

APPENDIX

A Extended Model: Posterior Distributions

A.1 Homoskedastic errors

Consider the basic contingent claim model b extended by the vector of observables x_2 :

$$\begin{aligned} \log C_i &= \beta_1 \log b(\sigma, x_{1i}) + \underline{\beta}'_2 x_{2i} + \eta_i, \quad i = 1, \dots, N & \eta_i &\sim N(0, \sigma_\eta) \\ R_t &= \mu + \xi_t, & t = 1, \dots, T_r & \xi_t \sim N(0, \sigma) \quad E(\xi_t \eta_i) = 0 \end{aligned}$$

The likelihood function is:

$$\ell(\sigma_\eta, \sigma, \beta_1, \beta'_2 | \underline{y}_t) \propto \frac{1}{\sigma_\eta^N} \times \exp \left\{ -\frac{\sum_1^N [\log C_i - \beta_1 \log b(\sigma, x_{1i}) - \beta'_2 x_{2i}]^2}{2\sigma_\eta^2} \right\} \quad (1)$$

Now define: $x'_i = (\log b(\sigma, x_{1i}), x'_{2i})$, $X' = (x_1, \dots, x_N)$, $\beta' = (\beta_1, \beta'_2)$, and $Y' = (\log C_1, \dots, \log C_N)$.

We formulate the following joint prior distribution for the parameters:

$$\begin{aligned} p(\sigma, \sigma_\eta, \beta) &= p(\sigma)p(\sigma_\eta)p(\beta | \sigma_\eta) \\ &= \text{IG}(\sigma : \nu_0, s_0^2) \text{IG}(\sigma_\eta : \nu_1, s_1^2) \text{N}(\beta : \beta_0, \sigma_\eta^2 V_0) \end{aligned}$$

The fact that $p(\sigma)$ is based on the T_r returns data, can be incorporated by setting $\nu_0 s_0^2 = \sum_1^{T_r} (R_t - \bar{R})^2$ and $\nu_0 = T_r - 1$. By Bayes theorem, the joint density of the parameters is

$$p(\sigma_\eta, \sigma, \beta | \underline{y}_t) \propto \frac{\exp \left\{ -\frac{\nu_0 s_0^2}{2\sigma^2} - \frac{\nu_1 s_1^2}{2\sigma_\eta^2} \right\}}{\sigma^{\nu_0+1} \sigma_\eta^{N+\nu_1+1}} \times \exp \left\{ -\frac{(Y - X\beta)' (Y - X\beta) + (\beta - \beta_0)' V_0^{-1} (\beta - \beta_0)}{2\sigma_\eta^2} \right\}$$

Define:

$$\hat{\beta} = (X'X)^{-1} X'Y, \quad V = [X'X + V_0^{-1}]^{-1}, \quad \bar{\beta} = V [X'X \hat{\beta} + V_0^{-1} \beta_0],$$

and

$$\nu_\eta = N - k + \nu_1, \quad \nu_\eta s_\eta^2 = (Y - X\bar{\beta})' (Y - X\bar{\beta}) + (\beta_0 - \bar{\beta})' V_0^{-1} (\beta_0 - \bar{\beta}) + \nu_1 s_1^2.$$

The joint density becomes

$$p(\sigma_\eta, \sigma, \beta | \underline{y}_t) \propto \frac{\exp \left\{ -\frac{\nu_0 s_0^2}{2\sigma^2} \right\}}{\sigma^{\nu_0+1}} \times \frac{1}{\sigma_\eta^{\nu_\eta+1+k}} \times \exp \left\{ \frac{-\nu_\eta s_\eta^2 - (\beta - \bar{\beta})' V^{-1} (\beta - \bar{\beta})}{2\sigma_\eta^2} \right\}. \quad (2)$$

It is analogous to that resulting from a standard regression model with the twist that X , $\hat{\beta}$, $\bar{\beta}$, V , and $\nu_\eta s_\eta^2$, are functions of σ . We can now break down the joint density in the conditionals of interest.

First,

$$p(\beta \mid \sigma, \sigma_\eta, \underline{y}_t) \sim N(\bar{\beta}, \sigma_\eta^2 V). \quad (3)$$

The joint density of σ and σ_η is then

$$p(\sigma_\eta, \sigma \mid \underline{y}_t) \propto \frac{1}{\sigma^{\nu_0+1}} \times \exp\left\{-\frac{\nu_0 s_0^2}{2\sigma^2}\right\} \times \frac{1}{\sigma_\eta^{\nu_\eta+1}} \times \exp\left\{-\frac{\nu_\eta s_\eta^2}{2\sigma_\eta^2}\right\} \times |V|^{1/2}. \quad (4)$$

The conditional posterior density of σ_η is

$$p(\sigma_\eta \mid \sigma, \underline{y}_t) \sim \text{IG}(\nu_\eta = N - k + \nu_1, \nu_\eta s_\eta^2(\sigma)) \quad (5)$$

The posterior density of σ is

$$p(\sigma \mid \underline{y}_t) \propto \frac{1}{\sigma^{\nu_0+1}} \exp\left\{-\frac{\nu_0 s_0^2}{2\sigma^2}\right\} \times [\nu_\eta s_\eta^2(\sigma)]^{-\nu_\eta/2} \times |V|^{1/2}. \quad (6)$$

The distribution in equation (6) is the marginal posterior distribution of σ . A draw from σ_η , σ , and β can be made by a metropolis draw from expression (6) and then a direct draw from (5) and (3).

So no Gibbs step is required for the homoskedastic model.

Alternatively, note that the densities $p(\sigma_\eta \mid \sigma, \beta, \underline{y}_t)$, $p(\beta \mid \sigma, \sigma_\eta, \underline{y}_t)$, and $p(\sigma \mid \sigma_\eta, \beta, \underline{y}_t)$ are readily obtained by inspection of the joint posterior (2) density. They are the basis for a (un-needed) Gibbs cycle. Even with a Gibbs cycle a Metropolis step is still needed for σ .

A.2 σ : The Metropolis Step

This appendix discusses the σ draws. Since a Gibbs cycle will be needed when the errors are heteroskedastic, this discussion is based upon the conditional posterior distribution of σ , not the unconditional posterior in (6). It is

$$p(\sigma \mid \beta, \sigma_\eta, \underline{y}_t) \propto \frac{\exp\left\{-\frac{\nu_0 s_0^2}{2\sigma^2}\right\}}{\sigma^{\nu_0+1}} \times \exp\left\{-\frac{\nu s^2(\sigma, \beta)}{2\sigma_\eta^2}\right\}, \quad (7)$$

where $\nu s^2(\sigma, \beta) = (Y - X(\sigma)\beta)'(Y - X(\sigma)\beta)$. For computational convenience, we introduce the sample statistic $\nu_\eta s_\eta^{2\star}$, the mode of the kernel, and rewrite the posterior density of σ as

$$p(\sigma \mid \beta, \sigma_\eta, \underline{y}_t) \equiv \frac{K}{\nu_\eta s_\eta^{2\star}} \times \text{IG}(\nu_0, \nu_0 s_0^2) \times \exp\left\{-\frac{\nu s^2(\sigma, \beta)}{2\sigma_\eta^2}\right\} \quad (8)$$

We draw in sequence $(\beta | \cdot)$, $(\sigma | \cdot)$, and $(\sigma_\eta | \cdot)$, building a chain of such draws. There is no analytical expression for K , but it could be computed numerically by importance sampling from the first kernel in (8). This is unrealistic as. First, we would need to recompute K for every draw of σ . This is because β and σ_η change after each draw. Second, even then, direct draws from (8) by conventional methods such as inverse CDF are unrealistic. The Metropolis algorithm does not require the computation of K .

The Metropolis algorithm (see Metropolis et al. (1953) and Tierney (1991)) nests a simpler algorithm, the accept/reject, (see Devroye (1987)), which requires the knowledge of K . We explain the accept/reject algorithm first. We cannot draw directly from the density $p(\sigma)$. There is a *blanketing* density $q(\sigma)$ from which we can draw, and which meets the condition that there exists a finite number c such that $cq(\sigma) > p(\sigma)$, for all σ . Draw from q a number σ and accept the draw with probability $p(\sigma)/cq(\sigma)$. The intuition of why this produces a sample of draws with distribution $p(\sigma)$ is simple: We draw from q and for each draw we know by how much cq dominates p . p/cq is not the same for every value of σ because p and q do not have the same shape. The smaller p/cq , the more q dominates p , the more likely we are to draw too often in this area, the less likely the draw is to be accepted. If the parameter space is unbounded, a finite c such that $cq(\sigma) > p(\sigma), \forall \sigma$, exists only if the tail of q drops at a slower rate than the tail of p . For density (8), this can be accomplished if q is an inverted gamma with parameter $\nu \leq T_r - 1$. Given that c exists, an ideal density is such that p/q is relatively constant over σ . Otherwise c needs to be very large, and we will waste time rejecting many draws. Experimentation shows that the inverted gamma may have a shape different from (7), particularly if the option kernel is more informative than the prior kernel. This is because q must have low degrees of freedom ($\nu \leq T_r - 1$) for c to exist. q is not allowed to tighten when the information in the options data increases. An extreme case of this occurs if we only use option data. Also, the calculation of c is non trivial. One must first calculate K rather precisely, and then solve for the minimum of p/q over σ . So the accept-reject algorithm alone is unsatisfactory.

However, for any candidate density q , we always find c such that $cq > p$, for **most** values of σ . For some values of σ , $cq < p$, i.e., the density q does not dominate p everywhere. In these regions, we do not draw often enough from q , and therefore, underestimate the mass under the density p . The Metropolis algorithm is a rule of how to *repeat* draws, i.e., build mass for values of σ , where q does not

draw often enough. This does not require dominance everywhere, and gives us more choices for the density q and the number c . For a given density q , too large a c leads to frequent rejections, and too low a c produces many repeats, but the algorithm is still valid. A c which trades off these two costs can be computed very quickly. Furthermore we do not need to compute K anymore. This is because the transition kernel of the Metropolis is a function of the ratio $p(y) / p(x)$, where x and y are the previous and the current candidate draws. K disappears from the ratio. Consider an independence chain with transition kernel $f(z) \propto \min\{p(z), cq(z)\}$. The chain repeats the previous point x with probability $1-\alpha$, where $\alpha(x, y) = \min\left\{\frac{w(y)}{w(x)}, 1\right\}$, and $w(z) \equiv p(z)/f(z)$. If $cq > p$, $w(z)=1$, and if $cq < p$, $w(z) >1$. The decision to stay or move is based upon $\frac{w(y)}{w(x)}$ which compares the (lack of) dominance at the previous and the candidate points.

We implement the Metropolis algorithm as follows. A truncated normal distribution was found to have a shape close to p . We choose it as blanketing density q . The truncation is effected by discarding negative draws. We have not encountered such draws even in the smallest samples where the mean is still more than 6 standard deviations away from 0. A possible alternative to the normal blanket would be the lognormal distribution. We set the blanket mean equal to the mode of $p(\sigma | \cdot)$. The mode is found quickly in about 10 evaluations of the kernel. We then set the variance of q to best match the shape of q to p . For this, we compute and minimize the function $p \star / q$, where $p \star$ is the kernel of p , at the mode and 1 point on each side of the mode, where p is half the height of p at the mode. These two points are found in a few evaluations of the kernel. The minimization requires an additional 10 evaluations. This brings q as close as possible to p in the bulk of the distribution where about 70 % of the draws will be made. Possible values for c are the ratios $p \star / q$ at these three points. We choose c so as to slightly favor rejections over repeats. The top left plots of figure 1 show that the ratio $p \star / cq$ is close to 1 almost everywhere. The intuition of the ratio $p \star / cq$ is as follows. If a candidate draw is at the mode, ratio = 1, and the previous draw is at the upper dotted line, ratio = 1.1, then there is a $1/1.1$ chance that the previous draw will be repeated rather than the candidate draw chosen. Also, a draw at 0.27, ratio = 0.93, has a 7% chance of being rejected. The efficiency of the algorithm is verified wby keeping track of the actual rejections and repeats in the simulation.

A.3 Heteroskedastic errors

We model heteroskedastic errors as:

$$\log C_i = \beta_1 \log b(\sigma, x_{1i}) + \beta_2' x_{2i} + \sigma_{\eta, j(i)} \eta_i, i = 1, \dots, N \quad \eta_i \sim N(0, 1) \quad (9)$$

where $j(i) = 1, \dots, J$ indicates the levels of volatility. J is a smaller number than N . The conditional posteriors for this model follow readily. Conditional on $\sigma_{\eta, j}, \sigma$, one divides the observations in (9) by $\sigma_{\eta, j}$, which results in a regression with unit variance errors. $p(\beta | \sigma_{\eta, j(i)}, \sigma)$ follows. Conditional on σ, β , the densities of the various $\sigma_{\eta, j}$ are Inverse Gammas. Conditional on $\sigma_{\eta, j}, \beta$, the likelihood kernel of σ becomes

$$\exp \left\{ -\frac{\nu s^2(\sigma, \beta)_1}{2\sigma_{\eta, 1}^2} - \dots - \frac{\nu s^2(\sigma, \beta)_J}{2\sigma_{\eta, J}^2} \right\}, \quad (10)$$

where the sums of squares j is taken over the quotes which errors have standard deviation $\sigma_{\eta, j}$

B Analysis of Market Error

Consider

$$\begin{aligned} \log C_i &= \beta_1 \log m(\sigma, x_{1i}) + \beta_2' x_{2i} + a_i, \text{ where } a_i = \eta_i + s_i \epsilon_i \\ &= \beta' x_i + a_i \\ R_t &= \mu + \xi_t \\ \eta_i &\sim N(0, \sigma_\eta), \quad \epsilon_i \sim N(0, \sigma_\epsilon), \quad \xi_t \sim N(0, \sigma) \\ s_i &= \begin{cases} 0 & \text{with prob. } 1 - \pi \\ 1 & \text{with prob. } \pi \end{cases} \end{aligned}$$

The variance of a_i is $\sigma_i^2 = \sigma_\eta^2 + s_i \sigma_\epsilon^2 \equiv \sigma_\eta^2(1 + s_i \omega)$. Introduce the state vector $s = \{s_1, \dots, s_N\}$, a sequence of independent Bernoulli trials. Consider the prior distributions

$$\begin{aligned} \pi &\sim B(a, b) \\ (\beta | \omega, \sigma_\eta) &\sim N\left(\beta_0, \sigma_\eta^2 \left(1 + \frac{a}{a+b} \omega\right) V_0\right) \end{aligned}$$

$$\begin{aligned}\sigma_\eta &\sim \text{IG}(\nu_1, s_1^2) \\ \sigma &\sim \text{IG}(\nu_0, s_0^2).\end{aligned}\tag{11}$$

where IG and B are the Inverted Gamma and the Beta distributions. As usual the prior on σ can be derived from the history of the underlying return if desired. The priors can be made arbitrarily diffuse by setting ν_0 and ν_1 to 0, and the diagonal elements of V_0 to large values. Note that σ_ϵ is modelled through the specification of ω . The goal is to obtain the posterior joint and marginal distributions of $\beta, \sigma, \sigma_\eta, \pi, \omega$, and s . The first conditional posterior is that of $(\beta, \sigma, \sigma_\eta | \underline{y}_t, \omega, S)$:

1:

$$\begin{aligned}p(\sigma_\eta, \sigma, \beta | \underline{y}_t, \omega, s) &\propto \frac{\exp\left\{\frac{-\nu_0 s_0^2}{2\sigma^2}\right\}}{\sigma^{\nu_0+T_r}} \times \frac{\exp\left\{\frac{-\nu_1 s_1^2}{2\sigma_\eta^2}\right\}}{\sigma_\eta^{\nu_1+1+k+N_0} (\sigma_\eta \sqrt{1+\omega})^{N-N_0}} \\ &\times \exp\left\{-\frac{(\beta - \beta_0)' V_0^{-1} (\beta - \beta_0)}{2\sigma_\eta^2 \left(1 + \frac{a}{a+b}\omega\right)}\right\} \times \exp\left\{-\frac{(Y^* - X^* \beta)' (Y^* - X^* \beta)}{2\sigma_\eta^2}\right\}\end{aligned}$$

where N_0 is the number of observations for which s_i is zero. $Y^* = (\log C_1^*, \dots, \log C_N^*)'$, where $\log C_i^* = \log C_i / \sqrt{1 + \omega s_i}$. The same transformation is applied to the vector X , i.e. each element is divided by $\sqrt{1 + s_i \omega}$. After this transformation, a draw of this posterior is made as shown in section 3. Now consider ω introduced above. When s is known, the likelihood function of ω depends only on the $N_1 = N - N_0$ observations for which $s_i = 1$. Consider for $\bar{\omega} = 1 + \omega$, a truncated inverted gamma prior distribution $\text{IG}(\nu_2, s_2^2) I_{\bar{\omega} > 1}$. The posterior distribution of $\bar{\omega}$ conditional on the other parameters is

2:

$$\begin{aligned}p(\bar{\omega} | \underline{y}_t, \beta, \sigma_\eta, s) &\propto \frac{1}{\bar{\omega}^{1+\nu_2+N_1}} \times \exp\left\{-\frac{\sum_{i \in N_1} (Y_i - \beta' x_i)^2}{2\sigma_\eta^2 \bar{\omega}^2} + \nu_2 s_2^2\right\} I_{\bar{\omega} > 1} \\ &\sim \text{IG}\left(\nu_2 + N_1, \nu_\omega^2 s_\omega^2 = \nu_2 s_2^2 + \sum_{i \in N_1} \left(\frac{Y_i - \beta' x_i}{\sigma_\eta}\right)^2\right) I_{\bar{\omega} > 1}\end{aligned}$$

where $I_{\bar{\omega} > 1}$ is the indicator function for $\bar{\omega} > 1$. A draw of ω is obtained directly from a draw of $\bar{\omega}$ since $\bar{\omega} = 1 + \omega$. We now need the conditionals $p(s_i | \underline{y}_t, s_{-i}, \cdot)$ where "." stands for all the other parameters, and s_{-i} refers to the state vector without s_i . Following McCulloch and Tsay (1993), they are written as

3:

$$\begin{aligned} p(s_i = 1 \mid y, s_{-i}, \cdot) &= \frac{\pi p(y_t \mid s_i = 1, \cdot)}{\pi p(y_t \mid s_i = 1, \cdot) + (1 - \pi) p(y_t \mid s_i = 0, \cdot)} \\ &= \frac{1}{1 + \frac{1-\pi}{\pi} \times \frac{p(y_t \mid s_i = 0, \cdot)}{p(y_t \mid s_i = 1, \cdot)}} \end{aligned}$$

For the set up considered here the denominator term is simply:

$$\frac{p(y_t \mid s_i = 0, \cdot)}{p(y_t \mid s_i = 1, \cdot)} = \sqrt{1 + \omega} \exp - \frac{(\log C_i - \beta' x_i)^2}{2\sigma_\eta^2} \times \frac{\omega}{1 + \omega}$$

We now need the last conditional posterior of π . It depends exclusively on N_1 the number of s_i 's equal to 1:

4:

$$p(\pi \mid s, \cdot) \sim B(a + N_1, b + N - N_1)$$