BAYESIAN ESTIMATION OF PROBABILITIES OF TARGET RATES USING FED FUNDS FUTURES OPTIONS

MARK FISHER

Preliminary and Incomplete

Abstract. This paper adopts a Bayesian approach to the problem of extracting the probabilities of the FOMC's target Fed Funds rate from option prices, extending the framework provided by Carlson et al. (2005) in a number of directions. Likelihood-related novelties include (*i*) the allowance for "slippage" between the target rate and the month-average rate and (*ii*) accounting for the nonnegativity of options prices by truncating the measurement error. In addition, we generalize the likelihood for target rates to include the likelihood for target-rate *paths*. Prior-related novelties include (*i*) enforcing nonnegativity constraints on the probabilities and (*ii*) the "informed ignorance" prior that allows for the inclusion of a large number of possible target rates (which is especially important for joint/path estimation).

1. Introduction

As Carlson et al. (2005) point out, the problem of extracting the probabilities of the FOMC's target Fed Funds rate from option prices is greatly simplified if one assumes the target rate can take only a small number of values. They use Classical least-squares regression techniques to extract the target-rate probabilities. The paper in hand adopts a Bayesian approach to inference instead, and extends the framework provided by Carlson et al. (2005) in a number of directions.[1]

The main advantage to the Bayesian approach is that one gets to specify a prior distribution. The prior allows us to deal with important features of the target-rate probabilities in a sensible way. First, the target-rate probabilities should satisfy an equality constraint (they should sum to one) and a number of inequality constraints (they should each be nonnegative). Although both Classical and Bayesian approaches handle the equality constraint easily, the Classical approach has difficulty handling the inequality constraints while the Bayesian approach handles them quite naturally via the prior.[2]

Second, the Bayesian approach makes it easy to allow for a large number of target rates (or target-rate paths) in the model. There is a simple prior for the unobserved probabilities that encapsulates what might be called "partially informed ignorance." In effect, this

---

Date: March 21, 2012 @ 11:09am.

I thank Jerry Dwyer, Mark Jensen, and Dan Wagonner for helpful discussions. The opinions expressed herein are those of the author and do not represent those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

[1]The reader is referred to Carlson et al. (2005) for omitted institutional details.

[2]For an early example of the Bayesian approach to handling inequality constraints, see Geweke (1986).

prior asserts that only a few of the target-rate probabilities are nonnegligible—without specifying which those are. Formally, this prior is a symmetric Dirichlet distribution with a "concentration" parameter that encourages parsimony. In addition, the concentration parameter itself is assumed to be unknown and therefore has its own prior distribution. (A closely-related prior is used in mixture models to allow for a potentially large number of mixture components while at the same time encouraging parsimony. One should keep in mind, however, the current exercise is one of functional-form fitting, rather than density estimation *per se*.) This prior allows one to include a large number of target rates (greater even than the number of observations) and still obtain a stable and well-estimated posterior distribution. This feature has the potential to allow robust estimation of paths of interest rate targets, where the number of paths (and the number of associated joint probabilities) is quite large.

Third, the Bayesian approach provides a coherent treatment of the possibility of unscheduled FOMC meetings.

The main disadvantage to the Bayesian approach is that the computational burden is significantly higher than for the Classical approach. (Regarding this point, however, one should recognize there are no free lunches, even in statistics.)

Novelties. Likelihood-related novelties include (*i*) the allowance for "slippage" between the target rate and the month-average rate and (*ii*) accounting for the nonnegativity of options prices by truncating the measurement error. In addition, we generalize the likelihood for target rates to include the likelihood for target-rate *paths*. Prior-related novelties include (*i*) enforcing nonnegativity constraints on the probabilities and (*ii*) the "informed ignorance" prior that allows for the inclusion of a large number of possible target rates (which is especially important for joint estimation).

Caveats. The options are American, not European. The risk-neutral measures for different horizons (different months) are not mutually compatible and consequently the proposed joint estimation is not completely coherent. Similarly, futures prices are used in place of forward prices. Liquidity is not taken into account in weighing the quotes. (This could enter via the uncertainty of the measurement error.)

Outline of paper. Section 2 presents on overview of the model, including the likelihood, the prior, and the posterior sampler. Section 3 introduces month-average target rates, deals with the joint distribution of multiple target rates (target rate paths), and outlines of how to extend the model to include unscheduled meetings. Section 4 describes some issues related to putting the available data into a form suitable for estimation. Section 5 discusses stochastic-process dynamics of the target-rate probabilities. Appendix A includes some additional material.

The **empirical section** of the paper is currently **absent**. Although a few "proof of concept" tests have been run, a systematic treatment of a substantial data set is only just beginning.

2. The big picture

The goal is to estimate the risk-neutral probability distribution for the target fed funds rate from put and call options data.

Let R denote the target rate as set by the Federal Open Market Committee (FOMC).[3] We assume the target rate is restricted to a finite set of known values $R \in r = \{r_1, \ldots, r_K\}$. Let the risk-neutral probabilities for the values be given by

$$p(R = r_j|\beta) = \beta_j, \tag{2.1}$$

where $\beta = (\beta_1, \ldots, \beta_K)$, $\beta_j \geq 0$ and $\sum_{j=1}^K \beta_j = 1$. We can express the (risk-neutral) density for R as a mixture of point masses:

$$p(R|\beta) := \sum_{j=1}^{K} \beta_j \, \delta(R - r_j), \tag{2.2}$$

where $\delta(\,\cdot\,)$ is the Dirac delta function.[4] We assume β is unknown. It is the focus of our investigation.

We model the month-average fed funds rate S as the sum of the target rate R and the "slippage" between S and R:

$$S = R + u. \tag{2.3}$$

Given (2.2) and (2.3), we can express the risk-neutral density for S as a related mixture of point masses:

$$p(S|\beta, u) := \sum_{j=1}^{K} \beta_j \, \delta(S - r_j - u). \tag{2.4}$$

We assume u is unknown as well.

Consider a European option written on the month-average rate S with a strike price of $\mathcal{K}$. The payoff to the option at expiration is

$$\varphi(\mathcal{K}, \gamma, S) := \big(\gamma\,(S - \mathcal{K})\big)^+, \tag{2.5}$$

where $x^+ := \max(x, 0)$ and

$$\gamma = \begin{cases} 1 & \text{for a call} \\ -1 & \text{for a put} \end{cases}. \tag{2.6}$$

Let B denote the value of a risk-free zero-coupon bond that matures on the expiration date. The value V of the option is the present value of the expected payout computed using the

[3] See Section 3 for institutional details.

[4] The main features of $\delta(\,\cdot\,)$ are (i) $\delta(x - x_0) = 0$ if $x \neq x_0$, (ii) $\int_{-\infty}^{\infty} \delta(x - x_0)\, dx = 1$, and (iii) $\int_{-\infty}^{\infty} f(x)\, \delta(x - x_0)\, dx = f(x_0)$.

risk-neutral probabilities:

$$\begin{aligned} V &= B \int_{-\infty}^{\infty} \varphi(\mathcal{K}, \gamma, S)\, p(S|\beta, u)\, dS \\ &= B \sum_{j=1}^{K} \beta_j \int_{-\infty}^{\infty} \varphi(\mathcal{K}, \gamma, S)\, \delta(S - r_j - u)\, dS \\ &= B \sum_{j=1}^{K} \beta_j\, \varphi(\mathcal{K}, \gamma, r_j + u). \end{aligned} \tag{2.7}$$

Equation (2.7) suggests the task ahead: Given some options data, fit a functional form using the basis functions $\varphi(\,\cdot\,, \cdot\,, \cdot\,)$.

Assume the data are composed of n observations, $\{\mathcal{K}_i, \gamma_i, V_i\}_{i=1}^{n}$, and suppose they are generated according to

$$y_i = \sum_{j=1}^{K} \beta_j\, \varphi(\mathcal{K}_i, \gamma_i, r_j + u) + \varepsilon_i \tag{2.8}$$

where $y_i := B^{-1}\, V_i$ is the deflated option value and ε_j is a measurement error. Let X denote the $n \times K$ matrix where

$$X_{ij} = \varphi(\mathcal{K}_i, \gamma_i, r_j + u) = \big(\gamma_i\, (r_j + u - \mathcal{K}_i)\big)^{+}. \tag{2.9}$$

(Note that X depends on the unknown u. We will write $X(u)$ when it is convenient to make this dependence explicit.) We can express (2.8) as

$$y_i = X_i \beta + \varepsilon_i, \tag{2.10}$$

where X_i denotes the the i-th row of X. Stacking the n observations, we have

$$y = X\beta + \varepsilon, \tag{2.11}$$

where $y = (y_1, \ldots, y_n)$ and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$. It is convenient to suppose $\varepsilon \sim \mathrm{N}(0_n, \sigma^2 I_n)$.[5] We assume σ^2 is unknown.

Let $\theta = (\beta, u, \sigma^2)$ denote the vector of unknown parameters. The likelihood for θ follows from (2.11) and the distribution for the measurement error:

$$p(y|\theta) = \mathrm{N}(y|X(u)\beta, \sigma^2 I_n) = (2\,\pi\,\sigma^2)^{-n/2}\, \exp\!\left(\frac{-S(\beta, u)}{2\,\sigma^2}\right), \tag{2.12}$$

where[6]

$$S(\beta, u) := (y - X(u)\,\beta)^{\top} (y - X(u)\,\beta). \tag{2.13}$$

Remark. The novelty in (2.12) relative to Carlson et al. (2005) is to allow for $u \neq 0$.

[5] One drawback of this assumption is that the sampling distribution for the data $p(y|\theta)$ [see (2.12)] implies we may observe negative option prices. This can be remedied by truncating the distribution for y. We incorporate this below.

[6] $A^{\top}$ denotes the transpose of A.

Predictive distributions. At this point it is convenient to make a few remarks about predictive distributions, both prior and posterior, for R and S.

Let $p(\beta)$ and $p(\beta|y)$ denote the marginal prior and posterior distributions for β and let $E[\beta_j] = \int \beta_j \, p(\beta) \, d\beta$ and $E[\beta_j|y] = \int \beta_j \, p(\beta|y) \, d\beta$ denote the prior and posterior expectations of β_j.

The prior and posterior predictive distributions for R are both mixtures of point masses where the mixture weights are the applicable expectations of β:

$$p(R) = \int p(R|\beta) \, p(\beta) \, d\beta = \sum_{j=1}^{K} E[\beta_j] \, \delta(R - s_j) \tag{2.14}$$

$$p(R|y) = \int p(R|\beta) \, p(\beta|y) \, d\beta = \sum_{j=1}^{K} E[\beta_j|y] \, \delta(R - s_j). \tag{2.15}$$

Note that $E[\beta|y]$ is the only feature of the posterior distribution that matters if the goal of the inferential exercise is to compute the posterior predictive distribution for R.[7] Also note that as we approach the end of the month, the posterior distribution for β should approach a point mass at $\beta = I_j$ for some j, where I_j is the j-th row the the identity matrix, so that $E[\beta|y] \to I_j$.

It is instructive to examine the predictive distributions for S as well. The prior predictive distribution for S inherits the form of the prior distribution for the slippage parameter u. We assume prior independence between β and u: $p(\beta, u) = p(\beta) \, p(u)$. The prior predictive distribution for S has the form of a mixture:

$$\begin{aligned} p(S) &= \iint p(S|\beta, u) \, p(\beta) \, p(u) \, du \, d\beta \\ &= \sum_{j=1}^{K} \int \beta_j \, p(\beta) \left(\int \delta(S - s_j - u) \, p(u) \, dv \right) d\beta \\ &= \sum_{j=1}^{K} E[\beta_j] \, p(S|j), \end{aligned} \tag{2.16}$$

where

$$p(S|j) := p(u)|_{u=S-s_j}. \tag{2.17}$$

Therefore, if $p(u)$ is continuous, then $p(S)$ will be a mixture of continuous distributions with different locations.

[7]In a decision-making setting where one can either act now wait for new information before acting, the posterior uncertainty regarding β may play a role.

The posterior predictive distribution for S is more complicated owing to potential dependence between β and u in the posterior distribution:

$$\begin{aligned} p(S|y) &= \iint p(S|\beta,u)\,p(\beta,u|y)\,du\,d\beta \\ &= \sum_{j=1}^{K}\int \beta_j\,p(\beta|y)\left(\int \delta(S-s_j-u)\,p(u|\beta,y)\,du\right)d\beta \\ &= \sum_{j=1}^{K}\int \beta_j\,p(\beta|y)\,p(S|j,\beta,y)\,d\beta \end{aligned} \tag{2.18}$$

where

$$p(S|j,\beta,y) = p(u|\beta,y)|_{u=S-s_j}. \tag{2.19}$$

If $p(S|j,\beta,y) \approx p(S|j,y)$, then

$$p(S|y) \approx \sum_{j=1}^{k} E[\beta_j|y]\,p(S|j,y). \tag{2.20}$$

As we approach the end of the month, the posterior predictive distribution should collapse to a point mass on the actual month-average funds rate.

The prior. It remains to adopt a prior distribution for the unknown parameters $\theta = (\beta, u, \sigma^2)$. We assume prior independence: $p(\theta) = p(\beta)\,p(u)\,p(\sigma^2)$. For σ^2 we adopt the Jeffreys prior: $p(\sigma^2) \propto 1/\sigma^2$.

Here we provide some introductory remarks regarding the prior for u. Typically we will adopt a symmetric prior for u, centered on zero. They may be times, however, when we will center it elsewhere (the beginning of the financial crisis, for example, when the month-average rate dropped significantly with no change in the stated target rate). The form of the prior for u should allow the scale to be chosen endogenously. One could for example adopts a Gaussian prior for u, perhaps a mixture to allow for differing regimes.

Now we focus on the prior for β. A benefit of the Bayesian approach is the ability to incorporate important considerations via the prior for β. In particular, the prior for β should embody two features. First the prior should ensure the constraint $\beta \in \Delta^{K-1}$, where Δ^{K-1} is the $(K-1)$-dimensional simplex. Second the prior should be capable of expressing the idea that while many target rates are possible, only a few should have nontrivial probability. The Dirichlet distribution embodies both of these features. Let

$$p(\beta|\alpha,\bar{\xi}) = \frac{\Gamma(\alpha)}{\prod_{j=1}^{K}\Gamma(\alpha\,\bar{\xi}_j)}\prod_{j=1}^{K}\beta_j^{\alpha\,\bar{\xi}_j-1}, \tag{2.21}$$

where $\alpha > 0$ and $\bar{\xi} \in \Delta^{K-1}$. Note $E[\beta|\alpha,\bar{\xi}] = \bar{\xi}$. We will refer to α as the concentration parameter. If $\bar{\xi}_j = 1/K$ and $\alpha = K$, then the prior is flat: $p(\beta|\alpha,\bar{\xi}) = (K-1)!$.

As K get large relative to n, the flat prior for β tends to dominate the likelihood (2.12) and the posterior can become quite flat itself. Setting $\alpha < K$ encourages parsimony, pushing the mass of the prior toward the vertices of the simplex, thereby implicitly suggesting that only a few of the components are nonnegligible. (This prior could be described as "partially informed ignorance.")

On the other hand, if $\bar{\xi}$ is well-informed (because, for example, it is based on closely-related data) then $\alpha > K$ may be suitable. Below we discuss incorporating uncertainty about α via a prior.

The posterior sampler. The main cost of the Bayesian approach is drawing from the posterior distribution when an analytical solution is absent (as it is here). The posterior distribution for the parameters can be expressed as

$$p(\theta|y) \propto p(y|\theta)\, p(\beta|\alpha, \bar{\xi})\, p(u)/\sigma^2, \tag{2.22}$$

where $\theta = (\beta, u, \sigma^2)$. We take a Gibbs sampler approach, applying Metropolis–Hastings where necessary within.

Drawing from the posterior for σ^2 (conditional on β and u) is straightforward. Note $\sigma^2|y, \beta, u \sim \text{Inv-}\chi^2(\nu, s^2)$, where $\nu = n$ and $s^2 = S(\beta, u)/n$.[8]

The posterior distribution for the slippage parameter u conditional on (β, σ^2) can be expressed as

$$p(u|y, \beta, \sigma^2) \propto \exp\left(\frac{-S(\beta, u)}{2\,\sigma^2}\right) p(u). \tag{2.23}$$

We can make draws from $p(u|y, \beta, \sigma^2)$ using a Metropolis scheme.

We now turn to drawing from the posterior for β (conditional on σ^2 and u). First we show how to draw from the posterior assuming the prior for β is flat; then we show how to draw from the posterior with the more general prior.

A technical detail. Since $\sum_{j=1}^{K} \beta_j = 1$, the conditional distribution $\beta_j|\beta_{-j}$ is degenerate: the value of β_j is fixed by β_{-j}. By "removing" one of the components, say β_k, we can then cycle through the remaining $K-1$ components via a one-at-a-time Gibbs sampler. However, the (average) magnitude of β_k will affect the efficiency of the sampler. Let $\beta_j|\beta^k_{-j}$ denote conditioning on $\beta_{-j} \setminus \{\beta_k\}$. If β_k is close to zero, we will find ourselves back in the previous trap with little or no wiggle room for $\beta_j|\beta^k_{-j}$. Therefore, for each sweep of the Gibbs sampler we will remove the largest component from the previous sweep and sample over the remaining components.

Back to the main thread. If the prior for β were flat, then the posterior for β (conditional on σ^2) would be a multivariate normal distribution truncated to the simplex. This truncated normal distribution can be hard to draw from when only a small fraction of the mass of the unrestricted distribution is contained in the simplex. Moreover, with K large relative to n, $X^\top X$ will be singular and hence not invertible. Nevertheless, a one-at-a-time Gibbs sampler works well.

The posterior distribution for $\beta_i|\beta^k_{-i}$ is a univariate truncated normal distribution:

$$p(\beta_i|y, \beta^k_{-i}, \sigma^2, u) = \text{N}_{[0,b^k_i]}(\beta_i|m_i, s_i^2), \tag{2.24}$$

where

$$b^k_i = \beta_i + \beta_k \tag{2.25}$$

[8]If the prior for σ^2 were different, we could use this as a proposal for a Metropolis step.

and (m_i, s_i^2) can be readily calculated. In particular, note $S(\beta, u) = c_{i0} + c_{i1}\,\beta_i + c_{i2}\,\beta_i^2$ for some coefficients (c_{i0}, c_{i1}, c_{i2}) that are functions of $(\beta^k_{-i}, X(u), y)$. Consequently, $m_i = -c_{i2}/(2\,c_{i3})$ and $s_i^2 = \sigma^2/c_{i3}$. Draws from the univariate truncated normal can be made using the inverse CDF method (for example).

We can accommodate the general Dirichlet prior for β by incorporating a Metropolis–Hastings step. We use the truncated normal distribution in (2.24) to draw the proposal β'_i. Because the proposal is a draw from the conditional posterior given a flat prior, the "Hastings ratio" equals the inverse of the likelihood ratio,

$$\frac{q(\beta, \beta')}{q(\beta', \beta)} = \frac{p(y|\beta, u, \sigma^2)}{p(y|\beta', u, \sigma^2)}. \tag{2.26}$$

Consequently, the acceptance condition depends only on the prior ratio:

$$\frac{p(\beta'|\alpha, \bar{\xi})}{p(\beta|\alpha, \bar{\xi})} = \left(\frac{\beta'_i}{\beta_i}\right)^{\alpha\,\bar{\xi}_i - 1} \left(\frac{\beta'_k}{\beta_k}\right)^{\alpha\,\bar{\xi}_k - 1} \geq u, \tag{2.27}$$

where $u \sim \mathrm{U}(0, 1)$ and where $\beta'_k = b_i^k - \beta'_i$. Note that if the prior for β were flat, then (of course) the proposal would always be accepted.

Uncertainty about α. The hyperparameter α plays an important role in the prior for β. Instead of specifying a fixed value for α, we can adopt a prior for α and incorporate an additional Metropolis step to sample α.

The likelihood for α is given by (2.21). Let the prior for α be given by[9]

$$p(\alpha|\zeta, \tau) = \frac{\alpha^{\frac{1}{\tau} - 1}\,\zeta^{\frac{1}{\tau}}}{\tau\left(\alpha^{\frac{1}{\tau}} + \zeta^{\frac{1}{\tau}}\right)^2}, \tag{2.28}$$

where $\zeta, \tau > 0$. The mean of the distribution does not exist. We can see how to interpret the parameters via the quantile function where $\alpha = Q(x)$ for $x \in (0, 1)$ and

$$Q(x) = \zeta \left(\frac{x}{1 - x}\right)^{\tau}. \tag{2.29}$$

Thus ζ is the median and τ determines the spread. For our purpose, it is natural to choose $\zeta = K$.[10]

For making draws from the posterior, it is convenient to change variables: let $z = \log(\alpha)$. Given (2.28), the prior for z is given by

$$p(z|\zeta, \tau) = \frac{\zeta^{1/\tau}\,e^{z/\tau}}{(\zeta^{1/\tau} + e^{z/\tau})^2\,\tau}, \tag{2.30}$$

[9] This is a special case of the Singh–Maddala distribution.

[10] The essence of this prior is captured by the following change of variables: Let $\eta := \log(\alpha/K)$ and let $\zeta = K$; then

$$p(\eta|\tau) = \frac{e^{\eta/\tau}}{(1 + e^{\eta/\tau})^2\,\tau},$$

which is symmetric around zero with a standard deviation of $(\pi/\sqrt{3})\,\tau$.

where $p(z|\zeta,\tau)$ is symmetric around its the mean of $\log(\zeta)$ with a standard deviation of $(\pi/\sqrt{3})\,\tau$. The Metropolis step proceeds as follows: Given some average step-size s, make random-walk proposals of $z' \sim \mathrm{N}(z, s^2)$ and accept z' if

$$\frac{p\big(\beta \mid \alpha', \overline{\xi}\big)\, p(z'|\zeta,\tau)}{p\big(\beta \mid \alpha, \overline{\xi}\big)\, p(z|\zeta,\tau)} \geq u, \tag{2.31}$$

where $u \sim \mathrm{U}(0,1)$ and where $\alpha = e^z$ and $\alpha' = e^{z'}$.

A typical setting for the hyperparameters is be $\overline{\xi} = (1/K, \ldots, 1/K)$, $\zeta = K$, and $\tau = 2$. However, in going from one observation day to the next, it may make sense to use $\overline{\xi}^t = E[\beta^{t-1}|y^{t-1}]$. (Limited experience indicates that the posterior distribution for α will be quite different for these two priors.)

The nonnegativity of options prices. A problem with the likelihood (2.12) is that it may imply $\Pr[y_i < 0|\beta, u, \sigma^2] \gg 0$. We believe one would never observe such out-of-bounds prices. (Such an observation would be discarded as a data mistake and not interpreted as the result of measurement error as we have construed it.) One way to fix this implication is to truncate the sampling distribution for the data at zero; this has the effect of imbuing the measurement error with an upward bias, where the bias gets larger as the option price gets closer to zero.

The likelihood (for a single observation) that does not take into account the positivity of option prices can be expressed as $p(y_i|\theta) = \mathrm{N}(y_i|\mu_i, \sigma^2)$ where $\mu_i := X_i(u)\beta$. By comparison, consider the likelihood given a truncated Gaussian distribution:

$$\widehat{p}(y_i|\theta) := 1_{[0,\infty)}(y_i)\, H(\mu_i, \sigma^2)\, \mathrm{N}(y_i|\mu_i, \sigma^2), \tag{2.32}$$

where[11]

$$H(\mu, \sigma^2) := \left\{1 - \Phi\left(\frac{-\mu}{\sigma}\right)\right\}^{-1}. \tag{2.33}$$

Note $\int_0^\infty \widehat{p}(y_i|\theta)\, dy_i = 1$ and

$$\widehat{E}[y_i|\theta] = \int_0^\infty y_i\, \widehat{p}(y_i|\theta)\, dy_i = \mu_i + \sigma^2\, H(\mu_i, \sigma^2)\, \mathrm{N}(0|\mu_i, \sigma^2) > \mu_i. \tag{2.34}$$

The inequality in (2.34) can be interpreted as an upward bias in the measurement error. This upward bias is largest at $\mu_i = 0$ and declines monotonically to zero as $\mu_i \to \infty$.

Let us now consider the likelihood of the entire data set:

$$\widehat{p}(y|\theta) := \prod_{i=1}^n \widehat{p}(y_i|\theta) = \mathbf{1}_{[0,\infty)^n}(y)\, C(\theta)\, p(y|\theta), \tag{2.35}$$

where $p(y|\theta) = \prod_{i=1}^n p(y_i|\theta)$ is given in (2.12) and

$$C(\theta) := \left(\int_{[0,\infty)^n} p(y|\theta)\, dy\right)^{-1} = \prod_{i=1}^n H\big(\mu_i, \sigma^2\big). \tag{2.36}$$

Note, for fixed σ, $H(\mu, \sigma^2)$ declines monotonically from $H(0, \sigma^2) = 2$ to $\lim_{\mu\to\infty} H(\mu, \sigma^2) = 1$. Therefore $1 \leq C(\theta) \leq 2^n$.

[11] $\Phi(\,\cdot\,)$ is the standard normal cumulative distribution function (CDF).

Let $M_{[0,\infty)^n}$ denote the model where $y \in [0,\infty)^n$. Assuming the data are all nonnegative, the likelihood of this model is

$$p(y|M_{[0,\infty)^n}) = \int C(\theta)\, p(y|\theta)\, p(\theta)\, d\theta. \tag{2.37}$$

Thus, the Bayes factor in favor of this model (relative to the unrestricted model) is

$$\text{BF} = \frac{\int C(\theta)\, p(y|\theta)\, p(\theta)\, d\theta}{\int p(y|\theta)\, p(\theta)\, d\theta} = E^{p(\theta|y)}[C(\theta)], \tag{2.38}$$

where the expectation is taken with respect to the posterior distribution of θ from the unrestricted model.

Given draws of $\{\theta^{(r)}\}_{r=1}^R$ from the posterior of the unrestricted model, it follows from the right-hand side of (2.38) that we can compute the Bayes factor as

$$\lim_{R\to\infty} R^{-1} \sum_{r=1}^{R} C\big(\theta^{(r)}\big) = \text{BF}. \tag{2.39}$$

In addition, the weights $z^{(r)} \propto C\big(\theta^{(r)}\big)$, where $\sum_{r=1}^R z^{(r)} = 1$, can be used to resample the draws. If the weights do not vary very much, then the two models will predict similar distributions for the parameters even if the Bayes factor is large.

The upshot is that proceed by first estimating without the option-price nonnegativity restriction and then to check afterwards to see how the restriction changes the inferences. (Limited experience suggests the restriction does not change the inferences noticeably.)

3. Month-average target rates, target-rate paths, and unscheduled meetings

The target fed funds rate is set by the Federal Open Market Committee (FOMC) at its scheduled meetings. There are eight scheduled meetings per year, so there are four months each year without a (scheduled) meeting. (Occasionally it is changed at unscheduled meetings. We will deal with this possibility in the Appendix.) It is set in increments of 25 basis points (0.25 percent).[12] [Currently, however, the target rate is a band: 0 to .25%.]

Consider a given meeting. We adopt the setup outlined in the previous section for the target rate R where there are K possibilities, $r = (r_1, \ldots, r_K)$, and the risk-neutral probability is $p(R = r_j|\beta) = \beta_j$. We say that r is *contiguous* if $r_j = r_1 + 0.25\%\,(j-1)$ for $j = 1, \ldots, K$.[13] We will often assume the model is contiguous.

Now consider a given month. The fed funds futures contract depends on the monthly average of fed funds rates. (For example, a Friday rate counts three days in the average, for Friday, Saturday, and Sunday. If the funds market were closed the following Monday due to a holiday, then the preceding Friday rate would count 4 days, assuming all four days were in the same month.)

Up to this point, we have implicitly assumed a single target applies for a given month. In general, however, the situation is more complicated. More often than not, there is a scheduled meeting within a given month. Prior to the meeting the previously established

[12] Cite some history.

[13] $0.25\% = 0.0025$.

target rate applies, while after meeting a new (possible different) applies. Assuming the Open Market Desk keeps the daily funds rate at the appropriate target rate throughout the month, then the month-average funds rate will be $\overline{R} = (1 - \omega)\, R_1 + \omega\, R_2$, where ω is the fraction of the days in the month for which R_2 is the target rate. If there were no meetings between the observation date and the meeting in the month in question, then R_1 can be identified with the current rate (assuming the absence of unscheduled meetings). More generally, both R_1 and R_2 will be unknown on the observation date and consequently the joint distribution of the two target rates come into play.

In addition, the daily effective funds rate does not equal to the target rate on a daily basis. More importantly for our purposes, the month-average funds rate S does not equal the month-average target rate $\overline{R}$. We account for this difference by letting $S = \overline{R} + u$, where v is the month-average slippage.

Notation. Let $\tau = (\tau_1, \ldots, \tau_w)$ denote a sequence of meeting dates and let R_{τ_i} denote the target rate established at time τ_i. Let T_m^0 denote the first day of month m. Let τ_m^- denote the date of the last meeting prior to the beginning of month m and let τ_m^+ denote the date of the following meeting: $\tau_m^- < T_m^0 \leq \tau_m^+$. Let ω_m denote the fraction of month m from τ_m^+ to the beginning of the month $m+1$:

$$\omega_m = \max\left\{\frac{T_{m+1}^0 - \tau_m^+}{T_{m+1}^0 - T_m^0},\, 0\right\} \tag{3.1}$$

In particular, if $\tau_m^+ = T_m^0$ then $\omega_m = 1$ and if $\tau_m^+ \geq T_{m+1}^0$ then $\omega_m = 0$. The month-average target rate for month m is

$$\overline{R}_m = (1 - \omega_m)\, R_{\tau_m^-} + \omega_m\, R_{\tau_m^+}. \tag{3.2}$$

Let t denote the quote date (also known as the observation date). If $t < \tau_i$ then R_{τ_i} is unknown while if $t \geq \tau_i$ then R_{τ_i} is known. Note τ_{m+1}^- is the date of the last meeting prior to the beginning of month $m+1$. If month m has a meeting then τ_{m+1}^- occurs in month m; otherwise τ_{m+1}^- occurs in an earlier month. If $t \geq \tau_{m+1}^-$, then $\overline{R}_m$ is known. If $t < \tau_{m+1}^-$, then $\overline{R}_m$ is unknown. If $t < \tau_m^-$ and there is a meeting during month m, the $\overline{R}_m$ involves two unknowns: $R_{\tau_m^-}$ and $R_{\tau_m^+}$.

A number of special cases. Now we consider a number of special cases of (3.2). In each case we show how the likelihood of an observation can be expressed as $y_i = X_i \beta + \varepsilon_i$, thereby allowing us to adopt both the likelihood and the prior described in Section 2.

Case 1. First consider a month that involves a single target rate. This occurs if either $\omega_m = 0$ (a month with no meeting) or $\omega_m = 1$ (a month with a meeting on the first day of the month). In either case, the month-average target rate involves a single target rate:

$$\overline{R}_m = \begin{cases} R_{\tau_m^-} & \omega_m = 0 \\ R_{\tau_m^+} & \omega_m = 1 \end{cases}. \tag{3.3}$$

Let us refer to this rate as R. Let $R \in r = \{r_1, \ldots, r_K\}$, $p(R = r_j | \beta) = \beta_j$, $S = R + u$, and $X_{ij} = \varphi(\mathcal{K}_i, \gamma_i, r_j + u)$. We have established $y_i = X_i \beta + \varepsilon_i$.

Case 2. Next suppose $\omega_m \in (0,1)$ but that $R_{\tau_m^-}$ is known. (Thus the month contains a single unknown target rate.) Let us refer to $R_{\tau_m^-}$ and $R_{\tau_m^+}$ as R_0 and R (respectively) so that $\overline{R}_m = (1-\omega_m)\,R_0 + \omega_m\,R$. Let $R \in r = \{r_1, \ldots, r_K\}$, $p(R = r_j|\beta) = \beta_j$, $S_m = \overline{R}_m + u$, and $X_{ij} = \varphi(\mathcal{K}_i, \gamma_i, \overline{r}_j + u)$ where $\overline{r}_j = (1-\omega_m)\,R_0 + \omega_m\,r_j$. Again, we have established $y_i = X_i\beta + \varepsilon_i$ (in which $\overline{r}_j$ has taken the place of r_j). In passing, note that the smaller ω_m is, the less information there will be regarding R.

Case 3. Now consider (3.2) more generally. Let us refer to $R_{\tau_m^-}$ and $R_{\tau_m^+}$ as R_1 and R_2, respectively, so that $\overline{R}_m = (1-\omega_m)\,R_1 + \omega_m\,R_2$. Assume $(R_1, R_2) \in r^1 \times r^2$ where $r^1 = (r^1_1, \ldots, r^1_{K_1})$ and $r^2 = (r^2_1, \ldots, r^2_{K_2})$. Let b denote a $K_1 \times K_2$ matrix of joint probabilities such that

$$p(R_1 = r^1_k, R_2 = r^2_\ell|b) = b_{k\ell} \tag{3.4}$$

where $b_{k\ell} \geq 0$ and $\sum_{k=1}^{K_1}\sum_{\ell=1}^{K_2} b_{k\ell} = 1$. Let β denote a vectorized version of b, where (using row-major order)

$$\beta_j = b_{k\ell} \quad \text{for} \quad j = (k-1)\,K_2 + \ell. \tag{3.5}$$

Thus $\beta \in \Delta^{K^*-1}$ where $K^* = K_1\,K_2$.

Assume $S_m = \overline{R}_m + u$ and let $x_{ik\ell} = \varphi(\mathcal{K}_i, \gamma_i, \overline{r}_{k\ell} + u)$ where

$$\overline{r}_{k\ell} = (1-\omega_m)\,r^1_k + \omega_m\,r^2_\ell. \tag{3.6}$$

Then

$$y_i = \sum_{k=1}^{K_1}\sum_{\ell=1}^{K_2} x_{ik\ell}\,b_{k\ell} + \varepsilon_i. \tag{3.7}$$

Let X_i denote the vectorized version of x_i. Then (3.7) can be expressed as $y_i = X_i\beta + \varepsilon_i$. Stacking the observations produces $y = X\beta + \varepsilon$. The matrix X has n rows/observations and K^* columns/paths and the vector β has K^* elements/joint probabilities.

There are a number of factors that militate against the ability to identify the probabilities in this case: the (potentially) large number of unknown probabilities relative to the number of observations, the (potentially) close spacings of the states $s_{k\ell}$ (relative to the magnitude of the slippage and the measurement error), the possibility that ω_m is near either zero or one. One possible way to overcome these problems is to use information from additional contract months. We now turn to that approach.

Two contract months at once. Suppose there is no meeting in month m ($\omega_m = 0$) and there is a meeting in month $m+1$. Let $t < \tau_m^-$ so that both $\overline{R}_m = R_{\tau_m^-}$ and $\overline{R}_{m+1} = (1-\omega_{m+1})\,R_{\tau_{m+1}^-} + \omega_{m+1}\,R_{\tau_{m+1}^+}$ are unknown. The month-average rates share a common target rate: $R_{\tau_{m+1}^-} = R_{\tau_m^-}$. We refer to $R_{\tau_m^-}$ and $R_{\tau_m^+}$ as R_1 and R_2, respectively, and we adopt the previous setup (including r^1, r^2, b, and its vectorization β). In addition, we refer to the months m and $m+1$ as months 1 and 2, respectively. Thus, $\overline{R}_1 = R_1$ and $\overline{R}_2 = (1-\alpha_2)\,R_1 + \alpha_2\,R_2$.

Assume each month has its own slippage: $S_m = \overline{R}_m + u_m$, for $m = 1, 2$. Let $x^m_{ik\ell} = \varphi(\mathcal{K}^m_i, \gamma^m_i, s^m_{k\ell} + u_m)$, where

$$s^1_{k\ell} = r^1_k \tag{3.8}$$

$$s^2_{k\ell} = (1 - \omega_2)\, r^1_k + \omega_2\, r^2_\ell. \tag{3.9}$$

(Note the matrix s^1 is composed of K_2 copies of the vector r^1.) Let X^m_i denote the vectorized version of $x^m_{ik\ell}$. Then we have

$$y^m_i = X^m_i \beta + \varepsilon^m_i. \tag{3.10}$$

Let

$$Y = \begin{bmatrix} y^1 \\ y^2 \end{bmatrix}, \quad \mathcal{X} = \begin{bmatrix} X^1 \\ X^2 \end{bmatrix}, \quad \text{and} \quad E = \begin{bmatrix} \varepsilon^1 \\ \varepsilon^2 \end{bmatrix}. \tag{3.11}$$

We can express the joint likelihood as $Y = \mathcal{X}\beta + E$, where $E \sim \mathrm{N}(0_{n^*}, \sigma^2 I_{n^*})$ and where $n^* = n_1 + n_2$. Consequently, we can use the posterior sampler described above.[14]

This approach can be generalized to include more than two months where β is a vectorized version of the tensor of joint probabilities and $\beta \in \Delta^{K^*-1}$, where $K^* = \prod_u K_u$. (See Appendix A for an example involving three contract months.)

Remark. Recall that α is the concentration parameter in the prior for β. Note that if $\alpha < K^* - \max\{K_u\}$ then there is no restricted model that supports independence among the probabilities.[15] In effect, a sufficient small value of α for the joint probabilities rules out independence.

Unscheduled meetings. In order to keep the analysis simple, consider a month with no scheduled meeting. Let R_1 denote the applicable target rate absent the unscheduled meeting and let R_2 denote the target rate set at the unscheduled meeting (if it occurs). Note that $R_2 \neq R_1$.

Let the month-average target rate be given by

$$\overline{R} = (1 - \omega)\, R_1 + \omega\, R_2. \tag{3.12}$$

The novelty is that ω is unknown. If $\omega = 0$, there is no unscheduled meeting, while if $0 < \omega \leq 1$ there is an unscheduled meeting. The distribution for ω is discrete because there are a finite number of days in any month. (The values of ω with positive probability are further restricted by the fact that unscheduled meetings "occur" on business days in the sense that the announcement of a new target rate occurs on business days.)

To treat the simplest case, suppose R_1 is known, in which case $\overline{r}_j = (1 - \omega)\, R_1 + \omega\, r^2_j$ (where $\overline{r}_j = R_1$ if and only if $\omega = 0$) and $X_{ij} = \varphi(\mathcal{K}_i, \gamma_i, \overline{r}_j + u)$. Note the X matrix depends on the unknown ω (as well as u). The (conditional) posterior distribution for ω is discrete where

$$p(\omega|\beta, u, \sigma^2, y) \propto \exp\left(\frac{-S(\beta, u, \omega)}{2\,\sigma^2}\right) p(\omega). \tag{3.13}$$

We can think of (3.12) as combining two competing models: the no-unscheduled-meeting model N characterized by $\omega = 0$ and the unscheduled-meeting model U characterized by

[14] See Appendix A for an alternative formulation in terms of marginal and conditional probabilities, with an illustration assuming $\omega_2 = 1$.

[15] At least no simple model that I can see.

$\omega \neq 0$. The Bayes factor in favor of model N relative to model U is given by the posterior odds ratio divided by the prior odds ratio:

$$\mathrm{BF}_U^N = \frac{p(\omega = 0|y)}{p(\omega \neq 0|y)} \div \frac{p(\omega = 0)}{p(\omega \neq 0)}. \tag{3.14}$$

Let the prior distribution $p(\omega|q)$ be given by $p(\omega = 0|q) = q$ and $p(\omega \neq 0|q) = 1-q$, where $1-q$ is distributed equally over the remaining possibilities. Suppose the hyperparameter q has a prior distribution, $p(q)$. Then $p(\omega = 0) = \int p(\omega = 0|q)\, p(q)\, dq = E[q]$ and $p(\omega = 0|y) = \int p(\omega|q)\, p(q|y)\, dq = E[q|y]$. For example, suppose the prior for q is a beta distribution parameterized as follows:

$$p(q) = \mathrm{B}\big(q|\alpha_q\, \overline{\xi}_q, \alpha_q\, (1 - \overline{\xi}_q)\big), \tag{3.15}$$

where $\overline{\xi}_q = E[q]$ and α_q is the concentration parameter. The flat prior is given by $\overline{\xi}_q = 1/2$ and $\alpha_q = 1$. If in fact we choose $\overline{\xi}_q = 1/2$, then

$$\mathrm{BF}_U^N = \frac{E[q|y]}{1 - E[q|y]}. \tag{3.16}$$

However, it probably makes more sense to choose $\overline{\xi}_q$ to match the unconditional probability that a month has no unscheduled meeting. If we choose $\alpha_q \ll 1$, then the prior expresses the following view: On the one hand it is probably very unlikely that there will be an unscheduled meeting, but on the other it is possible that it is very likely.

4. Data

Here we describe the data and what is required to put it in a form suitable for estimation. For each quote date there are quotes for puts and calls on a number of contracts.

The strike prices are expressed in terms of an index P, where $P = 100\,(1 - S)$.[16] For example, $P = 97$ refers to a month-average funds rate of $S = 1 - .97 = .03$. Note that the payoff at expiration to a call option in terms of P translates into the payoff of a put in terms of S. To see this, let $\mathcal{K}_P = 100\,(1 - \mathcal{K}_S)$ denote the two expressions of a given strike price and note

$$(P - \mathcal{K}_P)^+ = \big(100\,(1 - S) - 100\,(1 - \mathcal{K}_S)\big)^+ = 100\,(\mathcal{K}_S - S)^+. \tag{4.1}$$

Similarly, put payoffs expressed in terms of P become call payoffs expressed in terms of S. Hereafter, assume the data are expressed in terms of S.

Let t denote the quote date and m denote the contract month. The data for a given quote date include quotes for a total of M_t contract months, and for each contract month m are in the form of n_t^m triples: $\{(\mathcal{K}_{ti}^m, \gamma_{ti}^m, V_{ti}^m)\}_{i=1}^{n_t^m}$. We deflate the option values: $y_{ti}^m = V_{it}^m / B_t^m$ for $i = 1, \ldots, n_t^m$. The result is a data set for the quote date t of the form $\{(\mathcal{K}_{ti}^m, \gamma_{ti}^m, y_{ti}^m)\}_{i=1}^{n_t^m}$ for $m = 1, \ldots, M_t$.

It remains to compute the appropriate X matrices from the strike prices $\{\mathcal{K}_{ti}^m\}$. We proceed as follows. (We will suppress the dependence on the quote date occasionally to reduce the notational clutter.) First, for each contract month in the data set we need to

[16] Because the relation between S and P is linear, there are no issues relating to Jensen's inequality in changing back and forth between the two representations.

find α_m, $R_{\tau_m^-}$ and $R_{\tau_m^+}$ in order to form $\overline{R}_m$ as given in (3.2). Next compute the union of the unknown rates across entire collection of months. Let R_0 denote the known target rate as of time t and let $\{R_u\}_{u=1}^{U_t}$ denote the set of unknown rates that appear in the month-average rates. (At this point, some discretion on the part of the researcher comes into play.) For each unknown rate, specify the set of possible values: $R_u \in r^u = \{r_1^u, \ldots, r_{K_u}^u\}$. Let $r_0 = \{R_0\}$. It would probably make sense to have $r_0 \subseteq r_1 \ldots \subseteq r_u$. The total number of target-rate paths is $K_t^* = \prod_{u=1}^{U_t} K_u$.

If all of the data were to be used in a single joint estimation, the X matrix would have $n_t^* = \sum_{m=1}^{M_t} n_t^m$ rows/observations and K_t^* columns/paths. It may not be possible to learn much for such a model, given the number of observations and the number of paths. Instead a subset of the months and a corresponding subset of the target rates may be used. Given these subsets, let $X_{ij}^m = \varphi(\mathcal{K}_i^m, \gamma_i^m, s_j^m + u_m)$, where the index $j = 1 \ldots, K^*$ is computed according the vectorization of the join probability tensor for the given subset (and K^* now denotes the product computed from the subset). The data are now in a form suitable for estimation.

Additional thoughts. We will want to compare futures prices (as proxies for forward prices) with $E[S|y]$ computed from the posterior predictive distribution [see (2.18)]. We may want to use the futures prices to help form the prior for β. We can use put–call parity to assess the magnitude of the measurement error.

Questions for me. What is the expiration day if the last day of the month falls on a nonbusiness day? Does the funds rate for a given day include trades after the FOMC announcement. What is the effective funds rate on meeting days, both with and without target rate changes. Use this information to decide whether to include the meeting day with the previous target rate or with the next target rate.

5. Dynamics

We can extend this approach to incorporate day-to-day dynamics. The idea here is that β is a stochastic process that evolves through time. Since β is a vector of probabilities, it must be a martingale, where[17] $E_{t-1}[\beta_t] = \beta_{t-1}$. Let $p(\beta_t|\beta_{t-1})$ denote the pdf for β_t given the parameter β_{t-1} and let $Y_{t-1} := (y_1, \ldots, y_{t-1})$. Then

$$p(\beta_t|y_t, Y_{t-1}) \propto p(y_t|\beta_t)\, p(\beta_t|Y_{t-1}) = p(y_t|\beta_t)\, p(\beta_t|\beta_{t-1})\, p(\beta_{t-1}|Y_{t-1}). \tag{5.1}$$

There are a number of ways to compute the updating in (5.1). The particle filter is discussed below (as an illustrative example).

The central idea here is that the probabilities on a given day (for a fixed set of target rates) will be closely related to the probabilities on the preceding and succeeding days. One consequence is that more efficient estimation could use the information in the data from the neighboring days (y_{t-1}, y_{t+1}). Another consequence is that if we wish to infer whether the probabilities have changed from one day to the next, we must compute their joint distribution $p(\beta_{t-1}, \beta_t)$.

[17] Warning: The notation in this section may conflict with that in other sections.

Particle filter. The updating in (5.1) can be computed using a *particle filter*. In particular, the draws from $p(\beta_{t-1}|Y_{t-1})$ constitute a *particle swarm*. There are a number of related ways to implement a particle filter.

The most straightforward is the sampling–importance resampling (SIR) approach. Following this approach, make a random draw from $p(\beta_t|\beta_{t-1}^{(i)})$ for each $\beta_{t-1}^{(i)}$ in the swarm. Denote these draws $\{\beta_t^{(i)}\}$. These draws represent draws from $p(\beta_t|Y_{t-1})$. At this point, importance sampling comes in to play. Compute $w_t^{(i)} := p(y_t|\beta_t^{(i)})$ and resample $\{\beta_t^{(i)}\}$ using these weights. A problem one encounters is when the bulk of $p(y_t|\beta_t)$ is located in the tail of $p(\beta_t|Y_{t-1})$ then the weights $\{w_t^{(i)}\}$ may be dominated by a few large values with the result that the swarm provides a poor representation of the posterior.

An alternative approach is the independent particle filter (IPF). Make draws $\{\beta_t^{(j)}\}_{j=1}^m$ from the likelihood $p(y_t|\beta_t)$ assuming it is proper.[18] We need to compute the weights

$$p(\beta_t^{(j)}|Y_{t-1}) = p(\beta_t^{(j)}|\beta_{t-1})\, p(\beta_{t-1}|Y_{t-1}).$$

The factor $p(\beta_{t-1}|Y_{t-1})$ is encoded in the posterior draws from the previous day.[19] For each draw from the second day, choose L draws from the first day, where $1 \leq L \leq m$ and m is the number of draws from the first day. Then compute $z_t^{(i,j)} = p(\beta_t^{(j)}|\beta_{t-1}^{(i)})$. Let $z_t^{(j)} = \sum_{i=1}^L z_t^{(i,j)}$ and use these weights to resample $\{\beta_t^{(j)}\}_{j=1}^m$. (By sampling from $p(\beta_{t-1}|Y_{t-1})$, we incorporate its weight implicitly.) We can tweak the IPF by drawing from $p(y_t|\beta_t)\, p(\beta_t)$ where $p(\beta_t)$ is the maximum entropy distribution with mean $= \frac{1}{M}\sum_{i=1}^M \beta_{t-1}^{(i)}$. (We could use the Dirichlet distribution as well.) This will tend to reduce the variation of the weights computed as $p(\beta_t|\beta_{t-1})/p(\beta_t)$.

In order to obtain the joint distribution $p(\beta_{t-1}, \beta_t|Y_{t-1}, y_t)$, we can apply the so-called smoothing step. We do this as follows. Make a draw from $\{\beta_t^{(j)}\}_{j=1}^m$ and conditional on that draw, make a draw from $\{\beta_{t-1}^{(i)}\}_{i=1}^m$ using the weights $\{z_t^{(i,j)}\}_{i=1}^m$ (fixing the j drawn). The pairs thus generated are from the joint distribution conditional on (Y_{t-1}, y_t). The marginal distribution for β_{t-1} encoded in these draws reflects the data from both days.

The computation of $\{z_t^{(i,j)}\}$ can be quite expensive. One can reduce the expense by discretizing—by binning and computing $p(\beta_t^{(j)}|\beta_{t-1}^{(i)})$ at the midpoints of the bins and using the bin counts to adjust. Let $\widetilde{z}_t^{(i,j)} := p(\beta_t^{(j)}|\beta_{t-1}^{(i)})\, n_t^{(i,j)}$ where $p(\beta_t^{(j)}|\beta_{t-1}^{(i)})$ is computed at the midpoint of bin (i,j) and $n_t^{(i,j)}$ denotes the corresponding bin count. Then $\widetilde{z}_t^{(j)} := \sum_i \widetilde{z}_t^{(i,j)}$.

Appendix A. Addition material

Marginal and conditional target-rate probabilities. Given the joint probabilities (3.4), we can compute marginal probabilities $p(R_1 = r_k^1|b) = \sum_{\ell=1}^{K_2} b_{k\ell} = \beta_k^1$ and $p(R_2 = r_\ell^2|b) =$

[18]The draws of β_t are made independent of β_{t-1}. For more on the IPF, see Lin et al. (2005).

[19]One way to proceed is to take a draw from $p(\beta_t|\beta_{t-1}^{(i)})$ for each i from the first day and then approximate the density.

$\sum_{k=1}^{K_1} b_{k\ell} = \beta_\ell^2$ and conditional probabilities

$$p(R_2 = r_\ell^2 | R_1 = r_k^1, b) = \frac{p(R_1 = r_k^1, R_2 = r_\ell^2 | b)}{p(R_1 = r_k^1 | b)} = \frac{b_{k\ell}}{\sum_{\ell=1}^{K_2} b_{k\ell}} =: B_{k\ell}. \tag{A.1}$$

Note $\beta^2 = B^\top \beta^1$. Also note $p(R_1 = r_k^1, R_2 = r_\ell^2 | B, \beta^1) = B_{k\ell}\, \beta_k^1$.

As an alternative to modeling the joint probabilities, one could model marginal and conditional probabilities. If we choose to model β^1 and B (instead of modeling b), the number of parameters remains $K_1 K_2 - 1$, since $\sum_{k=1}^{K_1} \beta_k^1 = 1$ and $\sum_{\ell=1}^{K_2} B_{k\ell} = 1$ for $k = 1, \ldots, K_1$. (Each of the K_1 rows of B sums to one.) We could adopt Dirichlet priors for β^1 and for the columns of B. This approach, however, offers no advantages in general relative to the approach of modeling the joint probabilities.

In order to illustrate this approach, let us modify the example involving two months by setting $\omega_2 = 1$. In this case each month is associated with a single target rate and consequently

$$\begin{bmatrix} X^1 \\ X^2 \end{bmatrix} \beta = \begin{bmatrix} \widetilde{X}^1 & 0 \\ 0 & \widetilde{X}^2 \end{bmatrix} \begin{bmatrix} \beta^1 \\ \beta^2 \end{bmatrix} \tag{A.2}$$

subject to $\beta^2 = B^\top \beta^1$, where

$$\widetilde{X}_{ij}^m = \varphi(\mathcal{K}_i^m, \gamma_i^m, r_j^m + u_m) \tag{A.3}$$

for $r_j^m \in r^m$.

Carlson et al. (2005) provide an example of this approach (with $u_m = 0$ in their setup). In the example (p. 1214), $K_1 = 2$ and $K_2 = 3$. Consequently there are $K^* = K_1 K_2 = 6$ possible paths and $K^* - 1 = 5$ independent probabilities. Using a prior information, they set three path probabilities to zero, leaving three paths and two independent probabilities remaining. The restrictions produce the following matrix of conditional probabilities:[20]

$$B = \begin{bmatrix} B_{11} & 0 & 0 \\ 0 & B_{22} & B_{23} \end{bmatrix}, \tag{A.4}$$

so that

$$\beta^2 = B^\top \beta^1 = \begin{bmatrix} B_{11}\, \beta_1^1 \\ B_{22}\, \beta_2^1 \\ B_{23}\, \beta_2^1 \end{bmatrix}. \tag{A.5}$$

Since each row of B must sum to one, we have $B_{11} = 1$ and $B_{22} + B_{23} = 1$. In addition $\beta_1^1 + \beta_2^1 = 1$. (Together these restrictions imply $\beta_1^2 + \beta_2^2 + \beta_3^2 = 1$.) This leaves a total of two free parameters.

[20]The matrix of joint probabilities is

$$b = B * \beta^1 = \begin{bmatrix} B_{11}\, \beta_1^1 & 0 & 0 \\ 0 & B_{22}\, \beta_2^1 & B_{23}\, \beta_2^1 \end{bmatrix},$$

where (row-by-row) $b_k = B_k\, \beta_k^1$.

An example involving three months. Here we provide an example involving three months. Let

$$\overline{R}_m = (1-\omega_m)\, R_{\tau_m^-} + \omega_m\, R_{\tau_m^+} \tag{A.6a}$$

$$\overline{R}_{m+1} = (1-\omega_{m+1})\, R_{\tau_{m+1}^-} + \omega_{m+1}\, R_{\tau_{m+1}^+} \tag{A.6b}$$

$$\overline{R}_{m+2} = (1-\omega_{m+2})\, R_{\tau_{m+2}^-} + \omega_{m+2}\, R_{\tau_{m+2}^+}. \tag{A.6c}$$

If $\omega_{m+1} = 1$, then there is no rate in common between $\overline{R}_m$ and $\overline{R}_{m+1}$. Similarly, if $\omega_{m+2} = 1$, then there is no rate in common between $\overline{R}_{m+1}$ and $\overline{R}_{m+2}$. Assuming $\omega_{m+1} < 1$,

$$R_{\tau_{m+1}^-} = \begin{cases} R_{\tau_m^-} & \omega_m = 0 \\ R_{\tau_m^+} & \omega_m > 0 \end{cases} \tag{A.7}$$

and assuming $\omega_{m+2} \neq 1$,

$$R_{\tau_{m+2}^-} = \begin{cases} R_{\tau_{m+1}^-} & \omega_{m+1} = 0 \\ R_{\tau_{m+1}^+} & \omega_{m+1} > 0 \end{cases}. \tag{A.8}$$

Suppose $\omega_m, \omega_{m+1} \in (0,1)$ and $\omega_{m+2} = 0$ so that

$$\overline{R}_m = (1-\omega_m)\, R_1 + \omega_m\, R_2 \tag{A.9}$$

$$\overline{R}_{m+1} = (1-\omega_{m+1})\, R_2 + \omega_{m+1}\, R_3 \tag{A.10}$$

$$\overline{R}_{m+2} = R_3, \tag{A.11}$$

where (R_1, R_2, R_3) represent the three distinct rates. Let $r^m = (r_1^m, \ldots, r_{K_m}^m)$ for $m = 1, 2, 3$ and let $p(R_1 = r_k^1, R_2 = r_\ell^2, R_3 = r_w^3 | \mathcal{B}) = \mathcal{B}_{k\ell w}$, where $\mathcal{B}$ is a rank-3 tensor of joint probabilities with dimensions $K_1 \times K_2 \times K_3$. Let β denote the vectorized version $\mathcal{B}$:

$$\beta_j = \mathcal{B}_{k\ell w} \quad \text{for} \quad j = (k-1)\, K_2\, K_3 + (\ell - 1)\, K_3 + w. \tag{A.12}$$

Note $\beta \in \Delta^{K^*-1}$, where $K^* = K_1\, K_2\, K_3$.

Assume $S_m = \overline{R}_m + u_m$. Let $x_{ik\ell w}^m = \varphi(\mathcal{K}_i^m, \gamma_i^m, s_{k\ell w}^m + u_m)$, where

$$s_{k\ell w}^1 = (1-\omega_m)\, r_k^1 + \omega_m\, r_\ell^2 \tag{A.13}$$

$$s_{k\ell w}^2 = (1-\omega_{m+1})\, r_\ell^2 + \omega_{m+1}\, r_w^3 \tag{A.14}$$

$$s_{k\ell w}^3 = r_\ell^3. \tag{A.15}$$

Then we have

$$y_i^m = X_i^m \beta + \varepsilon_i^m \tag{A.16}$$

where X_i is a vectorized version of the tensor x_i. Again we can express the joint likelihood as $Y = \mathcal{X}\beta + E$, where $E \sim \mathrm{N}(0_{n^*}, \sigma^2 I_{n^*})$ and where $n^* = n_1 + n_2 + n_3$.

References

Carlson, J. B., B. R. Craig, and W. R. Mellick (2005). Recovering market expectations of FOMC rate changes with options on federal funds futures. *Journal of Futures Markets 25*, 1203–1242.

Geweke, J. (1986). Exact inference in the inequality constrained constrained normal linear regression model. *Journal of Applied Econometrics 1*, 127–142.

Lin, M. T., J. L. Zhang, Q. Cheng, and R. Chen (2005). Independent particle filters. *Journal of the American Statistical Soceity 100*, 1412–1421.

Federal Reserve Bank of Atlanta, 1000 Peachtree St. N.E., Atlanta, GA 30309