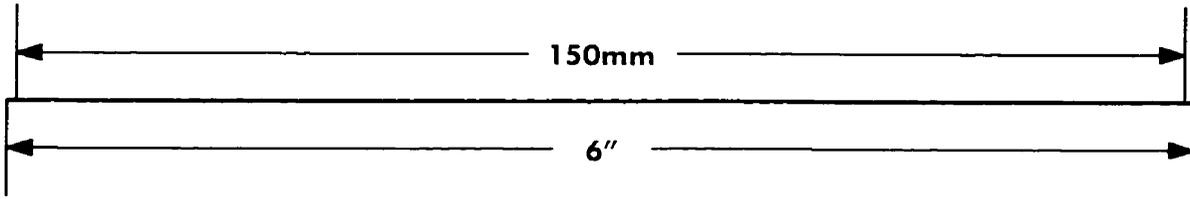
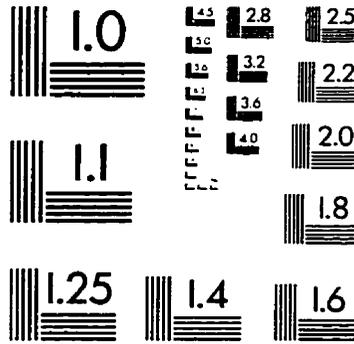
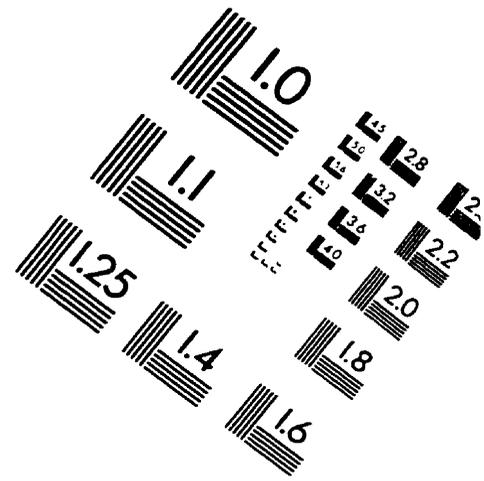
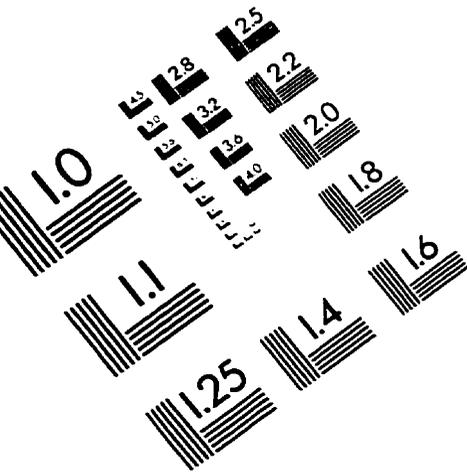
- [1] Amemiya, T. (1985): *Advanced Econometrics*. Cambridge, MA: Harvard University Press.
- [2] Bauwens, L. and Lubrano, M. (1996): "Bayesian Inference on GARCH Models Using the Gibbs Sampler." *Working Paper*. CORE at Université Catholique de Louvain.
- [3] Berger, J. and Pericchi, L.R. (1994): "The Intrinsic Bayes Factor." *Technical Report*. Department of Statistics, Purdue University.
- [4] Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992): "ARCH modeling in Finance." *Journal of Econometrics* **52**, 5-59.
- [5] Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* **31**, 307-327.
- [6] Brennan, M. (1986): "A Theory of Price Limits in Futures Markets." *Journal of Financial Economics* **16**, 213-233.
- [7] Calzolari, G. and Fiorentini, G. (1996): "Tobit Model with GARCH errors." Working Paper, Università di Firenze.
- [8] Chib, S. (1995): "Marginal Likelihood from the Gibbs Output." *Journal of the American Statistical Association* **90**, 1313-1321.
- [9] Davidson, R., and MacKinnon, J.G. (1993): *Estimation and Inference in Econometrics*. New York: Oxford University Press.
- [10] Dumas, B., Fleming, J., and Whaley, R. (1995): "Implied Volatility Functions: Empirical Tests." H.E.C. and Duke Working Paper.
- [11] Engle, R.F. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation." *Econometrica* **50**, 987-1008.

- [12] Evans, M. and Swartz, T. (1995): "Methods for Approximating Integral in Statistics with Special Emphasis on Bayesian Integration Problems." *Statistical Science* Vol. **10**. No. 3. 254-272.
- [13] Gelfand, A.E. (1996): "Model Determination Using Sampling-based Methods." in *Markov Chain Monte Carlo in Practice* eds Gilks W. R., S. Richardson and D. J. Spiegelhalter. 145-158. London: Chapman & Hall.
- [14] Gelfand, A.E. and Dey, D.K. (1994). "Bayesian Model Choice: Asymptotics and Exact Calculations." *Journal of the Royal Statistical Society Series B* **56**: 501-514.
- [15] Gelfand, A.E., Dey, D.K. and Chang, H. (1992). "Model determination using predictive distributions with implementation via sampling-based methods." in *Bayesian Statistics 4* eds Bernardo J. M. J. O. Berger, A. P. Dawid and A. F. M. Smith. 147-167. Oxford: Oxford University Press.
- [16] Gelfand, A.E. and Smith, A.F.M. (1990): "Sampling Based Approaches to Calculating Marginal Densities." *Journal of the American Statistical Association* **85**. 398-409.
- [17] Geweke, J. (1995): "Bayesian Comparison of Econometric Models." University of Minnesota and Federal Reserve Bank of Minneapolis. *Working Paper* **532**.
- [18] Hodrick, R. and Srivastava, S. (1987): "Foreign Currency futures." *Journal of International Economics* **22**. 1 - 24.
- [19] Kass, R. and Raftery, A. (1995): "Bayes factors." *Journal of the American Statistical Association* **90**. 773-795.
- [20] Kodres, L.E. (1988): "Tests of Unbiasedness in Foreign Exchange Futures Markets: The Effects of Price Limits." *Review of Futures Markets* **7**. 138-166.

- [21] Kodres, L.E. (1993): "Tests of Unbiasedness in the Foreign Exchange Futures Markets: An Examination of Price Limits and Conditional Heteroscedasticity." *Journal of Business*, Vol**66**, no. 3, 463-490.
- [22] Koop, G. and Potter, S.M. (1996): "Bayes Factors and Nonlinearity: Evidence from Economic Time Series." *Manuscript*, Department of Economics, University of Toronto.
- [23] Maddala, G.S. (1987): *Limited Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press.
- [24] McCurdy, T.H. and Morgan, I.G. (1987): "Tests of the Martingale Hypothesis for Foreign Currency Futures." *International Journal of Forecasting* **3**, 131-148.
- [25] McCurdy, T.H. and Morgan, I.G. (1988): "Testing the Martingale Hypothesis in Deutsche Mark Futures with Models Specifying the Form of Heteroscedasticity." *Journal of Applied Econometrics* **3**, 187-202.
- [26] Morgan, I.G. and Trevor, R.G. (1996): "Limit Moves as Censored Observations of Equilibrium Futures Prices in GARCH Processes." *Working Paper*, Queen's University
- [27] Poirier, D.J. (1995): *Intermediate Statistics and Econometrics: A Comparative Approach*. MIT Press.
- [28] Ritter, C. and Tanner, M.A. (1992): "Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler." *Journal of the American Statistical Association*, **87**, 861-868.
- [29] Tanner, M.A. and Wong, W.H. (1987): "The Calculation of Posterior Distributions by Data Augmentation." *Journal of the American Statistical Association* **82**, 528-550.
- [30] Tsionas, E.G. (1996): "Bayes Factors for General Likelihoods via Posterior Simulation." *Manuscript*, Department of Economics, University of Toronto.

- [31] Verdinelli. I. and Wasserman. L. (1995): "Computing Bayes factors using a generalization of the Savage-Dickey density ratio." *Journal of the American Statistical Association* **90**, 614-618.
- [32] Wei. S.X. (1995): "A Bayesian Approach to Dynamic Tobit Models". *Manuscript*. Department of Economics, University of Toronto.
- [33] Wei. S.X. (1996): "An Empirical Analysis of the Fed's Open Market Operations" . *Manuscript*. Department of Economics, University of Toronto.
- [34] Yu. B. and Mykland, P. (1994): "Looking at Markov Samplers through Cusum Path Plots: A Simple Diagnostic Idea." *Technical Report 413*. Department of Statistics, University of California at Berkeley.
- [35] Zellner. A. and Min. C. (1995): "Gibbs Sampler Convergence Criteria." *Journal of the American Statistical Association* **90**, 921-927.

IMAGE EVALUATION TEST TARGET (QA-3)



APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved