

Market Probability Tracker: Estimation specifics

In this note we supply additional detail about the way in which the probability distributions are estimated. This note is an adjunct to the document “Simplex Regression” to which we refer throughout.

The model is structured as a regression characterized by:

$$y_i = \lambda \sum_{j=1}^K X_{ij} \beta_j + \varepsilon_i,$$

where y_i is the price of the i -th option, λ is the discount factor, $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ is a measurement error, and

$$X_{ij} = \int_{-\infty}^{\infty} g_i(x) f_{ij}(x) dx.$$

In addition, $\beta = (\beta_1, \dots, \beta_K)$ is a collection of nonnegative weights that sum to one. The option’s payoff is given by $g_i(x)$ and basis distributions are given by

$$f_{ij}(x) = \text{Beta}(H(x)|j, K - j + 1)h(x),$$

where $H(x)$ denotes the CDF (cumulate distribution function) for a continuous random variable on the real line and $h(x) = H'(x)$ denotes the associated density.

We set H to be the normal distribution: $H = N(\mu, \sigma^2)$, for some mean μ and standard deviation σ . In particular,

$$H(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

where $\Phi(\cdot)$ is the CDF for the standard normal distribution. In addition,

$$h(x) = N(x|\mu, \sigma^2).$$

We center this distribution on the forward rate: $\mu = F$. We choose σ to be large enough to allow for a substantial amount of probability away from the center. After some experimentation, we have chosen to set σ to be three times the implied standard deviation from fitting the option premiums to a Gaussian distribution. This choice of σ means that H is fairly wide; consequently, we have chosen $K = 80$, which allows the basis distributions to be narrow enough to capture the kinds of detail we are trying to infer.

There are two ways we obtain the discount factor B and forward rate F . One way is to use put-call parity on pairs of options that have common strike prices. Put-call parity can be stated as:

$$P_i - C_i = B(F - K_i),$$

where P_i is the put price, C_i is the call price, and K_i is the common strike price for the i -th pair of options. Given a collection of such pairs, we can use OLS to estimate B and F .

If we do not have sufficient overlapping strikes, we estimate F and B from market data as follows. We take the forward rate directly from the futures contract (we do not adjust for the difference between forward and futures prices), and we take the discount factor from the Treasury curve: we linearly interpolate the discount rate

from constant maturity rates and convert to a discount factor via continuous compounding.

As noted above, we use the forward rate F to center H . By contrast, we treat the discount rate λ as a latent variable to be estimated and we use the “observed” discount rate B only as a starting value for the estimation (which is described in “Simplex Regression”). We assume the prior for λ is a truncated normal with a mean and standard deviation of 1 and 10^6 , respectively.

The prior distribution for β is Dirichlet with a mean of $\xi = (\frac{1}{K}, \dots, \frac{1}{K})$ and a concentration parameter α . The prior for α is lognormal with both parameters set to one.

We use the first posterior sampler for β described in Section 5 (not the alternative approach).