

ISSN 1936-5098

CAE Working Paper #09-12

**Correlated Defaults, Temporal Correlation, Expert Information  
and Predictability of Default Rates**

by

Nicholas M. Kiefer

June 2009

# Correlated Defaults, Temporal Correlation, Expert Information and Predictability of Default Rates

Nicholas M. Kiefer<sup>1</sup>  
Cornell University

June 25, 2009

<sup>1</sup>Cornell University, Departments of Economics and Statistical Science, 490 Uris Hall, Ithaca, NY 14853-7601, US. email:nicholas.kiefer@cornell.edu; US Department of the Treasury, Office of the Comptroller of the Currency, Risk Analysis Division, and CRE-ATES, funded by the Danish Science Foundation, University of Aarhus, Denmark. *Disclaimer: The statements made and views expressed herein are solely those of the author, and do not necessarily represent official policies, statements or views of the Office of the Comptroller of the Currency or its staff.*

## Abstract

Dependence among defaults both across assets and over time has proven to be an important characteristic of financial risk. A Bayesian approach to default rate estimation is proposed and illustrated using a prior distributions assessed from an experienced industry expert. Two extensions of the binomial model, most common in applications, are proposed. The first allows correlated defaults yet remain consistent with Basel II's asymptotic single-factor model. The second adds temporal correlation in default rates through autocorrelation in the systemic factor. Implications for the predictability of default rates are considered. The single-factor model generates more forecast uncertainty than does the parameter uncertainty. A robustness exercise, weakening the prior on the asset correlation, illustrates that the correlation indicated by the data is much smaller than that specified in the Basel II regulations. The application shows that econometric methods can be useful even when data information is sparse.

**Keywords:** Bayesian inference, Basel II, risk management, prior elicitation, maximum entropy, time series

# 1 Introduction

Estimation of long-run default rates for groups of homogeneous assets is essential for determining adequate capital requirements. Forecasting default rates, which may vary over time, is essential to prudent risk management. Recent events have illustrated the importance of systemic risk - namely correlated defaults. The Basel II (B2) framework (Basel Committee on Banking Supervision (2006)) for calculating minimum capital requirements provides for banks to use models to assess credit (and other) risks. In response to the credit crisis, the Basel Committee has stressed in a document for comment the continuing importance of quantitative risk management, see Basel Committee on Banking Supervision (2009). Underlying the B2 model for required capital is a single-factor model in which asset values for homogeneous assets are correlated due to a systematic shock affecting all assets in the portfolio bucket. We consider a series of models. The first is the simple binomial model for defaults in which the underlying asset value processes are uncorrelated across assets and time. This is by far the most commonly used model and invariably provides a starting point for more sophisticated analyses. Next, we follow B2 and add asset correlation over time through a systematic shock affecting the values of all assets in the portfolio. This is typically the next step a bank using Internal Ratings-Based (IRB) methods would take. Finally, we add autocorrelation in the systematic factor. This does not seem to be contemplated in the B2 model, but is certainly plausible, is consistent with models of asset values in other settings and can be modeled. Midportfolio asset groups, mostly commercial loans, typically to unrated companies, for which default is still rare typically have some default experience. These groups can have sufficient data to support models allowing simple forms of default dependence. Expert information on loan credit risk is important and a Bayesian approach is adopted throughout. Thus, uncertainty about the default probability should be modeled the same way as uncertainty about defaults – represented in a probability distribution. A future default either occurs or doesn't, given the definition. In fact both the definition and consequences of default can vary, particularly across countries - see Franks and Davydenko (2008). Since we do not know in advance whether default occurs or not, we model this uncertain event with a probability distribution. Similarly, the default probability is unknown. But there is information available about the default rate in addition to the data information. The simple fact that loans are made shows that some risk assessment

is occurring. This information should be organized and incorporated in the analysis in a sensible way, specifically represented in a probability distribution. This information can then be combined coherently with data information as represented in the relevant likelihood function. This combination of information is easy to do once the information is represented in probability distributions. The final distribution represents both data and expert information about the default probability.

Much effort in recent econometric research is devoted to methods for extracting data information with minimal assumptions. Examples include semiparametric and nonparametric models as well as GMM models avoiding to some extent distributional assumptions. These approaches are appealing when there is a great deal of informative data, as illustrated by the fact that most results available and sought are asymptotic. However there is a large class of important practical problems for which there are some, but not much data and there are also experts whose experience and knowledge is relevant. Despite the lack of conclusive data evidence, decisions must be made. In many applied settings professional judgment is expected and discussed at the same level as the statistical model (which is a result of judgment) or the relevance of the data (more judgment). This inference situation is unfamiliar to economists due perhaps to the ambiguity surrounding the way in which expert and data information are combined. Much of this uncomfortable ambiguity can be eliminated by adopting a formal Bayesian approach. In this approach the econometrician is faced with two modeling tasks. The first is the familiar task of specifying the likelihood function. The second is specifying the expert information, usually consisting of a limited number of assessments, in terms of a statistical distribution. As a practical matter a rather small set of assessments is used as the basis for fitting a probability distribution matching these assessments as well as possible. Just as the data distribution is specified in terms of a small number of parameters which imply a complete distribution, so the prior is based on a small number of assessments.

Section 2 discusses the statistical model for defaults. Section 3 concerns the prior distribution for defaults for a mid-portfolio "bucket" of loans elicited from an experienced industry expert. A smoothed maximum entropy representation is used to provide a statistical description of the properties elicited from the expert. Section 4 turns to inference and the posterior distributions are obtained. Section 5 concerns implications of both the economic model and parameter uncertainty for the predictability of default rates. The temporal variation in forecast default rates

implied by the single-factor model is substantially greater than the variation implied by parameter uncertainty. Section 6 considers an alternative, "looser" prior for the asset correlation parameter - a parameter whose value is specified in B2. In fact, the data seem to imply much less correlation than suggested by the B2 regulations. Section 7 concludes.

## 2 Statistical Models for Defaults

The simplest and most common probability model for defaults of assets in a homogeneous segment of a portfolio is the Binomial, in which the defaults are assumed independent across assets and over time, and occur with common probability  $\theta$ . The Basel requirements demand an annual default probability, estimated over a sample long enough to cover a full cycle of economic conditions. Thus the probability should be marginal with respect to external conditions. Perhaps this marginalization can be achieved within the binomial specification by averaging over the sample period, thus many discussions of the inference issue have focussed on the binomial model and the associated frequency estimator. Suppose the value of the  $i$ th asset in time  $t$  is

$$v_{it} = \epsilon_{it}$$

where  $\epsilon_{it}$  is the time and asset specific shock and default occurs if  $v_{it} < T^*$ , a default threshold value. A mean of zero is attainable through translation without loss of generality. We assume the shock is standard normal. Let  $d_i$  indicate whether the  $i$ th observation was a default ( $d_i = 1$ ) or not ( $d_i = 0$ ). The distribution of  $d_i$  is  $p(d_i|\theta) = \theta^{d_i}(1 - \theta)^{1-d_i}$ , where  $\theta = \Phi(T^*)$ , our Binomial parameter. Let  $D = \{d_i, i = 1, \dots, n\}$  denote the whole data set and  $r = r(D) = \sum_i d_i$  the count of defaults. Then the joint distribution of the data is

$$\begin{aligned} p(D|\theta) &= \prod_{i=1}^n \theta^{d_i}(1 - \theta)^{1-d_i} \\ &= \theta^r(1 - \theta)^{n-r} \end{aligned} \tag{1}$$

Since this distribution depends on the data  $D$  only through  $r$  ( $n$  is regarded as fixed), the sufficiency principle implies that we can concentrate attention on the distribution of  $r$

$$p(r|\theta) = \binom{n}{r} \theta^r (1 - \theta)^{n-r} \tag{2}$$

a Binomial( $n, \theta$ ) distribution.

The Basel II guidance suggests there may be heterogeneity due to systematic temporal changes in asset characteristics or to changing macroeconomic conditions. There is some evidence from other markets that default probabilities vary over the cycle. See Nickell, Perraudin, and Varotto (2000) and Das, Duffie, Kapadia, and Saita (2007). The B2 capital requirements are based on a one-factor model due to Gordy (2003) that accommodates systematic temporal variation in asset values and hence in default probabilities. This model can be used as the basis of a model that allows temporal variation in the default probabilities, and hence correlated defaults within years. The value of the  $i$ th asset in time  $t$  is modeled as

$$v_{it} = \rho^{1/2}x_t + (1 - \rho)^{1/2}\epsilon_{it} \quad (3)$$

where  $\epsilon_{it}$  is the time and asset specific shock (as above) and  $x_t$  is a common time shock, inducing correlation  $\rho$  across asset values within a period. The random variables are assumed to be standard normal and independent. The overall or marginal default rate we are interested in is  $\theta = \Phi(T^*)$ . However, in each period the default rate  $\theta_t$  depends on the realization of the systematic factor  $x_t$ . The model implies a distribution for  $\theta_t$ . Specifically, the distribution of  $v_{it}$  conditional on  $x_t$  is  $N(\rho^{1/2}x_t, 1 - \rho)$ . Hence the period  $t$  default probability is

$$\theta_t = \Phi[(T^* - \rho^{1/2}x_t)/(1 - \rho)^{1/2}] \quad (4)$$

Thus for  $\rho \neq 0$  there is random variation in the default probability over time. The distribution is given by

$$\begin{aligned} \Pr(\theta_t \leq A) &= \Pr(\Phi[(T^* - \rho^{1/2}x_t)/(1 - \rho)^{1/2}] \leq A) \\ &= \Phi[((1 - \rho)^{1/2}\Phi^{-1}[A] - \Phi^{-1}[\theta])/ \rho^{1/2}] \end{aligned} \quad (5)$$

using the standard normal distribution of  $x_t$  and  $\theta = \Phi(d)$ . Differentiating gives the density  $p(\theta_t|\theta, \rho)$ . This is the Vasicek distribution. The parameters are  $\theta$ , the marginal or mean default probability and the asset correlation  $\rho$ . The conditional distribution of the number of defaults in each period is (from 2)

$$p(r_t|\theta_t) = \binom{n_t}{r_t} \theta_t^{r_t} (1 - \theta_t)^{n_t - r_t} \quad (6)$$

from which we obtain the distribution conditional on the underlying parameters

$$p(r_t|\theta, \rho) = \int p(r_t|\theta_t)p(\theta_t|\theta, \rho)d\theta_t$$

The distribution for  $R = (r_1, \dots, r_T)$  is

$$p(R|\theta, \rho) = \prod_{t=1}^T p(r_t|\theta, \rho) \tag{7}$$

where we condition on  $(n_1, \dots, n_T)$ . Regarded as a function of  $(\theta, \rho)$  for fixed  $R$ , 7 is the likelihood function.

This formulation allows clumping of defaults within time periods, but not correlation across time periods. This is the next natural extension. Specifically, let  $x_t$  follow an  $AR(1)$  process

$$x_t = \tau x_{t-1} + \eta_t$$

with  $\eta_t$  iid standard normal. Now the formula for  $\theta_t$  (4) still holds but the likelihood calculation is different and cannot be broken up into the period-by-period calculation. Write using (6)

$$p(R|\theta_1, \dots, \theta_T) = \prod_{t=1}^T p(r_t|\theta_t(x_t, \theta, \rho))$$

emphasizing the functional dependence of  $\theta_t$  on  $x_t$  as well as  $\theta$  and  $\rho$ . Now we can calculate the desired unconditional distribution

$$p(R|\theta, \rho, \tau) = \int \dots \int \prod_{t=1}^T p(r_t|\theta_t(x_t, \theta, \rho))p(x_1, \dots, x_T|\tau)dx_1 \dots dx_T \tag{8}$$

where  $p(x_1, \dots, x_T|\tau)$  is the density of a zero-mean random variable following an  $AR(1)$  process with parameter  $\tau$ . Regarded as a function of  $(\theta, \rho, \tau)$  for fixed  $R$ , 8 is the likelihood function.

This one-factor model for asset value and hence default correlation is quite simple but it does have empirical support. Recall that the techniques here are applied to "buckets" of homogeneous assets. In a study of a large sample of nonfinancial and nonutility corporate bonds, Das, Duffie, Kapadia, and Saita (2007) examine the suitability of a model based on correlation determined by observable factors



(a T-bill rate and lagged S&P returns). Since their sample is not a bucket of homogeneous assets they also control for asset characteristics. They find that most of the default correlation is explained by the observable factors but there is a small, significant remaining correlation which is consistent with a common unobserved factor. We tuck all of the correlation into the unobserved factor, noting that it may be correlated with observed factors and noting that our assets are considerably more homogeneous.

### 3 Prior Distribution

I have asked an expert to consider a portfolio bucket consisting of loans that might be in the middle of a bank's portfolio. These are typically commercial loans to unrated companies. If rated, these might be about S&P Baa or Moody's BBB. The method included a specification of the problem and some specific questions followed by a discussion. The elicitation method included a specification of the problem and some specific questions over e-mail followed by a discussion. Elicitation of prior distributions is an area that has attracted attention. General discussions of the elicitation of prior distributions are given by Garthwaite, Kadane, and O'Hagan (2005), O'Hagan, Buck, Daneshkhah, Eiser, Garthwaite, Jenkinson, Oakley, and Rakow (2006) and Kadane and Wolfson (1998). I merely sketch the elicitation and representation of the prior here: details are given in a previous application Kiefer (2008). The elicitation took place in 2006. The expert found it easier to think in terms of the probabilities directly than in terms of defaults in a hypothetical sample. This is not uncommon in this technical area, as practitioners are accustomed to working with probabilities. The mean value was 0.01. The minimum value for the default probability was 0.0001 (one basis point). The expert reported that a value above 0.035 would occur with probability less than 10%, and an absolute upper bound was 0.3. The upper bound was discussed: the expert thought probabilities in the upper tail of his distribution were extremely unlikely, but he did not want to rule out the possibility that the rates were much higher than anticipated (prudence?). Quartiles. were assessed by asking the expert to consider the value at which larger or smaller values would be equiprobable given the value was less than the median, then given the value was more than the median. The median value was 0.01. The former was 0.0075. The latter, the .75 quartile, was assessed at .0125. The expert, who has long experience with this category of assets, seemed to be thinking of

a distribution with a long and thin upper tail but otherwise symmetric. After reviewing the implications, the expert added a .99 quantile at 0.02, splitting up the long upper tail. This information was used to fit a truncated Beta distribution in Kiefer (2009). Here, we instead take a maximum entropy (ME) approach.

The ME approach provides a method to specify the distribution that meets the expert specifications and otherwise imposes as little additional information as possible. Thus, we maximize the entropy in the distribution subject to the constraints indexed by  $k$  given by the assessments. This gives a piecewise uniform distribution for  $\theta$ . It can be argued that the discontinuities in the ME densities are unlikely to reflect characteristics of expert information and indeed this was the view of the expert. Smoothing was accomplished using the Epanechnikov kernel with several bandwidths  $h$  chosen to offer the expert choices on smoothing level (including no smoothing). Specifically, with  $p_S(\theta)$  the smoothed distribution with bandwidth  $h$  we have

$$p_S(\theta) = \int_{-1}^1 K(u)p_{ME}(\theta + u/h)du \quad (9)$$

with  $K(u) = 3(1 - u^2)/4$  for  $-1 < u < 1$ . Since the density is defined on bounded support there is an endpoint or boundary "problem" in calculating the kernel-smoothed density estimator. Specifically,  $p_S(\theta)$  as defined in 9 has larger support than  $p_{ME}(\theta)$ , moving both endpoints out by a distance  $1/h$ . We adjust for this using reflection,  $p_{SM}(\theta) = p_S(\theta) + p_S(a - \theta)$  for  $a \leq \theta < a + 1/h$ ,  $p_{SM}(\theta) = p_S(\theta)$  for  $a + 1/h \leq \theta < b - 1/h$ , and  $p_{SM}(\theta) = p_S(\theta) + p_S(2b - \theta)$  for  $b - 1/h \leq \theta \leq b$ . The resulting smoothed density has support on  $[a, b]$  and integrates to 1. See Schuster (1985). The prior distribution for  $\theta$  is shown in Figure 1.

To specify the prior on the asset correlation  $\rho$ , note that B2 provides guidance. For this portfolio bucket, B2 recommends a value of 0.20. I did not assess further details from an expert on this parameter. There appears to be little experience with correlation, relative to expert information available on default rates. There is agreement that the correlation is positive (as it has to be asymptotically if there are many assets). Consequently, I choose a Beta prior with mean equal to 0.20 for  $\rho$ . Since the B2 procedure is to fix  $\rho$  at that value, any weakening of this constraint is a generalization of the model we first choose a rather tight prior around the mean, namely a Beta(12.6, 50.4) distribution. As a check (a robustness check? a validation?) I also consider in Section 6 a looser prior, Beta(1.5, 6.0). These

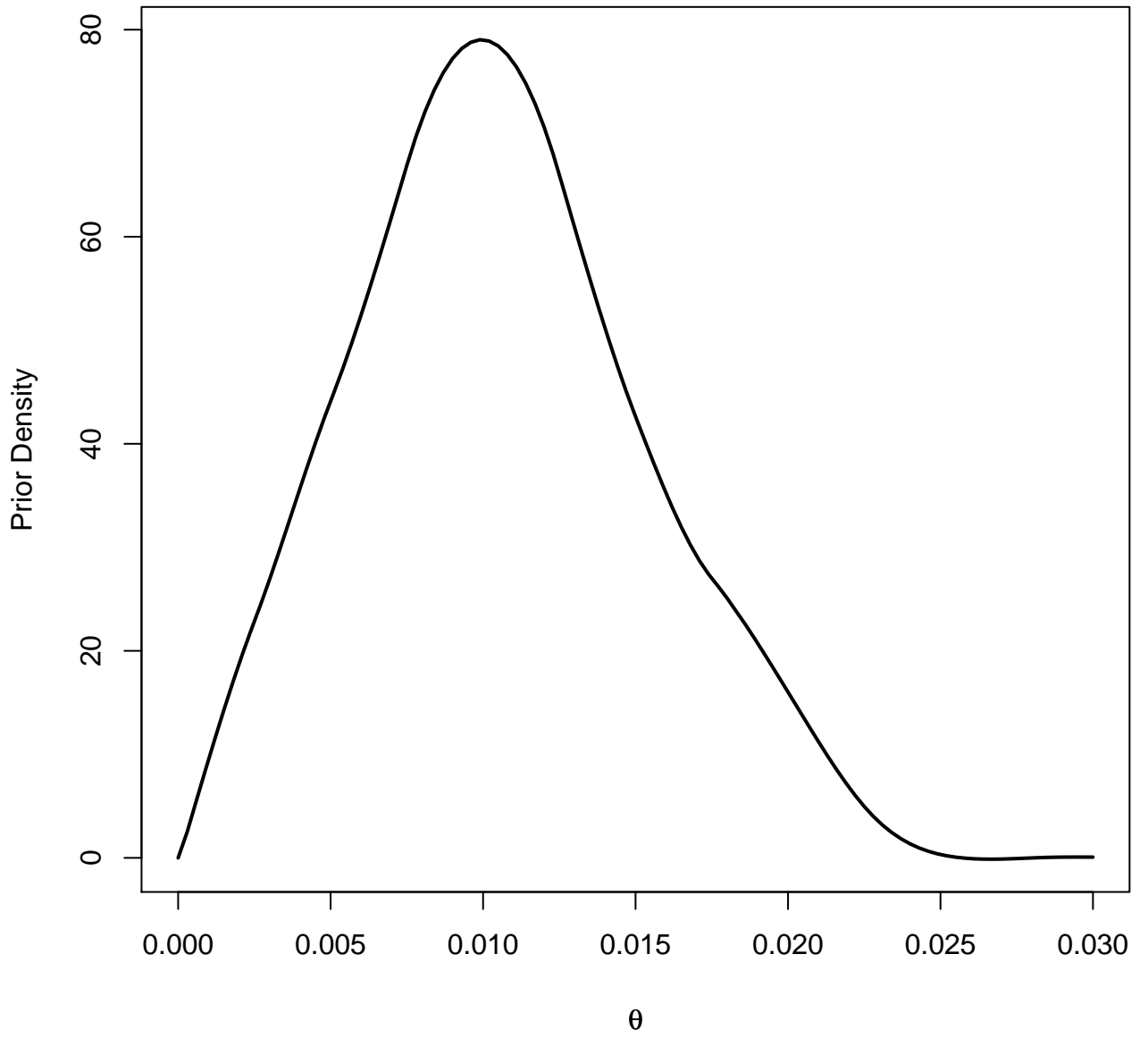


Figure 1: Prior on the long-run default probability  $\theta$

priors are illustrated in Figure 2. For the autocorrelation parameter  $\tau$ , we take the uniform on  $(-1,1)$ , reflecting the zero mean corresponding to the B2 model (without autocorrelation). Thus, the prior specifications on the parameters for which we have no expert information beyond that given in the B2 guidelines reflect the guidelines as means and little else. The joint prior is obtained as the product, which is the maximum-entropy combination of the given marginals. Here, it does not seem to make sense to impose correlation structure in the absence of expert information.

## 4 Inference

Writing the likelihood function generically as  $p(R|\phi)$  with  $\phi \in \{\theta, (\theta, \rho), (\theta, \rho, \tau)\}$  depending on whether we are referring to the likelihood function (2), (7), or (8), and the corresponding prior  $p(\phi)$ , inference is a straightforward application of Bayes rule. The joint distribution of the data  $R$  and the parameter  $\phi$  is

$$p(R, \phi) = p(R|\phi)p(\phi)$$

from which we obtain the marginal (predictive) distribution of  $R$ ,

$$p(R) = \int p(R, \phi) d\phi \tag{10}$$

and divide to obtain the conditional (posterior) distribution of the parameter  $\phi$  :

$$p(\phi|R) = p(R|\phi)p(\phi)/p(R) \tag{11}$$

Given the distribution  $p(\phi|R)$ , we might ask for a summary statistic, a suitable estimator for plugging into the required capital formulas as envisioned by Basel Committee on Banking Supervision (2006). A natural value to use is the posterior expectation,  $\bar{\phi} = E(\phi|R)$ . The expectation is an optimal estimator under quadratic loss and is asymptotically an optimal estimator under bowl-shaped loss functions. When applied to calculating minimum required capital, interest centers on the component  $\bar{\theta} = E(\theta|R)$ . For default prediction, taken up in Section 5, interest centers on  $\overline{\theta_{T+1}}$ , which depends on all components of  $\phi$ .

We construct a bucket of mid-portfolio corporate bonds from S&P-rated firms in the KMV North American Non-Financial Dataset. Default rates were computed for

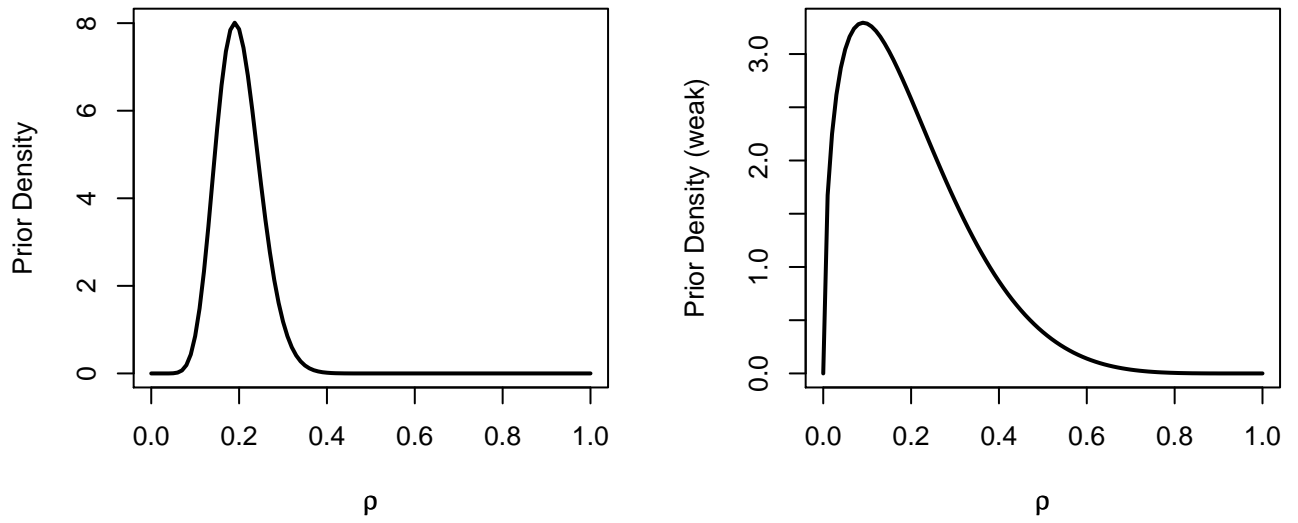


Figure 2: Prior and weakened prior on the asset correlation  $\rho$

cohorts of firms starting in September 1993 and running through September 2004. In total there are 2197 asset/years of data and 20 defaults, for an overall empirical rate of 0.00913. The analysis of the binomial model is straightforward using direct calculations involving numerical integration to calculate the predictive distribution and various moments (recall we are not in a conjugate-updating framework due to the flexible form of the prior representation). Model 2, with asset-value correlation but no temporal correlation, can be analyzed with direct numerical integration, though this is time consuming. Model 3 is more difficult. This seems an ideal setting to use Markov Chain Monte Carlo. Specifically, I use a random-walk Metropolis algorithm with a normal proposal distribution. In each case, 10000 samples from the joint posterior distribution were taken after a 5000-sample burnin. Scaling of the proposal distribution allowed an acceptance rate between 18 and 24 percent. This procedure was used for Model 2 and for Model 3. Calculation of posterior distributions of the parameters and the functions of parameters considered in Section 5 are based on these samples. The software was the mcmc package (Geyer (2009)) used in R (R Development Core Team (2009)).

The posterior distribution for the binomial model is shown in Figure 3.

This density has  $E(\theta|R) = 0.0096$  and  $\sigma_\theta = 0.0012$ .

Model 2 has asset value correlation within periods, allowing for heterogeneity in the default rate over time (but not correlated over time) and clumping of defaults. The marginal posterior distributions are shown in Figure 4.

The densities have a single maximum and the default probability is well determined, though not as precisely as in the single-parameter model. The posterior summary statistics are:  $E(\theta|R) = 0.0101$  and  $\sigma_\theta = 0.0019$ ,  $E(\rho|R) = 0.0960$  and  $\sigma_\rho = 0.0260$ . The correlation parameter is not as well-determined, even with the rather sharp prior. The data seem to be leaning toward rather lower correlation than the a priori expectation generated by the B2 prescription. Of course, the data evidence is rather weak as reflected in the posterior standard error. The posterior correlation is  $corr(\theta, \rho) = 0.078$ .

Allowing the systematic shock to be autocorrelated leads to Model 3. The marginal posterior densities are shown in Figure 5.

Again the default probability is well determined, though the other parameters are not. The posterior summary statistics are:  $E(\theta|R) = 0.0102$  and  $\sigma_\theta = 0.0019$ ,  $E(\rho|R) = 0.0914$  and  $\sigma_\rho = 0.0259$ , and  $E(\tau|R) = 0.7420$  and  $\sigma_\tau = 0.0880$ . Adding autocorrelation in the systematic factor very slightly reduces the within-period as-

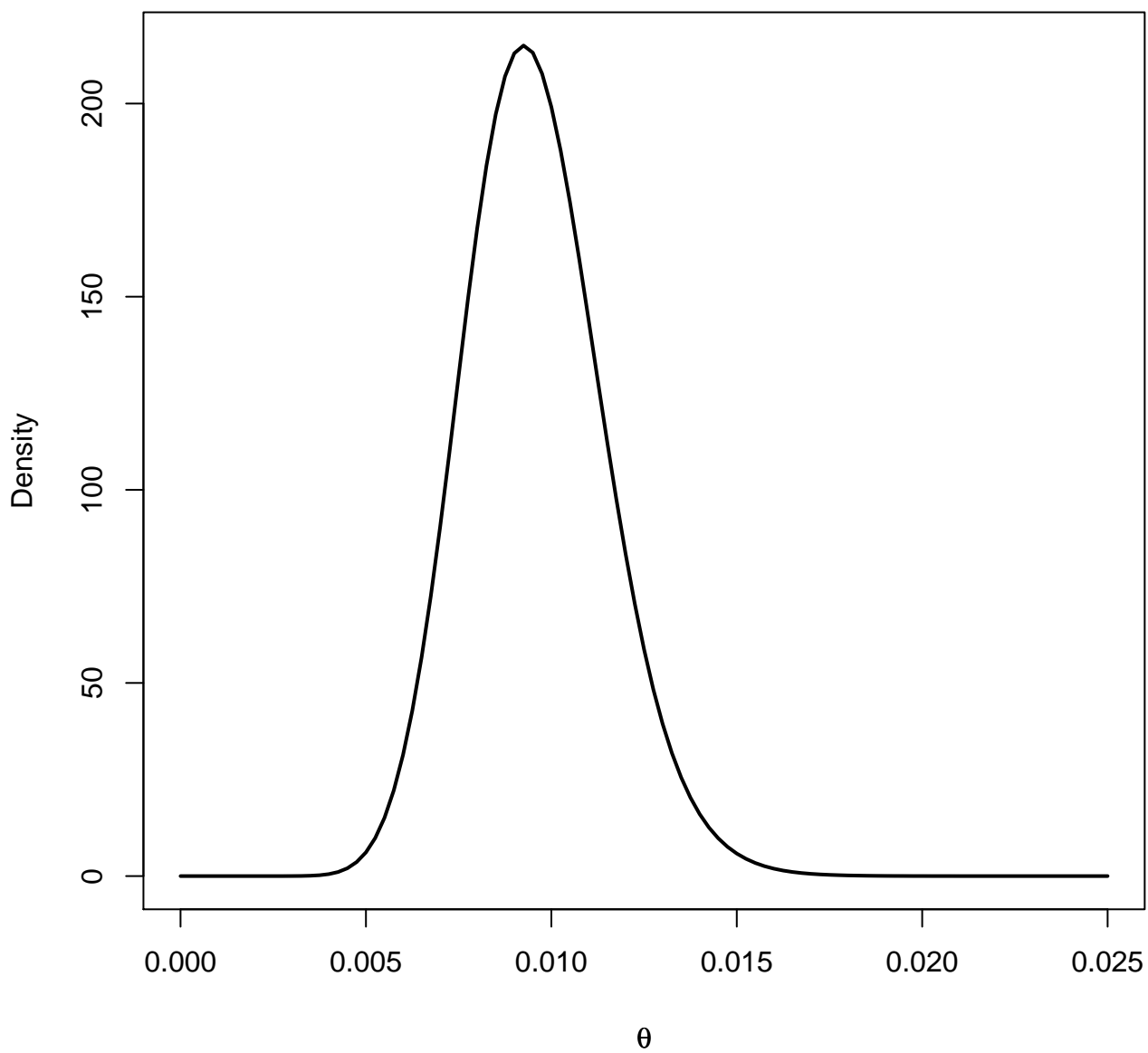


Figure 3: Posterior density for  $\theta$ , in the Binomial model (Model 1).

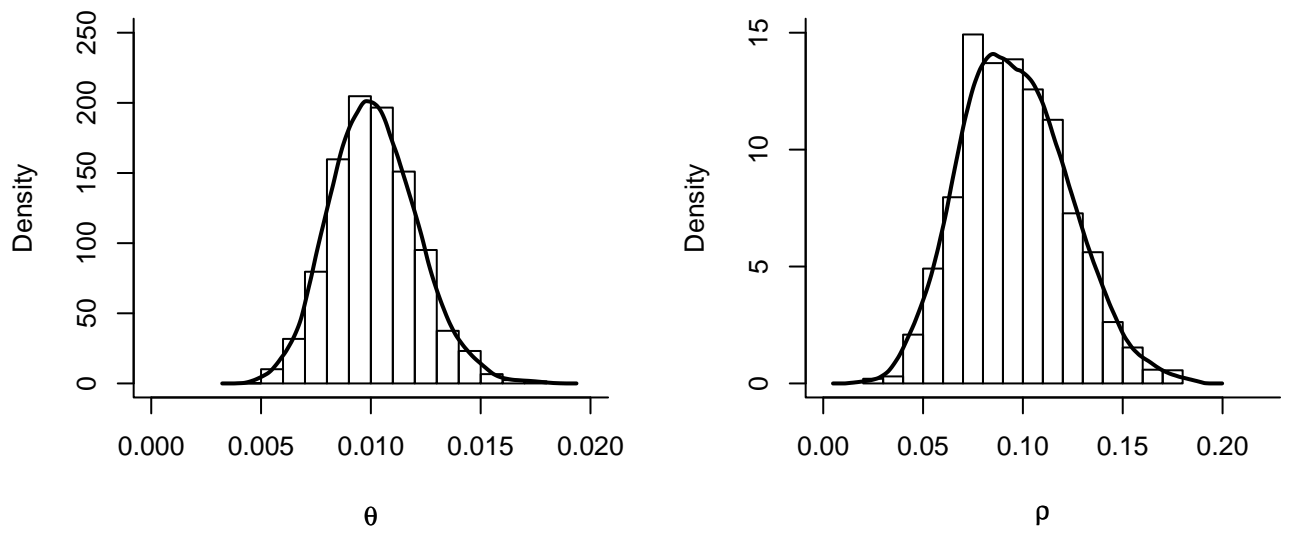


Figure 4: Posterior densities for  $\theta$  and  $\rho$  in Model 2.



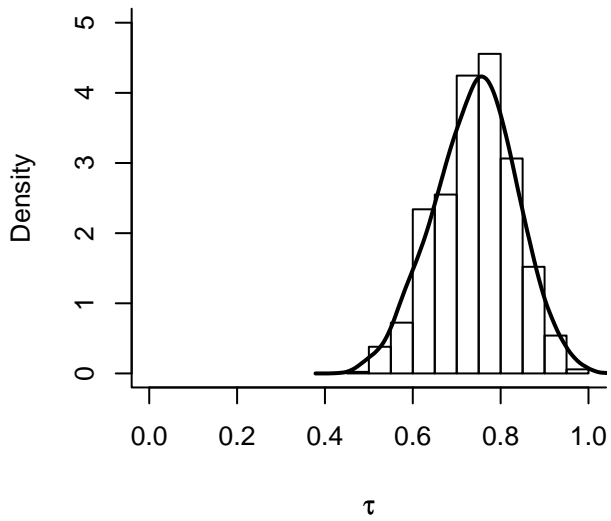
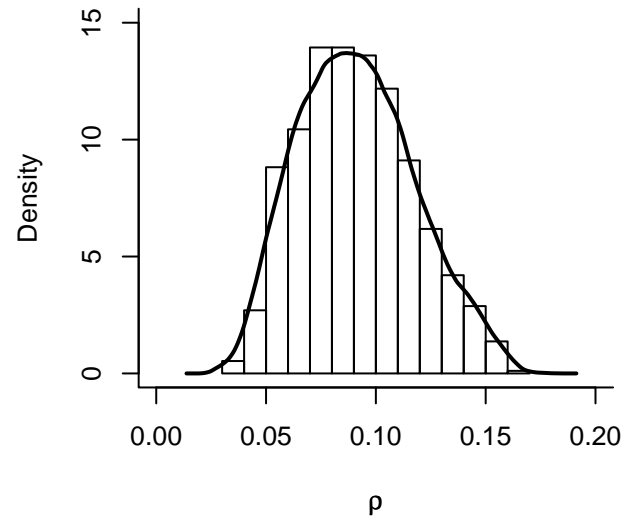
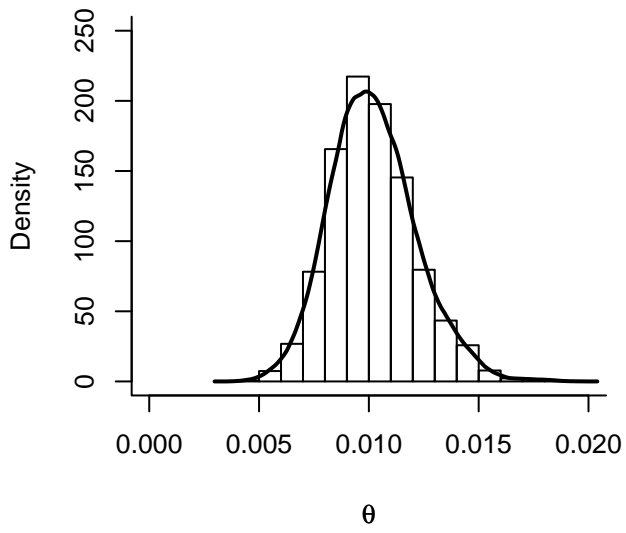


Figure 5: Posterior densities for  $\theta$ ,  $\rho$  and  $\tau$  in Model 3.

set correlation parameter. The posterior correlations, induced by the likelihood function since the parameters were specified a priori independent, are  $corr(\theta, \rho) = 0.113$ ,  $corr(\theta, \tau) = 0.010$ , and  $corr(\rho, \tau) = 0.058$ .

## 5 Predictability of Default Rates

Estimation of the long run default rate  $\theta$  is crucial for calculating minimum capital requirements under B2. But banks and other lenders are also interested in predicting defaults in the next year and outer years. The model underlying the B2 capital formula has implications for the predictability of future default rates. In fact, banks regularly predict default rates, and report to the supervisory authorities on the accuracy of their predictions. In a typical validation, a model is used to forecast  $\theta_{T+1}$ , that forecast,  $\theta_{T+1}^F$ , is then regarded as a fixed number, and the sampling distribution of the frequency estimator of the period  $T + 1$  default rate  $\widehat{\theta_{T+1}}$  is used to test  $H_0 : \theta_{T+1} = \theta_{T+1}^F$ . It is widely thought that this procedure leads to too many rejections, not in the sense that the sampling distribution understates the nominal size, but in the sense that departures small enough to have no business implications are sometimes statistically significant. Glennon and Kiefer (2008), argue that the appropriate null hypothesis is not the point null but an interval null, where the interval is determined by business considerations. Hanson and Schuermann (2006) study confidence intervals for estimated default probabilities from a sampling theory point of view and propose a bootstrap method. Here I investigate two different sources of uncertainty about  $\theta_{T+1}$ , that generated by the economic model given parameter values, and that generated by parameter uncertainty, as reflected in the posterior distribution for the parameters.

### 5.1 Uncertainty from the economic model

In this section we abstract from parameter uncertainty and suppose that the parameter  $\eta$  is known. In Model 1, the Binomial, the default rate is constant over time so the  $T + 1$  forecast default rate  $\theta_{T+1}^F$  is simply  $\theta$ , the known long-run default rate. The other models are more interesting.

In Model 2, the distribution of the default rate  $\theta_{T+1}$  conditional on  $\eta = (\theta, \tau)$  is from 5

$$\Pr(\theta_{T+1} < A) = \Phi[((1 - \rho)^{1/2}\Phi^{-1}[A] - \Phi^{-1}[\theta])/\rho^{1/2}]$$

and the density  $p(\theta_{T+1}|\theta, \rho)$  is obtained by differentiating. This density evaluated at the posterior mean values for  $\eta$  is graphed in Figure 6.

This distribution has  $E(\theta_{T+1}|\theta, \rho, R) = 0.010$  and  $\sigma_{\theta_{T+1}} = 0.009$ . Thus the Vasicek distribution generated by the B2 single-factor model implies considerable uncertainty about default forecasts even conditionally on parameters. In fact, it is widely thought that this distribution substantially understates the predictability of default rates. Tests based on whether  $\widehat{\theta_{T+1}}$ , (often regarded as a fixed number) are within a confidence interval generated by the Vasicek distribution, a procedure which has been suggested but is not widely used, are regarded as unsatisfactory because they "never" reject, even for quite bad models from a business point of view.

In Model 3 the realized period  $T$  default rate is useful in predicting  $\theta_{T+1}$  because of the dynamics of the systemic factor  $x$ . From 4 we can write

$$\begin{aligned} x_t &= (T^* - (1 - \rho)^{1/2}\Phi^{-1}(\theta_t))\rho^{-1/2} \\ &= \tau(T^* - (1 - \rho)^{1/2}\Phi^{-1}(\theta_{t-1}))\rho^{-1/2} + \epsilon_t \end{aligned}$$

Hence

$$\begin{aligned} \Pr(\theta_t \leq A|\theta_{t-1}) & & (12) \\ &= \Pr(\Phi[(T^* - \rho^{1/2}(\tau(T^* - (1 - \rho)^{1/2}\Phi^{-1}(\theta_{t-1}))\rho^{-1/2} + \epsilon_t))/(1 - \rho)^{1/2}] \leq 1) \\ &= \Pr(\epsilon_t < \rho^{-1/2}(\Phi^{-1}(A) - T^* + \tau(T^* - (1 - \rho)^{1/2}\Phi^{-1}(\theta_{t-1}))) \end{aligned}$$

using the fact that  $\epsilon_t$  is symmetric around zero. This is just a standard normal integral and the density  $p(\theta_{T+1}|\theta_T, \theta, \rho, \tau)$  is obtained by differentiation. The density at the posterior mean values for  $\eta = (\theta, \rho, \tau)$  is given in Figure 7 for  $\theta_{t-1} = 0.004$  and for  $\theta_{t-1} = 0.015$ . Conditioning on the recent default rate clearly has a substantial effect. Summary statistics are  $E(\theta_{T+1}|\theta_T = 0.004, \theta, \rho, R) = 0.007$  with  $\sigma_{\theta_{T+1}} = 0.006$ , and  $E(\theta_{T+1}|\theta_T = 0.015, \theta, \rho, R) = 0.016$  with  $\sigma_{\theta_{T+1}} = 0.013$ .

HPD regions for  $\theta_{T+1}$  for all models are given in Table 1.

## 5.2 Uncertainty from the model and the parameters

Conditioning on parameter values is useful in understanding the relative contributions of parameter uncertainty and the uncertainty generated by the model. In an

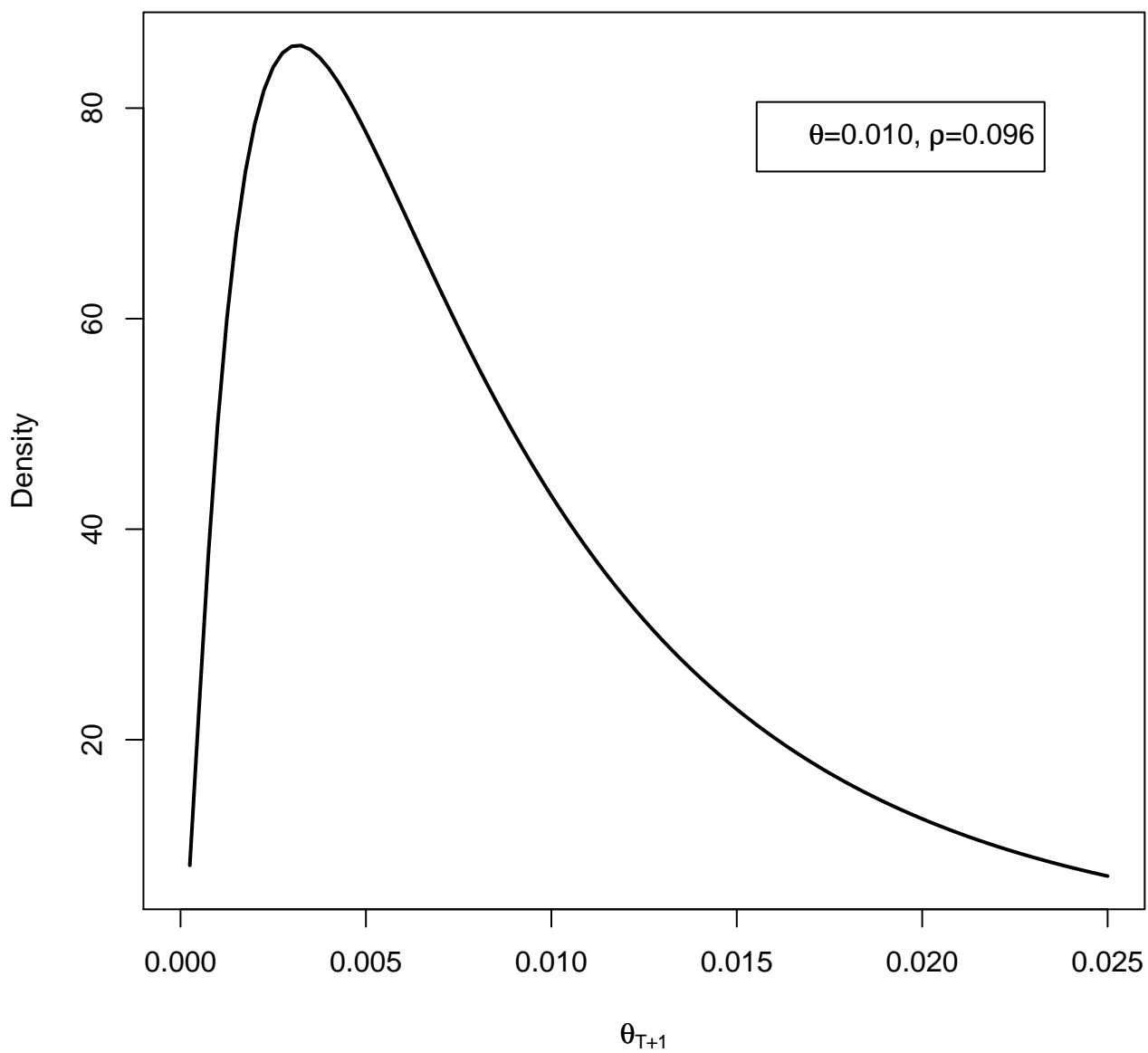


Figure 6: Predictive density  $p(\theta_t|\theta, \rho, R)$  with  $(\theta, \rho)$  fixed at their mean values.

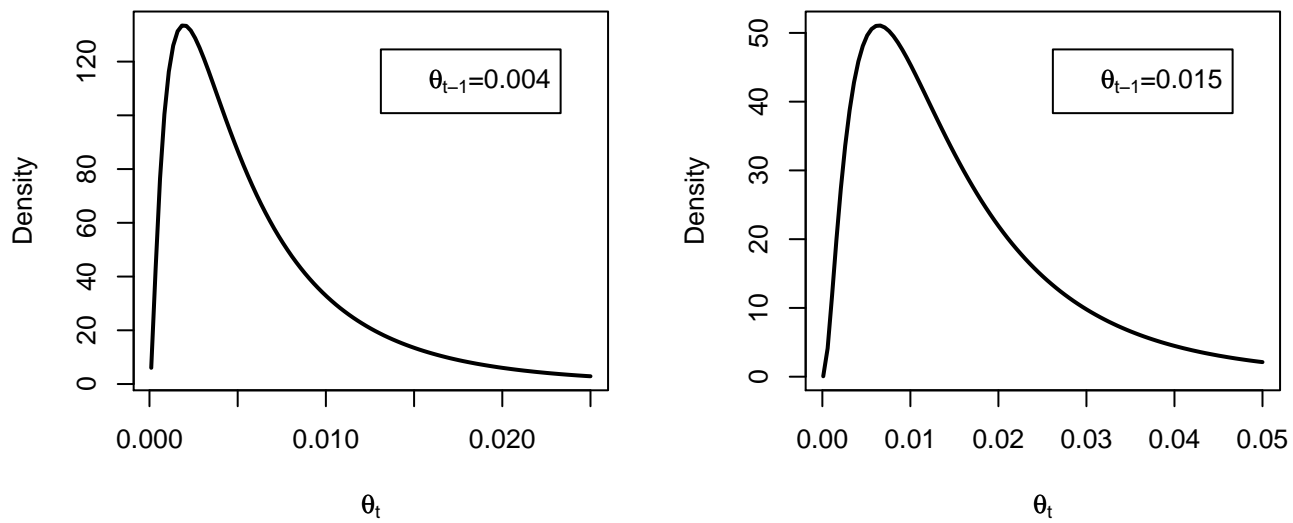


Figure 7: Predictive densities  $p(\theta_t|\theta_{t-1},\theta, \rho, \tau, R)$  with  $(\theta, \rho, \tau)$  fixed at their mean values.

application however it is clearly preferable to marginalize with respect to parameters rather than conditioning on particular values (here the expected values, but it could have been medians or modes). In this case, even Model 1 implies a predictive distribution, due to the uncertainty about the value of  $\theta$ . Of course, the predictive distribution for  $\theta_{T+1}$  from Model 1 is simply the posterior distribution of  $\theta$  given in Figure 3. Recall that this has mean 0.0096 and standard deviation 0.0012, so the uncertainty in the forecast due to parameter uncertainty in Model 1 is much less than the uncertainty generated by the economic model in Model 2, even with fixed parameters.

Turning to Model 2, the relevant density is

$$p(\theta_{T+1}|R) = \iint p(\theta_{T+1}|\theta, \rho)p(\theta, \rho|R)d\theta d\rho$$

where the definite integrals are over the supports of  $\theta$  and  $\rho$ . This density is shown in Figure 8. It has  $E(\theta_{T+1}|R) = 0.010$  and  $\sigma_{\theta_{T+1}} = 0.010$ . Thus, even accounting for parametric uncertainty, incorporating the variation predicted by the one-factor model increases the prediction standard error relative to the Binomial model by a factor of 8. Comparing with the conditional Model 2, unconditioning only adds 0.001 to the forecast standard error (about 11%).

For Model 3 the conditional density (on lagged defaults, not on parameters) is

$$p(\theta_{T+1}|\theta_T, R) = \iiint p(\theta_{T+1}|\theta_T, \theta, \rho, \tau)p(\theta, \rho, \tau|R)d\theta d\rho d\tau$$

where the integrals are definite. This density for the two trial values of lagged  $\theta$ , namely 0.004 and 0.015, are graphed in Figure 9. Summary statistics are  $E(\theta_{T+1}|\theta_T = 0.004, \theta, \rho, R) = 0.006$  with  $\sigma_{\theta_{T+1}} = 0.004$ , and  $E(\theta_{T+1}|\theta_T = 0.015, \theta, \rho, R) = 0.014$  with  $\sigma_{\theta_{T+1}} = 0.010$ . Here marginalizing with respect to parameters has reduced both the forecast values and the forecast standard errors.

Differences among these models can be summarized by considering the highest posterior density intervals for the forecast. These are the shortest intervals with the specified posterior probability. These are shown in Table 1.

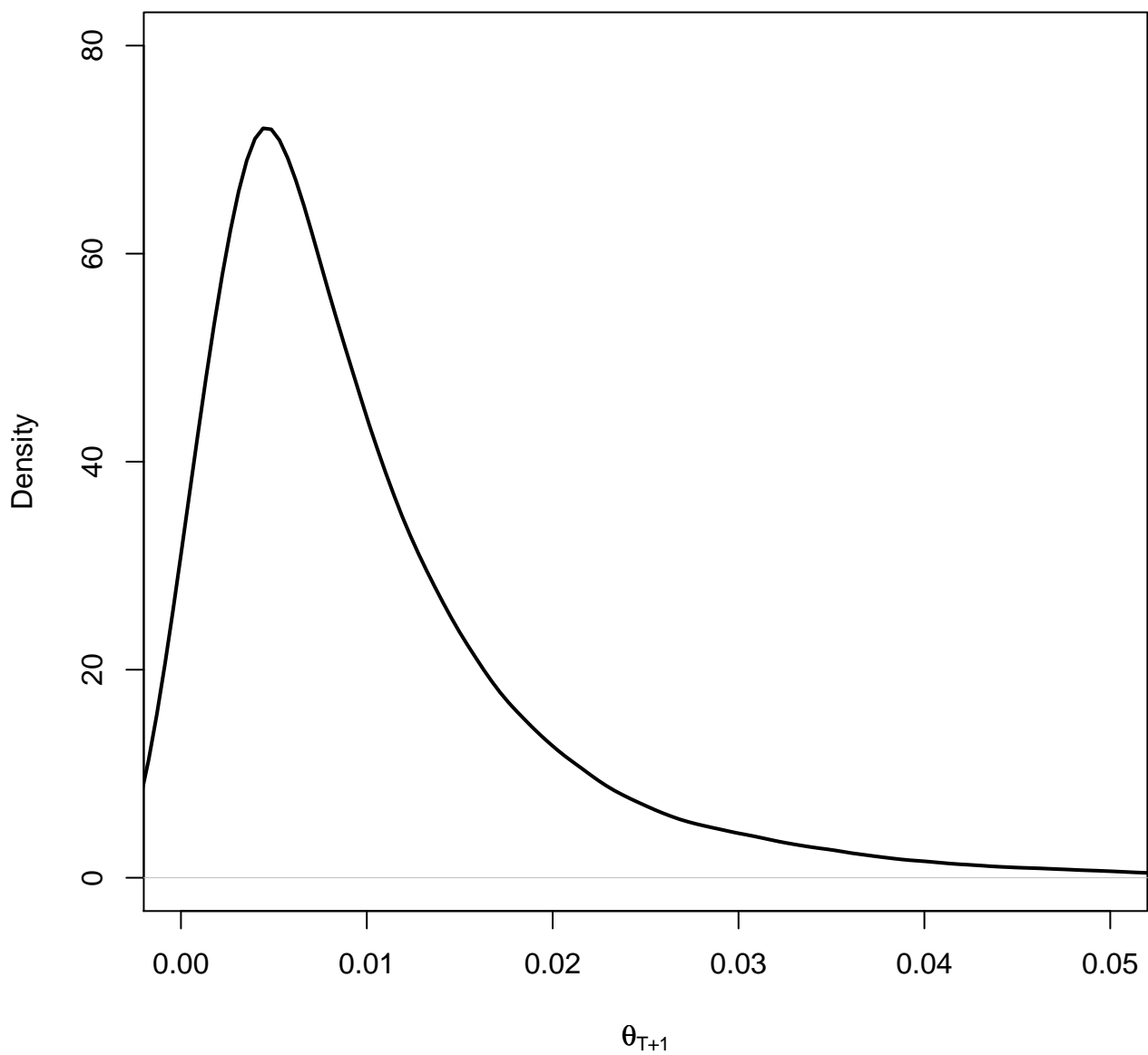


Figure 8: Predictive density  $p(\theta_{T+1}|R)$  from Model 2.

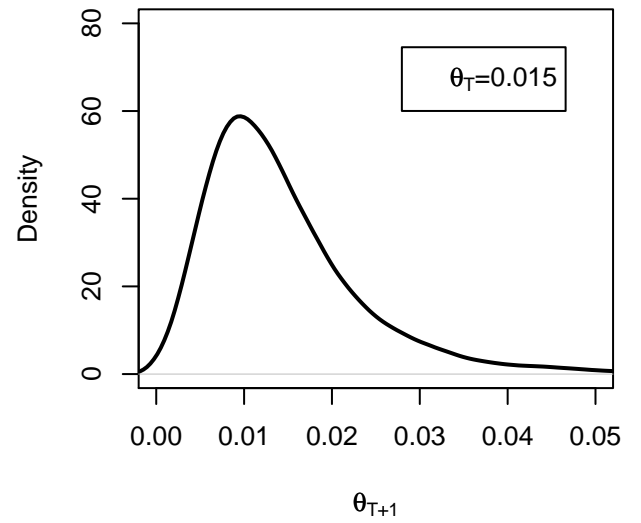
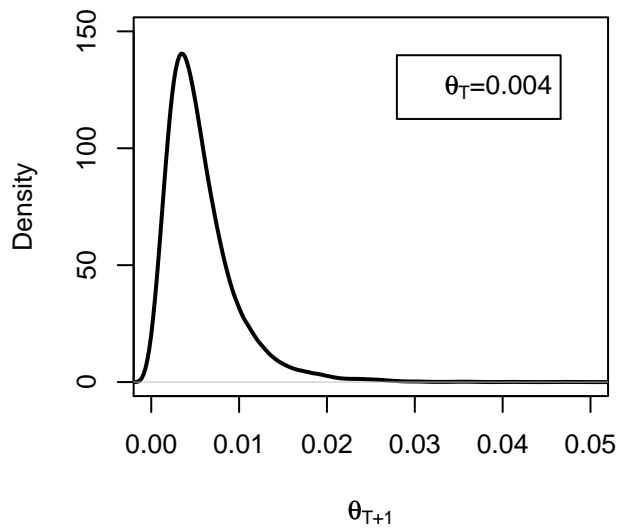


Figure 9: Predictive densities  $p(\theta_{T+1}|\theta_T, R)$  from Model 3.



**Table 1: Highest Posterior Density Intervals**

<i>Model</i>	$p = .9$	$p = .75$	$p = .5$
<i>Binomial</i>	0.0096	0.0096	0.0096
<i>2 - parameter</i>	0.0008, 0.0241	0.0008, 0.01506	0.0012, 0.0089
<i>3 - p, <math>\theta_T = 0.004</math></i>	0.0002, 0.0143	0.0007, 0.0093	0.0006, 0.0053
<i>3 - p, <math>\theta_T = 0.015</math></i>	0.0009, 0.0346	0.0016, 0.0218	0.0031, 0.0138
<i>Binomial(m)</i>	0.0063, 0.0121	0.0077, 0.0118	0.0089, 0.0111
<i>2 - parameter(m)</i>	0.0002, 0.0217	0.0005, 0.0133	0.0011, 0.0078
<i>3 - p, <math>\theta_T = 0.004(m)</math></i>	0.0006, 0.0109	0.0010, 0.0075	0.0017, 0.0053
<i>3 - p, <math>\theta_T = 0.015(m)</math></i>	0.0022, 0.0267	0.0038, 0.0194	0.0056, 0.0143

Note:  $(m)$  denotes marginal with respect to parameters.

## 6 Robustness - less prior information on rho

Our prior distribution on the asset correlation  $\rho$  strictly weakened the B2 prior, point mass on a specified value. Nevertheless, we chose a fairly tight distribution, reflecting some of the confidence apparent in the B2 specification. It is prudent however to run the analysis with a much looser prior as a robustness or validation exercise. There are guidelines for validation for the methods typically seen in practice - see OCC (2000). These required procedures note that any analysis depends on subjective business judgement as well as data and computations and stress the importance of rigorous evaluation. The Bayesian approach encourages making subjective information explicit. Although the report does not specifically address Bayesian methods, it seems likely that a robustness analysis using alternative, perhaps less informative priors would be a part of a Bayesian validation exercise. Here, we use the weaker prior shown in Figure 2. The prior on  $\tau$  cannot really be weakened - it is uniform on  $[-1,1]$ , weakening the B2 specification (zero) as much as possible while preserving the zero mean. The prior on  $\theta$  actually depends on an elicitation and is in accord with the data so weakening it is unlikely to be informative. Thus, we proceed with the joint prior which is given by the product of the marginals using the weak prior on  $\rho$ . The posterior marginal distributions are given in Figure 10. The moments are:  $E(\theta|R) = 0.0097$  and  $\sigma_\theta = 0.0019$ ,  $E(\rho|R) = 0.018$  and  $\sigma_\rho = 0.014$ , and  $E(\tau|R) = 0.641$  and  $\sigma_\tau = 0.166$

The posterior mean for the long-run default rate  $\theta$  is almost unchanged and the posterior mean for the autocorrelation  $\tau$  is reduced slightly (relative to the

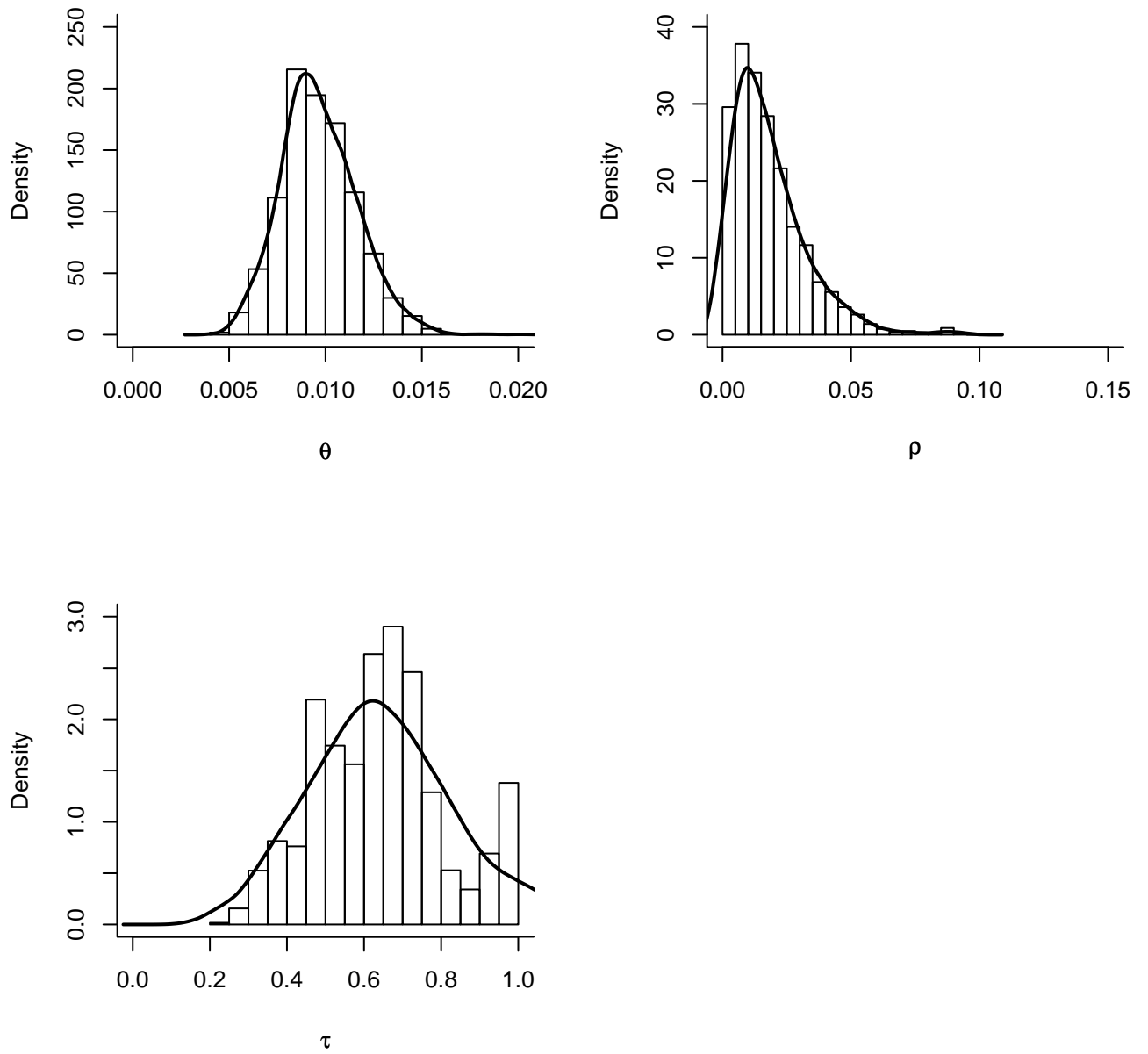


Figure 10: Posterior marginals using the weak prior.

standard deviations). The posterior mean for the asset correlation  $\rho$  is reduced dramatically from 0.091. This is evidence that the B2 specification for this parameter is in conflict with at least this data set. The reduction has strong implications for the predictability of annual default rates. The predictive distributions with the parameter values set at their posterior means, showing only uncertainty generated by the model, are shown in Figure 11. The summary statistics are  $E(\theta_{T+1}|\theta_T = 0.004, \theta, \rho, R) = 0.0055$  with  $\sigma_{\theta_{T+1}} = 0.0023$ , and  $E(\theta_{T+1}|\theta_T = 0.015, \theta, \rho, R) = 0.0135$  with  $\sigma_{\theta_{T+1}} = 0.0048$ . The predictive distribution is much tighter as shown by the graphs and by the calculated standard deviations.

Turning to the unconditional (on parameters) case the forecast densities reflecting both uncertainty due to the model and that due to parameter uncertainty are shown in figure 12. The summary statistics are  $E(\theta_{T+1}|\theta_T = 0.004, \theta, \rho, R) = 0.0056$  with  $\sigma_{\theta_{T+1}} = 0.0016$ , and  $E(\theta_{T+1}|\theta_T = 0.015, \theta, \rho, R) = 0.0129$  with  $\sigma_{\theta_{T+1}} = 0.0036$ .

The message here is that the data favor values of the asset correlation much smaller than those suggested by B2. The implication is that default rates are substantially more predictable than the B2 values imply. It is widely thought that the B2 model suggests that default rates are more variable over time than is the case in normal times.; our approach provides formal evidence on this issue. Note the substantial reduction in the sizes of the HPD intervals shown in Table 2 compared with those in Table 1.

**Table 2: Highest Posterior Density Intervals: Weak Prior**

<i>Model</i>	$p = .9$	$p = .75$	$p = .5$
$3 - p, \theta_T = 0.004$	0.0022, 0.0089	0.0033, 0.0078	0.0037, 0.0062
$3 - p, \theta_T = 0.015$	0.0061, 0.0205	0.0079, 0.0175	0.0087, 0.0141
$3 - p, \theta_T = 0.004(m)$	0.0030, 0.0082	0.0038, 0.0073	0.0044, 0.0063
$3 - p, \theta_T = 0.015(m)$	0.0073, 0.0177	0.0088, 0.0154	0.0102, 0.0138

Note:  $(m)$  denotes marginal with respect to parameters.

## 7 Conclusion

Given an economic model for default rates, forecast default rates are uncertain because the generating model is stochastic (if indeed it is a stochastic model) and because of uncertainty about the parameters of the model. The former is a much larger source of uncertainty in the application and dataset considered in this paper.

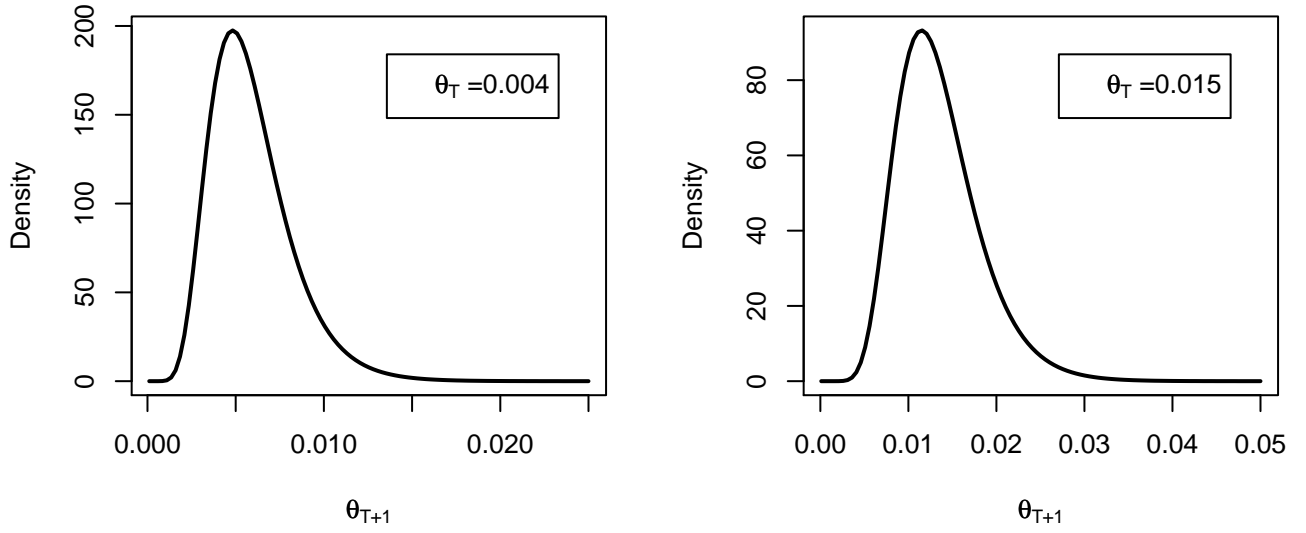


Figure 11: Predictive densities  $p(\theta_{T+1}|\theta_T, \theta, \rho, \tau, R)$  from Model 3 with the weak prior.

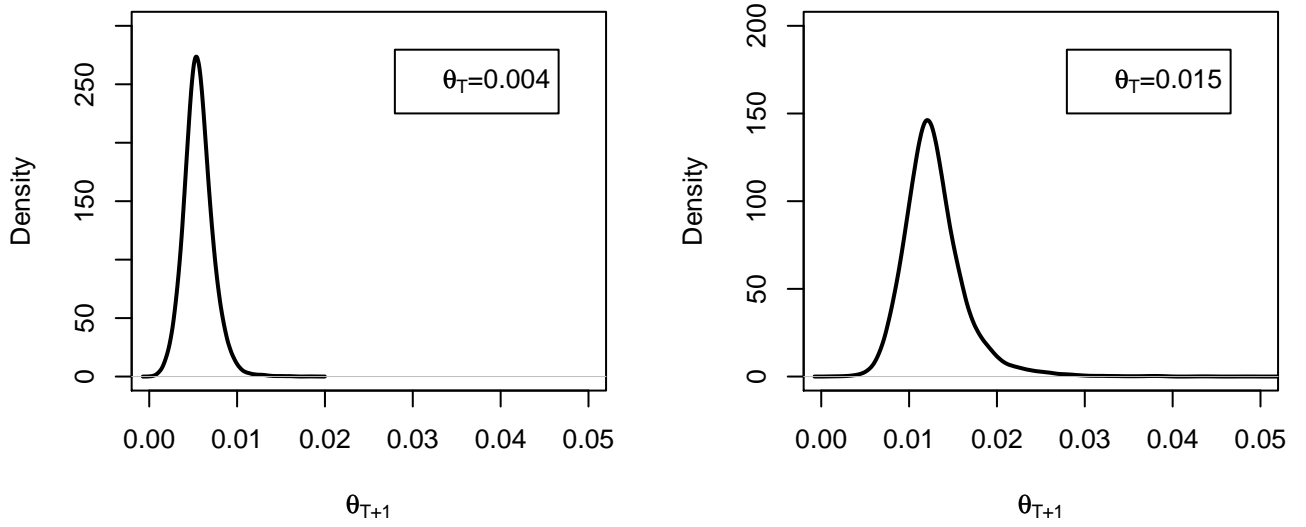


Figure 12: Predictive densities  $p(\theta_{T+1}|\theta_T, R)$  from Model 3 with the weak prior.

The model is the single-factor model underlying the B2 minimum capital calculations. Two variations are considered: a simplification - the binomial model with no systematic factor, widely used in applications and the starting point for deeper analyses, and a generalization allowing temporal correlation in the systematic factor. The latter allows conditional forecasting, and it is clear that, within the model, this period's default rate is useful in forecasting next period's. A Bayesian approach is taken with an assessed prior on the long-run default probability, a prior on the within-period asset correlation that reflects and weakens the B2 recommended specification (B2 fixes a value for the correlation, we choose a prior with mean equal to the B2 value but introduce some variance), and a uniform prior on the temporal correlation in the systemic risk factor. This too strictly generalizes the B2 model, which sets this value equal to zero. As a robustness check, we also consider a much weaker prior on the asset correlation. This turns out to be an interesting exercise, as it appears the data favor a much smaller value than that specified in B2. This has important implications for the predictability of default rates. Looking across industry experience, it appears that the stochastic model generates too much uncertainty, in that default rates are not as difficult to predict as the model implies. The model with the weakened prior gives a lower asset correlation, implying that defaults are more predictable over time. Nevertheless, there is still rather a lot of prediction variance implied by the model. This introduces another and more difficult source of forecast uncertainty: model uncertainty. Modeling model uncertainty is a difficult but important task and certainly merits further attention. The usual procedure of considering a sequence of increasingly general models is probably a good start.

In this and related applications the econometrician faces the dual chore of modeling the data distribution with a specification of a statistical distribution and modeling expert information with a statistical distribution. Adding the latter task substantially increases the range of applicability of econometric methods. This is clearly an area for further research.

## References

BASEL COMMITTEE ON BANKING SUPERVISION (2006): "International Convergence of Capital Measurement and Capital Standards: A Revised Framework, Comprehensive Version," Bank for International Settlements.

- (2009): “Proposed enhancements to the Basel II framework: Consultative Document,” Discussion paper, BIS.
- DAS, S. R., D. DUFFIE, N. KAPADIA, AND L. SAITA (2007): “Common Failings: How Corporate Defaults are Correlated,” *Journal of Finance*, 62, 93–117.
- FRANKS, J. R., AND S. A. DAVYDENKO (2008): “Do Bankruptcy Codes Matter? A Study of Defaults in France, Germany, and the U.K.,” *The Journal of Finance*, 63(2), 565–608.
- GARTHWAITE, P. H., J. B. KADANE, AND A. O’HAGAN (2005): “Statistical Methods for Eliciting Probability Distributions,” *Journal of the American Statistical Association*, 100, 780–700.
- GEYER, C. J. (2009): *mcmc: Markov Chain Monte Carlo*. R package version 0.6.
- GLENNON, D., AND N. M. KIEFER (2008): “Tests for Credit Model Performance,” Discussion paper, Cornell University.
- GORDY, M. B. (2003): “A Risk-Factor Model Foundation for Ratings-Based Bank Capital Rules,” *Journal of Financial Intermediation*, 12, 199–232.
- HANSON, S. G., AND T. SCHUERMANN (2006): “Confidence Intervals for Probabilities of Default,” *Journal of Banking and Finance*, 30, 2281–2301.
- KADANE, J. B., AND L. J. WOLFSON (1998): “Experiences in Elicitation,” *The Statistician*, 47(1), 3–19.
- KIEFER, N. M. (2008): “Default Estimation, Correlated Defaults and Expert Information,” Discussion paper, CAE, Cornell University.
- (2009): “Default Estimation and Expert Information,” *Journal of Business and Economic Statistics*, (forthcoming).
- NICKELL, P., W. PERRAUDIN, AND S. VAROTTO (2000): “Stability of Rating Transitions,” *Journal of Banking and Finance*, 24, 203–227.
- OCC (2000): “OCC Bulletin 2000-16, Subject: Risk Modeling, Description: Model Validation,” Discussion paper, Office of the Comptroller of the Currency.

O'HAGAN, A., C. E. BUCK, A. DANESHKHAH, J. R. EISER, P. GARTHWAITE, D. J. JENKINSON, J. E. OAKLEY, AND T. RAKOW (2006): *Uncertain Judgments: Eliciting Experts' Probabilities*. Chichester: John Wiley & Sons.

R DEVELOPMENT CORE TEAM (2009): *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

SCHUSTER, E. F. (1985): "Incorporating support constraints into nonparametric estimators of densities," *Communications in Statistical Theory and Methods*, 14, 1123–1136.