Gibbs estimation of microstructure models: Teaching notes

Joel Hasbrouck

October, 2006

Program: EMM Gibbs v0100.nb
Bibliography: EMM Gibbs v0100.doc

1. Introduction

This note discusses Gibbs estimation of the Roll model and various modifications. The goal is a more discursive and heuristic treatment of material covered in Hasbrouck (2004, 2006). Other applications of Gibbs samplers in market microstructure include Hasbrouck (1999) and Ball and Chordia (2001).

The techniques discussed here follow an approach that relies on simulation to characterize model parameters. Applied to microstructure models, there are three key elements:

- *Bayesian analysis*
- *Simulation*
- *Characterization of microstructure data generating processes by their conditional probabilities.*

In more detail:

Bayesian analysis

The models are estimated in a Bayesian framework. The differences between Bayesian and classical analysis are continually debated and discussed. The framework here is definitely Bayesian, but it should be noted that even if one doesn't buy the full Bayesian philosophy, the techniques discussed here can be motivated on grounds of estimation simplicity and computational efficiency.
This is an unusual claim. Bayesian analyses are usually more complex (both conceptually and computationally) than their classical counterparts. This is sometimes cited by Bayesian adherents as the prime barrier to their widespread adoption. Most microstructure models, though, are dynamic (over time) and they have latent (hidden, unobservable) quantities. The classic Roll model is dynamic, and the trade direction indicator ("buy or sell") variables are not observed.

Dynamic latent variable models can be formulated in state-space form and estimated via maximum likelihood. The latent variables are often non-Gaussian (e.g., again, the trade direction indicator variables), and if one wants to go beyond the techniques of multivariate linear models (like VARs), estimation involves nonlinear filtering. The Gibbs estimates are usually quicker and simpler.

There are presently a number of Bayesian statistics textbooks available. In my opinion the most useful for financial econometricians, are those that discuss econometrics from a Bayesian perspective. Lancaster (2004) and Geweke (2005) are both excellent. Lancaster's treatment is particularly accessible; Geweke presents more results. Nelson and Kim (2000) is a good introduction to the techniques in the context of a specific problem (regime switching models). In financial econometrics, the heaviest use of Bayesian simulation has been in modeling stochastic volatility. Shephard (2005) is a good survery of this area. Tanner (1996) and Carlin and Louis (2004) consider a broader range of Bayesian statistical tools and applications.

Simulation.

The output of a classical procedure (e.g., OLS) is usually a statement about the distribution of a parameter. E.g., "θ is asympototically normally distributed with mean $\bar{\theta}$ and variance σ_{θ}^2," where the mean and variance quantities are computed directly. But we could also characterize a distribution by a sample of draws from that distribution. This is what most of modern Bayesian analysis does. The output of the estimation procedures discussed here is a stream of random draws of the parameters of interest (conditional on the model and the data). From this stream we can construct an estimate of the full distribution (via kernel smoothing) or simply a summary measure (like the mean or median).

Among other things, simulation facilitates characterization of distributions for functions of random variables. For example, suppose that $x \sim N(\mu, \sigma^2)$ and we'd like to characterize the distribution of $y = f(x)$ where f is sufficiently complicated that we can't get closed-form results. We simply generate random values x_i and empirically examine the distribution of $y_i = f(x_i)$.

The link between simulation and Bayesian analysis is strong for the following reason. The distributions that arise in Bayesian analysis often describe many random variables (i.e., they are of high dimension). It also often happens that they have no closed form representation. Instead, they are characterized by simulation. The Gibbs procedure belongs to a class of random number generators called Markov Chain Monte Carlo (MCMC) techniques. They work by setting up rules for moving from one realization (draw) of the random variables to a subsequent realization. These draws are viewed as "states" in a Markov chain, and the rules define the transition probabilities. The limiting distribution of the states is identical to the distribution of the variables of interest, and is approximated by repeated application of the transition function.

Conditional probabilities

To set up a Gibbs estimate, we need to compute conditional densities for all of the unknowns (parameters *and* latent data). The conditional distributions for the parameters are usually identical to those found in many other applications (e.g., the normal Bayesian linear regression model). This note merely summarizes these distributions, refering the reader elsewhere for a fuller treatment. The conditional distributions for the latent data, though, are specific to the problem at hand. Although not particularly complicated, they are non-standard, and this note covers them in greater detail.

Programs

This note is written in *Mathematica*. Some *Mathematica* code and results are embedded. Most of the results, though, are computed using SAS code that is available on my web site. To obtain these programs, go to the link for *Empirical Market Microstructure* (2006, Oxford University Press), off of my home page at www.stern.nyu.edu/~jhasbrou. Follow the link for SAS programs and data. The unzipped files contain the programs used in this note (and other programs related to the book).

The programs make heavy use of SAS/IML ("Interactive Matrix Language"). This is not the language I've used for most of my papers, but it is widely available. Anyone who has a copy of SAS should be able to run the programs. These programs are not "industrial strength". I've played around with them in generating the results for this note, but they haven't been tested against all the things that might come up in, say, the CRSP daily file. I haven't done any performance benchmarks, but I suspect that they run slower than comparable code in OX or Matlab.

The main programs used here are:

RollGibbs 2-trade case.sas
RollGibbs Analyze q.sas
RollGibbs 01.sas
RollGibbsBeta 01.sas

These programs call two macros: *RollGibbs.sas* and *RollGibbsBeta.sas.* These macros, in turn, make use of IML subroutines contained in a library called "imlstor". To set up this library, run the program *RollGibbs Library 01.sas* (which contains the code for the subroutines).

■ 2. Overview

This note illustrates the estimation approach for the Roll (1984) model of transaction prices. In this model, the "efficient price" (m_t) is assumed to follow a Gaussian random walk:

$$m_t = m_{t-1} + u_t \text{ where } u_t \sim N(0, \sigma_u^2) \quad (1)$$

The transaction price (p_t) is the efficient price, plus or minus a cost component that depends on whether the customer is buying or selling:

$$p_t = m_t + c\, q_t \quad (2)$$

where c is the cost parameter and $q_t = \pm 1$. (If the customer is buying, $q_t = +1$; if selling, $q_t = -1$). The trade prices are observed. The q_t and the m_t are not.
By taking first differences:

$$\Delta p_t = c\,\Delta q_t + u_t \tag{3}$$

This specification is important because if the Δq_t were known, this would be a simple regression.

- Bayesian estimation of normal linear regressions is well understood. The discussion (in the next section) starts with a review of these procedures.
- There are two parameters in this "regression": c (the coefficient) and σ_u^2. It is fairly easy to compute (in closed form) the posterior distributions $f(c \mid \sigma_u^2,\ p_1,\ \ldots,\ p_T)$ and $f(\sigma_u^2 \mid c,\ p_1,\ \ldots,\ p_T)$.
- It is not possible to compute in closed form the joint posterior $f(c,\ \sigma_u^2 \mid p_1,\ \ldots,\ p_T)$. This motivates the next section, which summarizes the Gibbs sampler.
- The Gibbs procedure is illustrated by applying it to a special case of the Roll model, one in which c and σ_u^2 are known, but the q_t are not.
- The note then turns to a full estimation of the Roll model ...
- ... and extensions.

3. *Mathematica* initializations

4. Bayesian analysis of the normal linear regression

The basic Bayesian approach

Bayesian analysis consists of using a model and data to update prior beliefs. The revised beliefs are usually called posterior beliefs, or simply "the posterior". Let y denote the observed data, and let the model be specified up to a parameter Θ (possibly a vector). Bayes theorem says:

$$f(\Theta \mid y) = \frac{f(\Theta, y)}{f(y)} = \frac{f(y \mid \Theta)\, f(\Theta)}{f(y)} \propto f(y \mid \Theta)\, f(\Theta) \tag{4}$$

$f(\Theta)$ is an assumed prior distribution for the parameter.
$f(y \mid \Theta)$ is the likelihood function for the observations, given a particular value of Θ.
The use of $\propto$ ("is proportional to") reflects the fact that it is usually not necessary to compute $f(y)$, at least not by computing the marginal $f(y) = \int f(\Theta,\ y)\, d\Theta$.
Instead, $f(y)$ can treated as a normalization constant, set so that the posterior integrates to unity.

Often a distribution of interest, say $f(x)$, can be written as $f(x) = k\, g(x)$, where $g(x)$ is a parsimonious function of x and k is some scaling factor. k might in fact be very complicated, possibly depending on other random variables and implicitly incorporating other distribution functions, but for purposes of characterizing the distribution of x, it is constant. In this case, $g(x)$ is said to be the *kernel* of $f(x)$.

Bayesian estimation of the normal linear regression model

The normal regression model is:

$$\underset{N\times 1}{y} = \underset{N\times K}{X}\,\underset{K\times 1}{\beta} + \underset{N\times 1}{u} \text{ where } u \sim N(0, \Omega_u) \tag{5}$$

X is a matrix of covariates (explanatory variables) possibly including a constant; β is the coefficient vector.

Estimation of coefficients (given the error variance)

Assume for the moment that σ_u^2 is known. It is particularly convenient to assume a multivariate normal prior distribution for the coefficients:

$$\beta \sim N(\mu_\beta^{Prior}, \Omega_\beta^{Prior}) \tag{6}$$

The posterior distribution, $f(\beta \mid y)$ is

$$N(\mu_\beta^{Post}, \Omega_\beta^{Post}) \tag{7}$$

where

$$\mu_\beta^{Post} = D\,d \tag{8}$$

$$\Omega_\beta^{Post} = D^{-1} \tag{9}$$

$$D^{-1} = X' \Omega_u^{-1} X + (\Omega_\beta^{Prior})^{-1} \tag{10}$$

$$d = X' \Omega_u^{-1} y + (\Omega_\beta^{Prior})^{-1} \mu_\beta^{Prior} \tag{11}$$

As Ω_β^{Prior} increases in magnitude, the posterior mean and variance converge toward the usual OLS values.
In this case, both the prior and posterior have the same form (multivariate normal). Such a prior is said to be conjugate.

Simulating the coefficients

We'll often have to make a random draw from the coefficient distribution. To make a random draw from $\underset{n\times 1}{x} \sim MVN(\mu, \Omega)$:

- Compute the Cholesky factorization $F : \Omega = F' F$, where F is an upper triangular matrix.
- Draw $\underset{n\times 1}{z}$ where the z_i are i.i.d. $N(0, 1)$.
- Set the random draw as $x = \mu + F' z$

Restrictions on the prior

The economic model sometimes imposes bounds on the coefficients. For example, in the Roll model, we'll often want to require $c > 0$. Suppose that the coefficient prior is

$$\beta \sim N(\mu_\beta^{Prior}, \Omega_\beta^{Prior}),\ \underline{\beta} < \beta < \bar{\beta} \tag{12}$$

Note that when we write this, $\mu_\beta^{Prior}, \Omega_\beta^{Prior}$ denote the formal parameters of the normal density. But since the distribution is truncated, they no longer denote the mean and covariance of the density.

With this prior, the posterior is simply $N(\mu_\beta^{Post}, \Omega_\beta^{Post})$, with μ_β^{Post} and Ω_β^{Post}computed as described above, restricted to the space $\underline{\beta} < \beta < \bar{\beta}$.

Simulation from restricted normals.

First suppose that we want to make a random draw z from a standard normal density, restricted to the interval $\underline{z} < z < \bar{z}$. The procedure is:

- Compute $\underline{\pi} = \Phi(\underline{z})$ and $\bar{\pi} = \Phi(\bar{z})$, where Φ is the c.d.f. of the standard normal.
- Draw u from the uniform distribution over $(\underline{\pi}, \bar{\pi})$
- Set $z = \Phi^{-1}(u)$.

Now suppose that we want to make a bivariate random draw from
$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim N(\mu, \Omega)$, $\underline{x} < x < \bar{x}$.

- Compute the Cholesky factorization $F : \Omega = F' F$, where F is an upper triangular matrix.
- Set $\underline{z} = \frac{\underline{x}_1 - \mu_1}{F_{11}}$ and $\bar{z} = \frac{\bar{x}_1 - \mu_1}{F_{11}}$
- Draw z_1 from the standard normal density, restricted to $(\underline{z}, \bar{z})$. Then $x_1 = \mu_1 + F_{11} z_1$ will have the properties required of x_1.
- Set $\eta = F_{11} z_1$.
- Set $\underline{z} = \frac{\underline{x}_2 - \mu_2 - \eta}{F_{22}}$ and $\bar{z} = \frac{\bar{x}_2 - \mu_2 - \eta}{F_{22}}$
- Draw z_2 from the standard normal density, restricted to $(\underline{z}, \bar{z})$. Then $x_2 = \mu_2 + F_{22} z_2$ will have the properties required of x_2.
- The random draw as $x = \mu + F' z$ will have the required joint properties

This method may be generalized to higher dimensions.

Estimation of error variance (given the coefficients)

Assuming that β is known, it is convenient to specify a inverted gamma prior for σ_u^2. One way of writing this is:

$$\frac{1}{\sigma_u^2} \sim \Gamma[a^{Prior}, b^{Prior}] \qquad (13)$$

Then the posterior is

$$\frac{1}{\sigma_u^2} \mid y \sim \Gamma[a^{Post}, b^{Post}] \qquad (14)$$

where

$$a^{Post} = a^{Prior} + \frac{N}{2} \text{ and } b^{Post} = b^{Prior} + \frac{\sum_{i=1}^{N} u_i^2}{2} \qquad (15)$$

The u_i are the regression residuals $u = y - X\beta$.

Further notes

The density of a gamma variate x with parameters a and b is:

```
PDF[GammaDistribution[a, λ], x] /. λ → 1 / b
```

$$\frac{\left(\frac{1}{b}\right)^{-a} e^{-b x} x^{-1+a}}{\text{Gamma}[a]}$$

```
Mean[GammaDistribution[a, λ]] /. λ → 1 / b
```

$$\frac{a}{b}$$

Note: In the statistics literature, the Gamma distribution with parameters a and b is usually expressed as immediately above. *Mathematica* parameterizes the distribution with the second parameter expressed as an inverse.

```
PDF[GammaDistribution[a, λ], x] /. λ → 1 / b /. x → 1 / z
```

$$\frac{\left(\frac{1}{b}\right)^{-a} e^{-\frac{b}{z}} \left(\frac{1}{z}\right)^{-1+a}}{\text{Gamma}[a]}$$

■ 5. The Gibbs recipe

The Gibbs procedure is a vehicle for simulating from a complicated joint distribution $f(x_1, \ldots, x_n)$, possibly one that possesses no closed form representation.

The draws are constructed by iterating over the *full conditional distributions*:

$$\begin{array}{c} f(x_1 \mid x_2, \ldots, x_n) \\ f(x_2 \mid x_1, x_3, \ldots, x_n) \\ \ldots \\ f(x_n \mid x_1, \ldots, x_{n-1}) \end{array} \tag{16}$$

Let $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$

Each iteration of the Gibbs sampler is called a sweep.
Let $x^{[i]}$ denote the value of x at the conclusion of the ith sweep.
The procedure is:

- Initialization. Set $x^{[0]}$ to any feasible value.
- Sweep i:
  Given $x^{[i-1]}$:
  Draw $x_1^{[i]}$ from $f\left(x_1 \mid x_2^{[i-1]}, \ldots, x_n^{[i-1]}\right)$
  Draw $x_2^{[i]}$ from $f\left(x_2 \mid x_1^{[i-1]}, x_3^{[i-1]}, \ldots, x_n^{[i-1]}\right)$
  ...
  Draw $x_n^{[i]}$ from $f\left(x_n \mid x_1^{[i-1]}, \ldots, x_{n-1}^{[i-1]}\right)$
- Repeat

In the limit, as $i \to \infty$, $x^{[i]}$ is distributed as $f(x)$.

Notes

- The $x^{[i]}$ are *not* independently distributed: $x^{[i]}$ takes $x^{[i-1]}$ as its starting point. If the degree of dependence is high, a large number of sweeps may be needed to ensure proper mixing.
- The dependence may affect the calculation of some summary statistics. Think of the analogy to standard time series analysis. If $z_1, \ldots, z_T$ are a sample of stationary time series data, $\Sigma z_i / T$ is a consistent estimate of $E[z_t]$. The standard error of this estimate, however, must be corrected for dependence.
- Convergence may be an issue. It is useful to graph the full sequence of draws.
- In analyzing the sequence of draws, it is common to throw out a few initial draws, so as to reduce the dependence on starting values. These discarded draws are sometimes called *burn in* draws.

- The Gibbs sampler also works when multiple variables are drawn at once. We might, for example, draw $x_1^{[i]}$ and $x_2^{[i]}$ from $f\left(x_1, x_2 \mid x_3^{[i-1]}, \ldots, x_n^{[i-1]}\right)$. This *block sampling* is often more computationally efficient.

Application to the normal regression model

From earlier results, we have $f(\beta \mid y, \sigma_u^2)$ and $f(\sigma_u^2 \mid \beta, y)$. To obtain the full posterior $f(\beta, \sigma_u^2 \mid y)$ via the Gibbs procedure:
Initialize $\sigma_u^{2[0]}$ to any feasible value. The ith sweep of the sampler is:

- Draw $\beta^{[i]}$ from $f(\beta \mid y, \sigma_u^{2[i-1]})$. (This will be a draw from a multivariate normal posterior.)
- Draw $\sigma_u^{2[i]}$ from $f[\sigma_u^2 \mid y, \beta^{[i]})$. (That is, draw $1/\sigma_u^{2[i]}$ from the gamma posterior.)

Proceed, iteratively drawing $\sigma_u^{2[i]}$ and $\beta^{[i]}$.

Notes

- The $f(\beta, \sigma_u^2 \mid y)$ is an exact small-sample distribution.

We now return to ...

■ 6. The Roll model

Recall that we're working with the price change specification:

$$\Delta p_t = c\,\Delta q_t + u_t \tag{17}$$

The sample is $p_1, p_2, \ldots, p_T$, and there are $T-1$ price changes.

The unknowns are the parameters c and σ_u^2, and the latent data: $q_1, \ldots, q_T$.
In the Bayesian perspective, parameters and latent data are treated identically, and "estimated" in a similar fashion.
We don't need to construct priors for the q_t. We can use the ones we assumed in the data generating process: $q_t = \pm 1$ with equal probabilities.
The prior on c is $c \sim N(\mu^{Prior}, \Omega^{Prior})$ restricted to $c > 0$. (I often take $\mu^{Prior} = 0$ and $\Omega^{Prior} = 1$. Remember that these are parameters of the truncated distribution, not the true mean and variance.)
The prior on σ_u^2 is inverted gamma with parameters a and b. (I often take $a = b = 10^{-6}$.)
The Gibbs sampler will look like this:

- Initialize $c^{[0]}$, $\sigma_u^{2[0]}$ and $q_1^{[0]}, \ldots, q_T^{[0]}$ to any feasible values. (I usually take $q_1 = 1$; $q_t = \text{Sign}(\Delta p_t)$ if $\Delta p_t \neq 0$, $q_t = q_{t-1}$ if $\Delta p_t = 0$. For US equities, $c = 0.01$ ("1%") is a good ballpark figure, if we're working in logs, and $\sigma_u^2 = 0.01^2$).

For the ith sweep of the sampler:

- Estimate the price change specification as a regression, assuming that $q_t = q_t^{[i-1]}$ and that $\sigma_u^2 = \sigma_u^{2[i-1]}$. Construct the posterior for c, and draw $c^{[i]}$ from this posterior.
- Using $c^{[i]}$, compute the residuals from the regression. Construct the posterior for σ_u^2 and draw $\sigma_u^{2[i]}$ from this posterior.

- Draw $q_1^{[i]}$ from $f\left(q_1 \mid c^{[i]}, \sigma_u^{2[i]}, q_2^{[i-1]}, q_3^{[i-1]}, \ldots, q_T^{[i-1]}\right)$
  Draw $q_2^{[i]}$ from $f\left(q_2 \mid c^{[i]}, \sigma_u^{2[i]}, q_1^{[i]}, q_3^{[i-1]}, \ldots, q_T^{[i-1]}\right)$
  ...
  Draw $q_T^{[i]}$ from $f\left(q_T \mid c^{[i]}, \sigma_u^{2[i]}, q_1^{[i]}, q_2^{[i]}, \ldots, q_{T-1}^{[i]}\right)$

The first two steps of the sweep are, as discussed, standard Bayesian procedures. We now turn to the third step: simulating the trade direction indicators

7. Estimating the q_t using a Gibbs sampler.

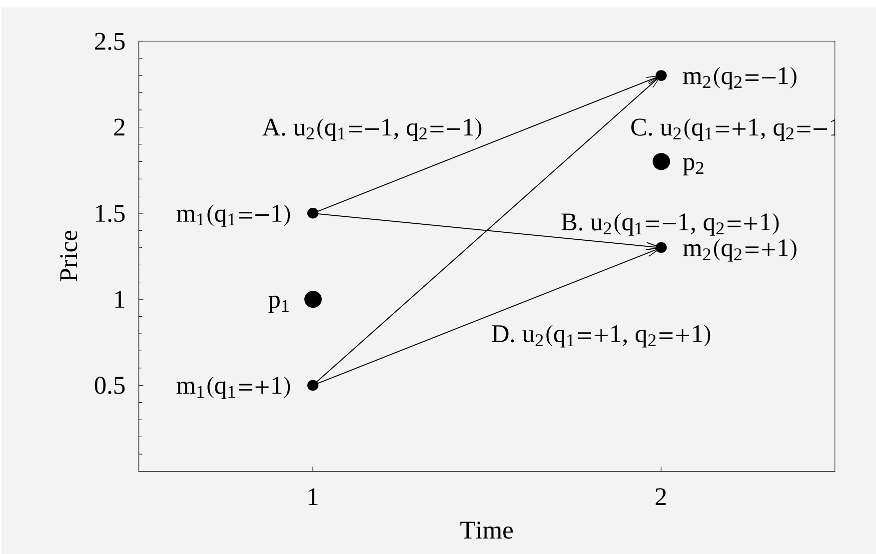
In this section, we'll be taking c and σ_u^2 as known. We'll first look at the simple case where $T = 2$. We can get closed-form results here, so we don't really need a Gibbs procedure, but it's a good starting point.

The distribution of q_1 and q_2 when T=2

Graph

Suppose:

```
p1 = 1; p2 = 1.8; c = .5;
```



Given p_1 and c, a choice of q_1 determines m, and similarly for p_2.
Therefore, setting q_1 and q_2 fixes $u_2 = m_2 - m_1$.
Since $q_1, q_2 \in \{-1, +1\}$, there are four possible paths (values) for u_2. These are labeled A, B, C and D in the figure. Since low values of $|u_2|$ are more likely than high values, the relative lengths of these paths indicate the relative likelihood of the (q_1, q_2) realizations that determined them:

- Intuitively, path B is the shortest, so it is most likely that $q_1 = -1, q_2 = +1$.
- Path C is the longest, so it is least likely that $q_1 = +1, q_2 = -1$.

- Paths A and D are of equal length, corresponding to the realizations $q_1 = q_2 = -1$ and $q_1 = q_2 = +1$.

We now turn to a more exact treatment.

The joint density of q_1 and q_2

The density function for u is $f(u_2) =$

$$\frac{e^{-\frac{u^2}{2\sigma_u^2}}}{\sqrt{2\pi}\,\sigma_u}$$

By rearranging the price change specification, $u_2 =$

```
u2[q1, q2]
```

$-c\,(-q_1 + q_2) + \triangle p_2$

The probability $\mathrm{Pr}[q_1, q_2] \propto f(u_2[q_1, q_2])$

```
Pr[q1, q2] /. PrRule
```

$$\frac{e^{-\frac{(-c\,(-q_1+q_2)+\triangle p_2)^2}{2\sigma_u^2}}}{\sqrt{2\pi}\,\sigma_u}$$

The possible outcomes are:

```
Outcomes = {{-1, -1}, {-1, +1}, {+1, -1}, {+1, +1}};
TableForm[Outcomes, TableHeadings → {None, {q1, q2}}]
```

q_1	q_2
−1	−1
−1	1
1	−1
1	1

We'll normalize by the sum of the probabilities:

```
PrSum = Plus @@ Apply[Pr, Outcomes, {1}]
```

Pr[−1, −1] + Pr[−1, 1] + Pr[1, −1] + Pr[1, 1]

```
PrNRule = PrN[q1 : _, q2 : _] :→ Pr[q1, q2] / PrSum;
```

The normalized probability is:

```
PrN[q1, q2] /. PrNRule
```

$$\frac{\mathrm{Pr}[q_1, q_2]}{\mathrm{Pr}[-1, -1] + \mathrm{Pr}[-1, 1] + \mathrm{Pr}[1, -1] + \mathrm{Pr}[1, 1]}$$

For demonstration purposes, here are some values:

```
nValues = {Δp2 → .8, σu → 1, c → 0.5};
```

With thes values, the normalized probabilities are:

q_1	q_2	Probability
−1	−1	0.276061
−1	1	0.372643
1	−1	0.0752353
1	1	0.276061

Gibbs sampler

We can draw the trade direction indicator variables jointly in this case. There's no need to use a Gibbs sampler. But for illustration purposes, let's build it.

The required conditional probabilities are $\Pr[q_1 \mid q_2]$ and $\Pr[q_2 \mid q_1]$. These may be computed directly from joint distribution given above, but it is usually computationally easier to work with the odds ratio.
For example, $u_2[q_1, q_2] =$

$$-c\,(-q_1 + q_2) + \Delta p_2$$

Its density is:

$$\frac{e^{-\frac{(-c\,(-q_1+q_2)+\Delta p_2)^2}{2\,\sigma_u^2}}}{\sqrt{2\,\pi}\,\sigma_u}$$

So given q_2, the odds in favor of a buy at time 1 are $\text{Odds}\,(\text{Buy}) = \frac{\Pr[q_1=+1|\ldots]}{\Pr[q_1=-1|\ldots]} =$

$$\textbf{OddsBuy1} = \frac{\textbf{f /. }q_1 \to +1}{\textbf{f /. }q_1 \to -1} \;\textbf{// Simplify}$$

$$e^{\frac{2\,c\,(c\,q_2-\Delta p_2)}{\sigma_u^2}}$$

Then $\Pr[\text{Buy}] = \frac{\text{Odds}\,(\text{Buy})}{1+\text{Odds}\,(\text{Buy})}$. We compute this probability and make a draw for q_1.

For the particular numeric values we worked with above $(\Delta p_2 = 0.8,\ c = 0.5,\ \sigma_u = 1)$, these odds, for $q_2 = +1$ and $q_2 = -1$ are:

$$\{0.740818,\ 0.272532\}$$

So, for example, if $q_2 = +1$, $\Pr[q_1 = +1 \mid \ldots] =$

$$0.425557$$

Similarly, given q_1, the odds in favor of a buy at time 2 are $\frac{\Pr[q_2=+1|\ldots]}{\Pr[q_2=-1|\ldots]} =$

$$\textbf{OddsBuy2} = \frac{\textbf{f /. }q_2 \to +1}{\textbf{f /. }q_2 \to -1} \;\textbf{// Simplify}$$

$$e^{\frac{2\,c\,(c\,q_1+\Delta p_2)}{\sigma_u^2}}$$

For the numeric values, these are (for $q_1 = +1$ and $q_1 = -1$):

$$\{3.6693,\ 1.34986\}$$

The Gibbs sampler involves the following steps:

We construct a series of realizations in the following fashion:

Initialize $q_1^{[0]},\ q_2^{[0]} = \pm 1$ (it doesn't matter which).

The ith sweep involves the following steps:

- Draw $q_1^{[i]}$ from $\Pr\left[q_1 \mid q_2^{[i-1]}\right]$
- Draw $q_2^{[i]}$ from $\Pr\left[q_2 \mid q_1^{[i]}\right]$

After *N* sweeps we'll have a series of *N* simulated realizations:
$q^{[0]},\ q^{[1]},\ \ldots,\ q^{[N]}$ where $q^{[i]} = \left(q_1^{[i]},\ q_2^{[i]}\right)$.
In the limit, as $N \to \infty$, the distribution of $q^{[N]}$ is the joint distribution $\Pr[q_1, q_2]$

The sampler was implemented in a SAS program (*RollGibbs 2-trade case.sas*), which was run for 10,000 sweeps. The tabulated frequencies of the simulated draws were:

	N	PctN
outcome		
q1= 1 q2= 1	2,785	27.9
q1= 1 q2=-1	752	7.5
q1=-1 q2= 1	3,751	37.5
q1=-1 q2=-1	2,712	27.1
All	10,000	100.0

Compare these to the computed probabilities above.

Why do we ever need to use Gibbs sampler when we can compute path probabilities directly?

We need to compute path probabilities for the entire sample. With two price observations, there are $2^2 = 4$ buy/sell paths.
A year contains about 250 trading days. The number of buy/sell paths is:

$$2^{250}$$

```
1809251394333065553493296640760748560207343510400633813116524\
750123642650624
```

◻ The T-trade case

A sweep of the sampler will involve:

- Draw $q_1^{[i]}$ from $\Pr\left(q_1 \mid q_2^{[i-1]}, q_3^{[i-1]}, \ldots, q_T^{[i-1]}\right)$
  Draw $q_2^{[i]}$ from $\Pr\left(q_2 \mid q_1^{[i]}, q_3^{[i-1]}, \ldots, q_T^{[i-1]}\right)$
  ...
  Draw $q_T^{[i]}$ from $\Pr\left(q_T \mid q_1^{[i]}, q_2^{[i]}, \ldots, q_{T-1}^{[i]}\right)$

In general, $\Pr[q_t \mid \ldots]$ depends only on the adjacent trades - those at times $t-1$ and $t+1$. So the sampler becomes:

- Draw $q_1^{[i]}$ from $\Pr\left(q_1 \mid q_2^{[i-1]}\right)$
  Draw $q_2^{[i]}$ from $\Pr\left(q_2 \mid q_1^{[i]}, q_3^{[i-1]}\right)$
  ...
  Draw $q_T^{[i]}$ from $\Pr\left(q_T \mid q_{T-1}^{[i]}\right)$

The first draw, for q_1, is the same as the draw for q_1 in the $T = 2$ case.

The last draw, for q_T, is the same as the draw for q_2 in the $T = 2$ case.

We now turn to the middle draws.

$$\mathbf{u_{t+1}\ /.\ u_{Rule}}$$

$$c\,(-q_t + q_{1+t}) + \Delta p_{1+t}$$

Since the u_t are assumed to be independent, the joint density $f\,(u_t, u_{t+1}) \propto$

```
PDF[NormalDistribution[0, σu], ut]
 PDF[NormalDistribution[0, σu], ut+1]
```

$$\frac{e^{-\frac{u_t^2}{2\sigma_u^2}-\frac{u_{1+t}^2}{2\sigma_u^2}}}{2\pi\sigma_u^2}$$

```
f = % /. uRule // Simplify
```

$$\frac{e^{-\frac{(-c\,q_{-1+t}+c\,q_t+\Delta p_t)^2+(-c\,q_t+c\,q_{1+t}+\Delta p_{1+t})^2}{2\sigma_u^2}}}{2\pi\sigma_u^2}$$

The odds ratio is:

$$\text{Odds} = \frac{\Pr[q_t = +1 \mid \ldots]}{\Pr[q_t = -1 \mid \ldots]} \tag{18}$$

```
Odds = (f /. qt → +1) / (f /. qt → -1) // Simplify
```

$$e^{\frac{2c\,(c\,q_{-1+t}+c\,q_{1+t}-\Delta p_t+\Delta p_{1+t})}{\sigma_u^2}}$$

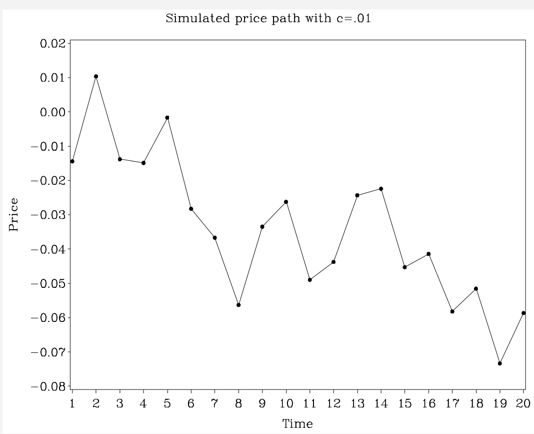
from which we may compute

$$\Pr[q_t = +1 \mid \ldots] = \frac{\text{Odds}}{1+\text{Odds}} \tag{19}$$

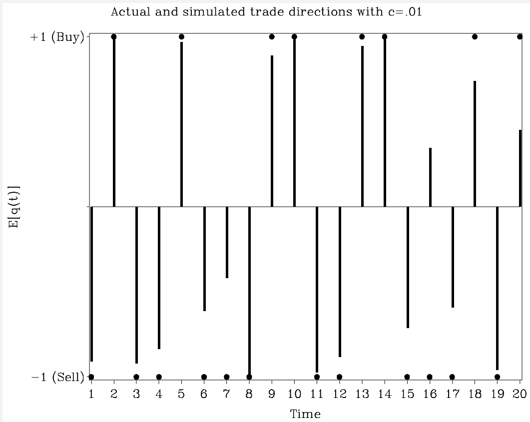
and make the desired draw.

□ Examples (SAS program *RollGibbs Analyze q.sas*)

I simulated twenty trades for a price process with $\sigma_u = 0.01$ and $c = 0.01$, and then ran the Gibbs sampler for 2,000 sweeps to estimate the trade directions. Here is a plot of the transaction prices:

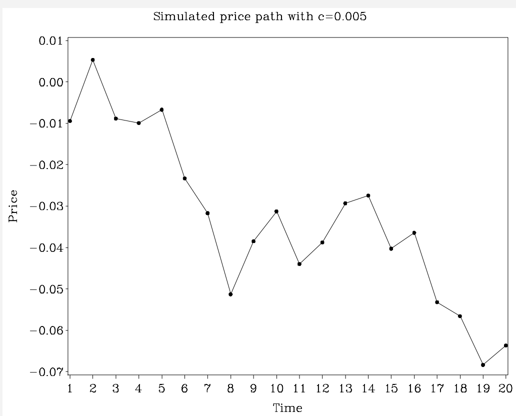


Below are the actual and average simulated trade directions. Actuals are indicated by a dot; estimated are indicated by the bars.

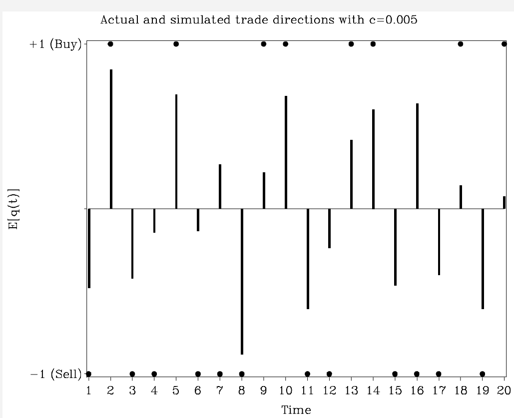


Note that estimates by and large agree with with actuals, at least in direction (q_{16} is the sole exception).

Now, consider the same analysis, with the cost parameter changed to $c = 0.005$, i.e., one-half the previous value. The figure below shows the prices. Notice that the bid-ask bounce is much less visually evident.

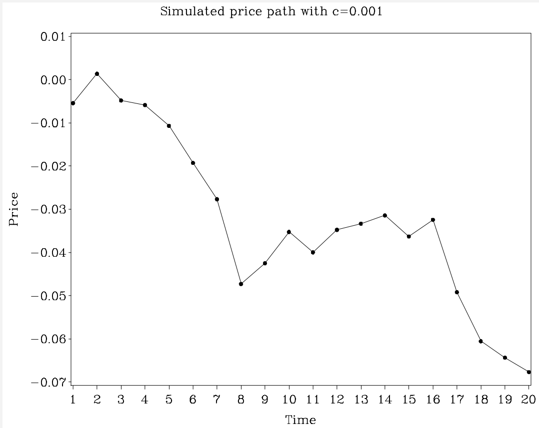


Here are the actual and estimated trade direction indicators:

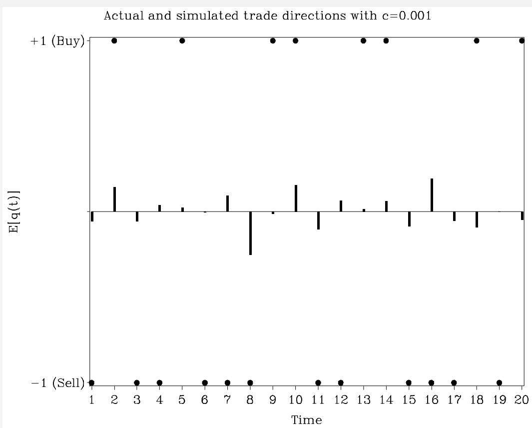


The buy/sell classification accuracy here is weaker. There are more directionally incorrect inferences, and the directions that are correctly identified are weaker. Just as we'd have a harder time picking out the buys and sells visually, the sampler has a tougher time classifying trades.

We'll now try things with $\sigma_u = 0.01$ and $c = 0.001$. Here's the price path:



And here are the actual and estimated trade directions.



The intuition is as follows. Intuitively, the Gibbs sampler tries to figure out how much of an oberserved price change is transient (due to bid ask effects) and how much is permanent (the efficient price innovation). When c is large relative to σ_u, bid-ask bounce generates reversals that are easy to pick out visually, and using the sampler. When c is small, though, bid-ask effects are swamped by the innovations in the efficient price, and are not easily discerned.

We'll see that this extends to the parameter estimates as well.

□ Modification when some of the q's are known.

In some samples, it might happen that the trade directions are known for some subset of the q_t. For these q_t, we don't simulate; we simply leave them at their known values.

This might seem to violate the assumed probability structure of the model in a fundamental way. After all, if the data generating process and our priors are that $q_t = \pm 1$, with equal probability, how can a definite realization be accomodated? The answer is that we're conditioning on the observed data, and the only thing that matters is the prior distribution of the q_t that we don't observe.

By way of a more formal justification, we could assume that the data generating process involves two steps:

- First q_t is drawn, ± 1, each with probability $\frac{1}{2}$.
- Next, an indicator variable O_t is drawn. With probability π, $O_t = 1$, and the actual q_t is observed. With probability $1 - \pi$, $O_t = 0$, and the actual q_t is unobserved.

As part of the sample, we "observe" the realizations of O_t. That is, we know which q_t are known for sure. If we don't care about modeling the O_t process, letting the observed q_t remain at their known values and simulating the rest corresponds to estimation conditional on the realized O_t. This is a sensible way to proceed.

In doing this, we are implicitly assuming that the O_t process is independent of q_t and u_t. If buys are more likely to be observed than sells, or if the realization of O_t depends on the magnitude of u_t ("Trades are more likely to be observed for large efficient price movements"), then the O_t process is informative, and we may wish to model it explicitly.

Modification when some of the q's are zero.

The U.S. daily CRSP dataset reports closing prices. But if there is no trade on day t, then the reported price is the midpoint of the closing bid and ask. This event is indicated in the data (using a negative price). Essentially, $p_t = m_t$.

Formally, we can incorporate this into the data generating process by noting that $p_t = m_t + c\, q_t = m_t$ when $q_t = 0$.

We can handle this situation in a fashion similar to the known-q_t case discussed above. If the price reported at time t is a quote midpoint, we set $q_t = 0$, and don't resample it during the sweeps.

Formally, this can be justified by letting O_t denote an indicator variable of whether or not there was a trade. Estimation can then proceed condtional on the O_t realizations. Here as well, we're implicitly assuming idependence. We're ruling out (or at least not modeling), for example, the possibility that trades are more likely to occur on days when there are large efficient price changes.

8. Full estimation of the Roll model

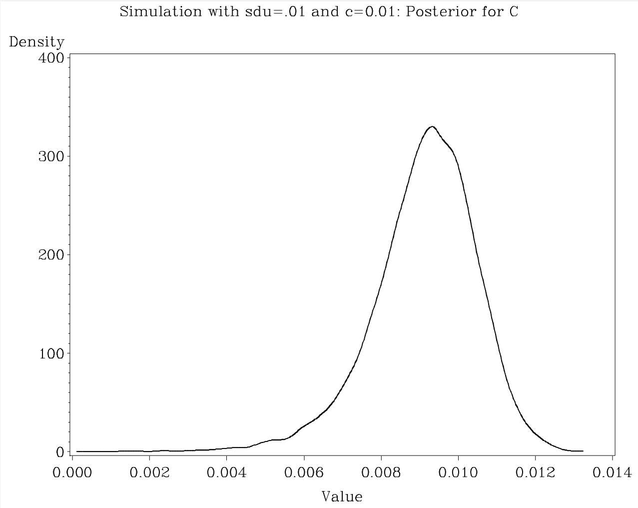
Sample runs from *Roll Gibbs 01.sas*

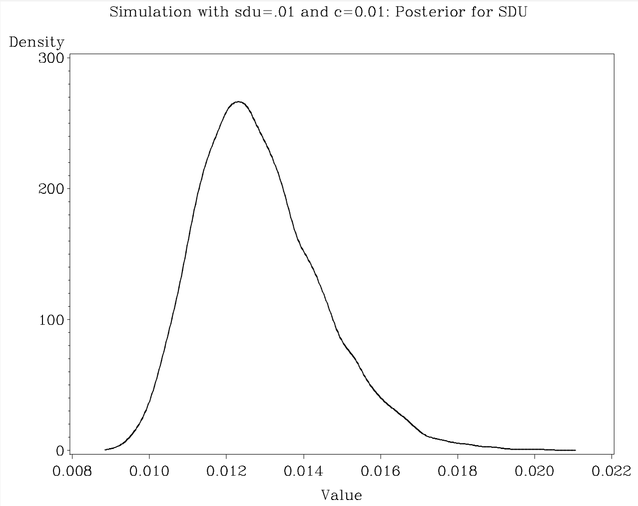
In all cases, the prior on c is $N(\mu_c^{Prior}, \Omega_c^{Prior})$, restricted to $c > 0$, with $\mu_c^{Prior} = 0$ and $\Omega_c^{Prior} = 1$. The prior on σ_u^2 is inverted gamma with $a = b = 1 \times 10^{-6}$.

$\sigma_u = .01$; $c = 0.01$, 100 observations, 20,000 sweeps (first 20% dropped)

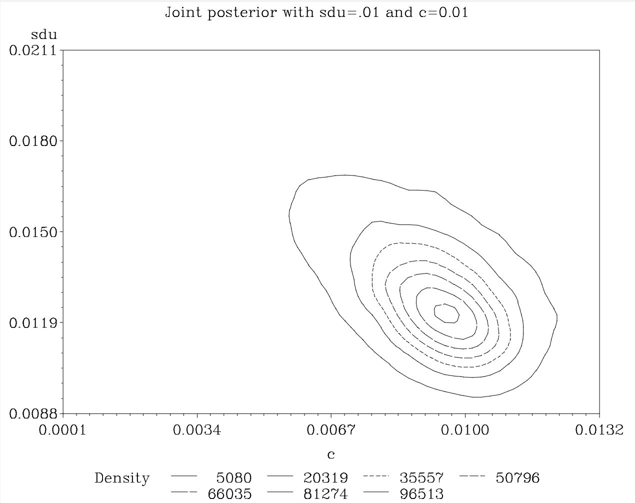
Posteriors:

Variable	N	Mean	Dev	Min	Max
SWEEP	16000	12001	4619	4001	20000
SDU	16000	0.0129	0.0016	0.0088	0.0211
C	16000	0.0091	0.0014	0.0001	0.0132





Contour plot of joint posterior:



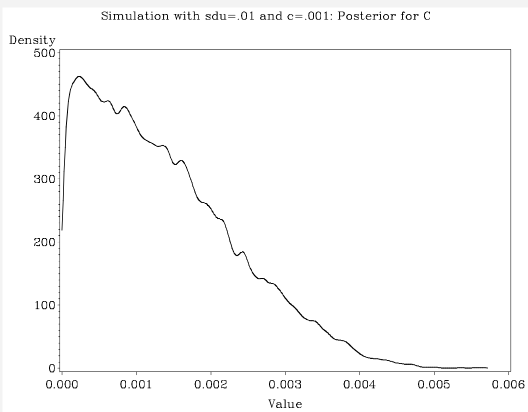
Note the downward slope of posterior. The procedure is trying to allocate volatility, either to the permanent (random-walk) component or to the transient component (effective cost). If more volatility is attributed to the `sdu` component, less is attributed to `c`.

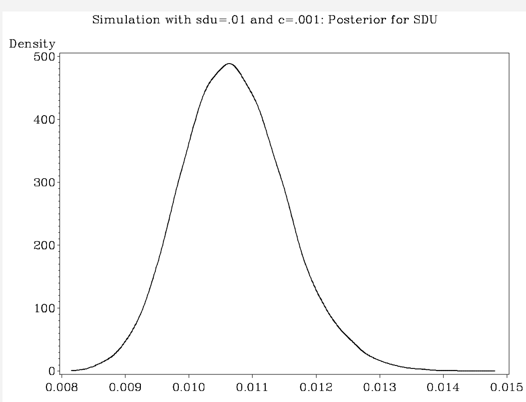
σ_u = .01; c = 0.001, 100 observations, 20,000 sweeps (first 20% dropped)

When `c` << σ_u, the transient cost effects (reversals) are difficult to disentangle from the random-walk component. We still get a reasonably sharp posterior for σ_u (= `sdu`), but the posterior for `c` is broad.

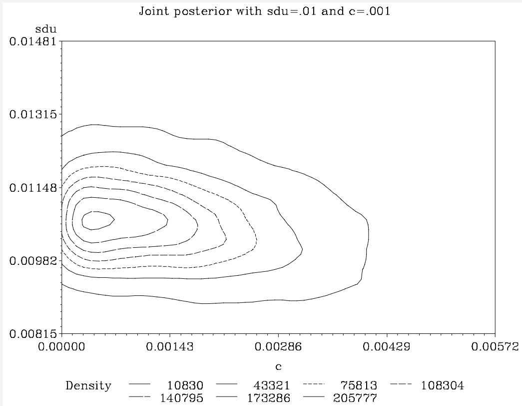
Posteriors:

Variable	N	Mean	Dev	Min	Max
SWEEP	16000	12001	4619	4001	20000
SDU	16000	0.0107	0.0008	0.0082	0.0148
C	16000	0.0014	0.0010	8E-8	0.0057





Contour plot of joint posterior:



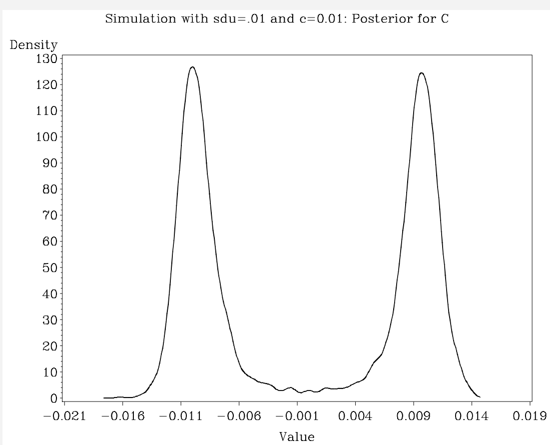
Use of an unrestricted prior for c

Is it really necessary to require $c > 0$? After all, if c really is non-negative, shouldn't the procedure pick it up?

When we run the last problem with no restrictions on c, two things happen.

- For all of the q_t, the sample draws average out to zero. (We can't tell whether a given trade is a buy or a sell.)
- The posterior for c is bimodal, and symmetric about zero.

Here's an example. 30 observations were simulated with $\sigma_u = 0.01$ and $c = 0.01$. The prior for c was not restricted to be non-negative. Here is the posterior:



What's happening is this. The Roll model with $c > 0$ is observationally equivalent to one in which $c < 0$, and the trade signs are reversed. Without the nonnegativity restriction on c, the posterior (and the sampler) will span both possibilities.

Using vague prior for c

In the preceding simulations, the prior for c is $N(0, 1)$, restricted to $c > 0$. This is fairly flat over the usual region of interest. (For a US equity, even extreme values of c, estimated from trade and quote data, are rarely above 0.05.) Nevertheless, there is some curvature. Why not remove all doubt and set the prior to, say, $N(0, 1000000)$?

The problem with this is that under some circumstances we may need to make a draw from the prior. This is not common, but in a small sample, with $c << \sigma_u$, over many draws, the following situation may arise. Suppose that on a particular sweep, the trade direction indicators are drawn to have the same sign: $q_1 = ... = q_T = +1$ or $q_1 = ... = q_T = -1$. In this case, all of the Δq_t are zero. But the Δq_t are the r.h.s. variables in the price change regression. If an explanatory variable in a regression has no variation, the regression is completely uninformative. In this case, the draw of c for that sweep must be made from the prior.

■ 9. Return factors

Return factors are logically introduced by adding them to the efficient price change specification, e.g.,

$$m_t = m_{t-1} + f_t' b + u_t \tag{20}$$

where f_t is a $K \times 1$ vector of known factor realizations and b is the vector of loadings (factor coefficients). The factor terms then appear in the trade price specification:

$$\Delta p_t = c\, \Delta q_t + f_t' b + u_t \tag{21}$$

For example, a market model for stock i might be specified as:

$$\Delta p_{it} = c_i\, \Delta q_{it} + b_i\, r_m + u_t \tag{22}$$

where r_m is the return on a market index.

The b are estimated as another coefficient in the Bayesian regression (the same regression in which c is estimated). The parameters of the coefficient prior μ_β^{Prior} and $\Omega_\beta^{\text{Prior}}$ are expanded to include the new coefficients, and they are drawn (simulated) in the same step as the draw for c.

The addition of a return factor will generally increase the resolution between permanent (random-walk) and transient (bid-ask) components. In practice, this accomplishes two things:

- We'll get a better estimate of c.
- We'll also generally get better estimates of b.

The first of these is pretty straightforward: the explanatory power of the factors reduces the residual variance. The second point may require some expansion. bs are conventionally estimated using daily price changes. These daily price changes are contaminated by bid-ask bounce. For some stocks, bid-ask bounce may be large compared to the factor-induced and idiosyncratic changes in the efficient price, leading to large estimation errors in the bs. Estimating a specification that includes a $c\, \Delta q_t$ term effectively allows us to estimate bs on price series that are purged of the bid-ask bounce.

Market model example (RollGibbsBeta 01.sas)

The specification is the one-factor model described above The parameters are:
$c = 0.01$; $r_m \sim N(0, \sigma_m^2)$ with $\sigma_m = 0.01$; $\sigma_u = 0.01$; $\beta = 1.1$.
100 observations were simulated; 10,000 sweeps (first 20% discarded).
The posterior summary statistics are:

Variable	N	Mean	Dev	Min	Max
SWEEP	8000	6001	2310	2001	10000
SDU	8000	0.0106	0.0012	0.0076	0.0174
C	8000	0.0109	0.0009	0.0070	0.0141

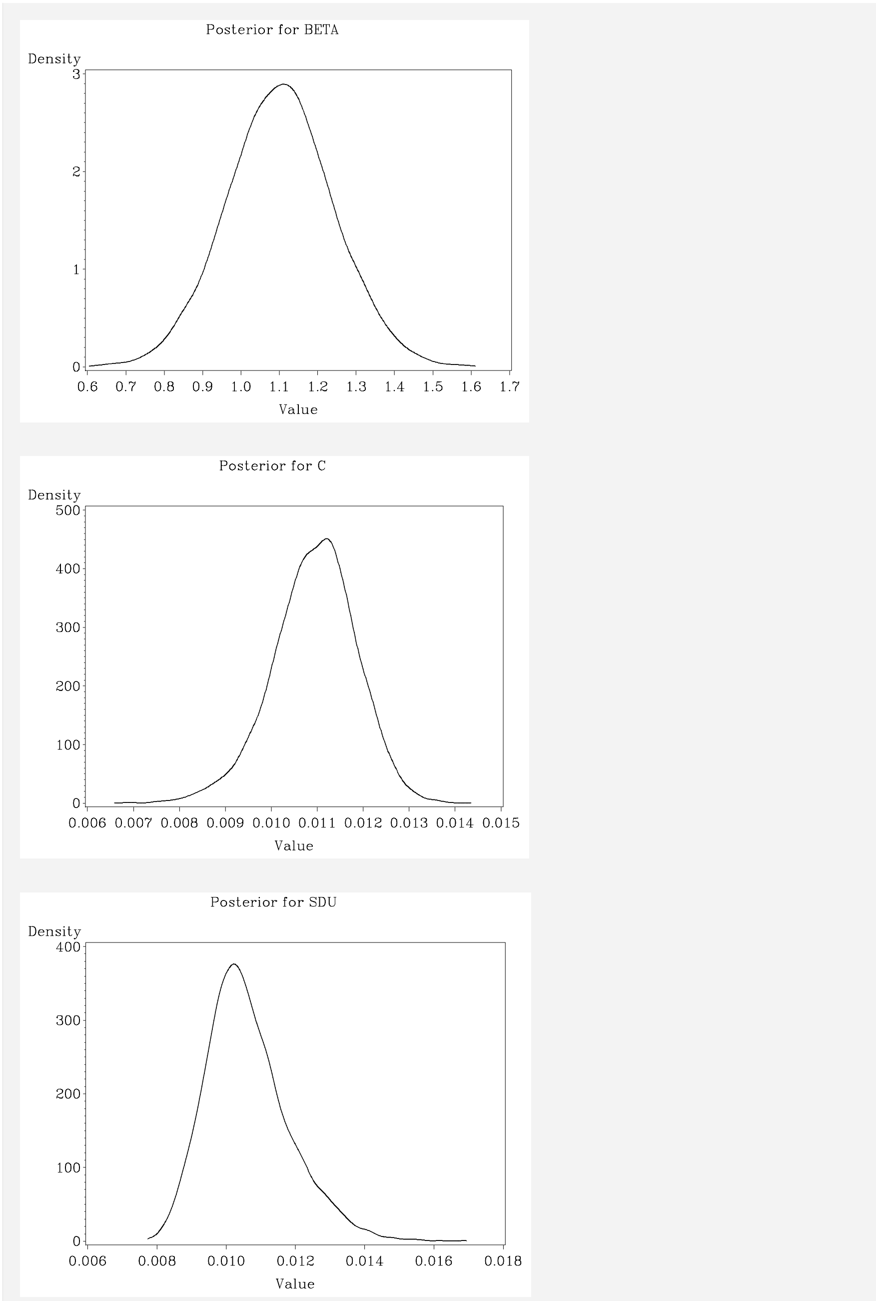
BETA	8000	1.100	0.139	0.603	1.607

For purposes of comparison, β was also estimated in the usual way (OLS). The estimated specification:

$$r = \underset{(0.186)}{1.132}\, r_m + e. \tag{23}$$

The standard error of the β estimate is 0.186 - modestly higher than the 0.138 value for the Gibbs estimate.

Here are the posteriors:



■ Extensions

Hasbrouck (2006) discusses variation in c. Hasbrouck (2004) discusses variations with asymmetric information.

■ Transformations of the Gamma distribution

When $a = n/2$ and $b = 1/2$, the Gamma becomes:

```
PDF[ChiSquareDistribution[n], x]
```

$$\frac{2^{-n/2}\, e^{-x/2}\, x^{-1+\frac{n}{2}}}{\text{Gamma}[\frac{n}{2}]}$$

We sometimes need to determine the pdf's of σ_u^2 and/or σ_u.

If $\frac{1}{\sigma_u^2}$ is Gamma, then what is the pdf of σ_u^2?

Recall that if $y = g(x)$, then $f(y) = f(g^{-1}(y))\,|\,g^{-1\prime}(y)\,|$

```
gRule = {g[x_] :→ x^-1, gi[y_] :→ y^-1};
```

```
fy = PDF[GammaDistribution[a, λ], x] * D[-gi[y] /. gRule, y] /.
     λ → 1 / b /. x → gi[y] /. gRule // Simplify
```

$$\frac{(\frac{1}{b})^{-a}\, e^{-\frac{b}{y}}\, (\frac{1}{y})^{1+a}}{\text{Gamma}[a]}$$

Verify that this integrates to unity:

```
Integrate[fy, {y, 0, ∞},
 Assumptions -> {a ∈ Reals, b ∈ Reals, a > 0, b > 0}]
```

1

Compute the expectation:

```
Integrate[y fy, {y, 0, ∞},
 Assumptions -> {a ∈ Reals, b ∈ Reals, a > 1, b > 0}]
```

$$\frac{b}{-1+a}$$

If $\frac{1}{\sigma_u^2}$ is Gamma, then what is the pdf of σ_u (the standard deviation)?

```
gRuleSD = {g[x_] :→ x^(-1/2), gi[y_] :→ y^-2};
```

```
fy = PDF[GammaDistribution[a, λ], x] * D[-gi[y] /. gRuleSD, y] /.
     λ → 1 / b /. x → gi[y] /. gRuleSD // Simplify
```

$$\frac{2\,(\frac{1}{b})^{-a}\, e^{-\frac{b}{y^2}}\, (\frac{1}{y^2})^{a}}{y\,\text{Gamma}[a]}$$

Verify that this integrates to unity:

```
Integrate[fy, {y, 0, ∞},
 Assumptions -> {a ∈ Reals, b ∈ Reals, a > 0, b > 0}]
```

1

Compute the expectation:

```
Integrate[y fy, {y, 0, ∞},
 Assumptions -> {a ∈ Reals, b ∈ Reals, a > 1, b > 0}]
```

$$\frac{\sqrt{b}\ \mathrm{Gamma}\left[-\frac{1}{2}+a\right]}{\mathrm{Gamma}[a]}$$

■ References

Ball, Clifford A. and Tarun Chordia, 2001, True spreads and equilibrium prices. Journal of Finance 56, 1801-35.

Carlin, Bradley P. and Thomas A. Louis, 2000, Bayes and Empirical Bayes Methods for Data Analysis. London, Chapman and Hall.

Geweke, John, 2005, Contemporary Bayesian Statistics and Econometrics. New York, John Wiley and Sons.

Hasbrouck, Joel, 1999, Security bid/ask dynamics with discreteness and clustering: Simple strategies for modeling and estimation. Journal of Financial Markets 2, 1-28; doi:10.1016/S1386-4181(98)00008-1

Hasbrouck, Joel, 2004, Liquidity in the futures pits: Inferring market dynamics from incomplete data. Journal of Financial and Quantitative Analysis 39.

Hasbrouck, Joel, 2006, Trading costs and returns for US equities: the evidence from daily data. New York University, Stern School of Business.

Kim, Chang-Jin and Charles R. Nelson, 2000, State-space models with regime switching. Cambridge, Massachusetts, MIT Press.

Lancaster, Tony, 2004, An Introduction to Modern Bayesian Econometrics. Malden (MA), Blackwell Publishing.

Shephard, Neil, 2005, Stochastic Volatility. Oxford, Oxford University Press.

Tanner, Martin A., 1996, Tools for Statistical Inference. New York, Springer-Verlag.