

# The distribution of defaults and Bayesian model validation

**Douglas W. Dwyer**

Moody's KMV, 100 Broadway, New York, NY 10005, USA;  
email: douglas.dwyer@mkmv.com

*Quantitative rating systems are increasingly being used for the purposes of capital allocation and pricing credits. For these purposes, it is important to validate the accuracy of the probability of default (PD) estimates generated by the rating system and not merely focus on evaluating the discriminatory power of the system. The validation of the accuracy of the PD quantification has been a challenge, fraught with theoretical difficulties (mainly, the impact of correlation) and data issues (eg, the infrequency of default events). Moreover, models – even “correct” models – will over-predict default rates most of the time. Working within the standard single-factor framework, we present two Bayesian approaches to the level validation of a PD model. The first approach provides a set of techniques to facilitate risk assessment in the absence of sufficient historical default data. It derives the posterior distribution of a PD, given zero realized defaults, thereby providing a framework for determining the upper bound for a PD in relation to a low default portfolio. The second approach provides a means for monitoring the calibration of a rating system. It derives the posterior distribution of the aggregate shock in the macro-economic environment, given a realized default rate. By comparing this distribution to the institution’s view of the stage of the credit cycle its borrowers are in, this approach provides useful insight for whether an institution should revisit the calibration of its rating system. This method allows one to determine that a calibration needs to be revisited even when the default rate is within the 95% confidence level computed under the standard approach.*

## 1 INTRODUCTION

A single number bet on an American roulette wheel pays out 36 times the initial bet with a probability of  $1/38$ . From the Casino’s perspective, they will win if after playing 36,000 times there are fewer than 1,000 losses. On average, they would expect 947.4 losses with a standard deviation of 30.4. The probability that a Casino

---

I am grateful to Joseph McQuown, Bill Morokoff, Martha Sellers, Jamie Stark, Roger Stein, Kenneth Wee, Sarah Woo and Andrew Zhang for helpful comments and suggestions. Any remaining errors are, of course, my own.

will lose money on 36,000 spins of the wheel is 4.2%.<sup>1</sup> The reason a Casino can afford such a narrow spread on a bet whose payoffs are so skewed is that a roulette table is carefully designed to yield independently, identically distributed random numbers. This fact makes it relatively easy to determine whether the roulette table is functioning properly. For example, the odds of the Casino experiencing a loss rate of greater than or equal 2.95% rather than the 2.63% expected loss rate on 36,000 bets is 1 in 10,000 for a properly functioning roulette table.

In credit risk management, the issues are fundamentally more complicated, owing to two characteristics. First, the “true default rate” of an actual exposure is more difficult to ascertain than the probability of losing on a roulette table owing to the limited number of existing default events and data availability issues. Second, there is correlation in default events. Put simply, there is some uncertainty regarding the underlying default probability of any given exposure and when defaults occur they tend to be cyclical, for example, a recession will trigger a cluster of default events. As an example, Figure 1 presents the one-year default rate from 1970–2005 for investment grade issuers as rated by Moody’s Investors Service. The average default rate over this time period was 0.068%. Nevertheless, in 22 of these years there were actually zero defaults. Occasionally, the default rates for investment grade credits are elevated and these periods generally correspond to recessionary periods in the economy such as in 1973, 1982, 1991 and 2000–2002.

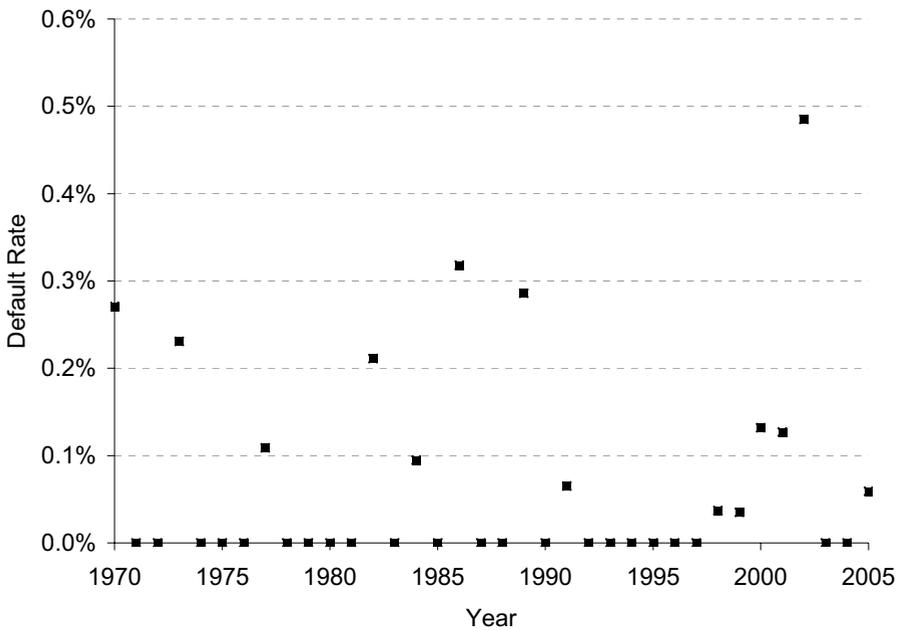
The implications of correlation for model validation are explicitly recognized by regulators in the Basel Accord as it pushes banks to improve their estimates of standalone risk in the form of borrower probability of defaults (PDs) and account for correlations in capital via the Asymptotic Single Risk Factor Model (cf, Basel Committee on Banking Supervision (2005)). Furthermore, regulators recognize that tests of PD model calibration that assume no correlation are flawed and that current tests which allow for correlation will only identify extreme issues with a model’s calibration (cf, Tasche (2005)).

This paper provides a means for assessing when the realized default rate will differ from the default rate forecasted by a model. We present the standard single factor framework that can be used to describe the distribution of defaults in a bucket of exposures with uniform PDs, given the correlation of exposures in the bucket and the predicted PD for the bucket.<sup>2</sup> According to this framework,

---

<sup>1</sup>These numbers are computed using the central limit theorem to approximate a binomial distribution with a normal distribution (cf, Stein (2006)).

<sup>2</sup>This framework – which has been called the Vasicek Model, a Gaussian Copula, or a Gaussian one-factor model – is commonly utilized by academics, practitioners, and regulators. One application is level validation of a PD model – showing that the observed default rates are consistent with the PDs produced by a model. To our knowledge, the first application of this framework to the level validation of a PD model was by Kurbat and Korablev (2002). Other examples include determining regulatory capital and pricing of collateralized debt obligations (CDOs). Nevertheless, the Gaussian assumption has been criticized for understating the risk of multiple extreme events relative to other distributions (cf, Embrechts *et al* (2001)). Furthermore, multi-factor models have been used to compute economic capital with greater realism.

**FIGURE 1** The default rate of investment grade issues 1970–2005.

Source: Exhibit 30 of Hamilton and Varma (2006).

correlations play a decisive role in the testing of the quality of model calibration. This framework can be extended to compute the distribution of defaults across multiple buckets and the distribution of the default rate for one bucket observed over multiple time periods.

The paper presents two Bayesian approaches for validating the accuracy of the forecasted PD estimates generated by rating models by building on the foundations of the single-factor framework. The first approach allows us to determine the posterior distribution of the PD given the realized sample default rate and an uninformed prior. Specifically, suppose we know nothing about what the PD for a bucket should be; this could imply a prior distribution for the PD that is uniformly distributed between 0 and 1.<sup>3</sup> Given an observation of zero defaults for the bucket, the posterior distribution of PDs can be solved for given

<sup>3</sup>In Bayesian inference, one combines a prior distribution with observed outcomes to determine a posterior distribution of the parameter in question. The prior distribution is intended to represent one's knowledge of the problem at hand. There is always a certain amount of judgment involved in this choice. One justification for an uninformed prior is that it imposes as little of the statistician's prejudice onto the data as possible. Furthermore, from a technical standpoint a uniform prior is analytically convenient, which makes it a common choice for a parameter that is known to have both an upper and lower bound (cf, Gelman *et al* (1995, Ch. 2)). In Appendix C, we show how the results change for an alternative prior.

a correlation assumption. Therefore, one can determine an upper bound for what the PD could be, such that we are 95% certain that the PD is below this value. This approach will prove useful in validating the level of the PD in relation to models for low-risk portfolios.

The second approach allows us to test the forecasted PD against our beliefs regarding the distribution of the aggregate shock in the macro-economic environment given the realized default rate in the bucket – assuming that the PD model is correct. This distribution can be compared with knowledge regarding the state of the economy over which the model was tested. For example, suppose one observes a 3% default rate in a bucket of 1,000 exposures that had a PD of 1%. With a correlation of 0.2, one will observe such an elevated level of defaults more than 5% of the time. Nevertheless, we show that this level of default is only consistent with the predicted PD of 1% if there was a large negative shock that occurred during this time period. If the general business conditions surrounding the episode are not indicative of elevated default rates, then one should consider revising the calibration of the model that produced the PD of 1%. The standard approach of computing the  $p$ -value of the default rate for a given correlation assumption would not reach such a conclusion.

In Appendix D, we present an application of this approach to the default rate time series for investment and speculative grade issuers from 1970 to 2005 as computed by Moody's Investors Service. We show that the model can be calibrated to these time series. Furthermore, the model is consistent with many properties of these time series. Finally, we show how the approach could be used to reject the hypothesis that the speculative grade default rate is 1.4% using data from either 2004 or 2005 even though the  $p$ -value associated with this hypothesis is less than 90% in both years under the standard approach.

The framework employed in this paper has been used widely throughout the credit risk literature. The framework perhaps began with Vasicek (1987). The main contribution of this paper is the application of Bayesian methods to model validation. For example, many authors have used classical statistics to derive the set of possible PDs that would not be rejected given  $N$  observations and zero defaults.<sup>4</sup> The first Bayesian application in this paper, in contrast, solves for the posterior distribution of the PD given  $N$  observations, zero defaults and a prior distribution of the PD, from which a Bayesian style confidence interval can be derived. The second Bayesian application in this paper uses the realized default rate to derive the posterior distribution of the aggregate shock to the portfolio – given a PD, a sample size and a correlation parameter. This distribution can then be compared to one's general knowledge of the factors impacting the economy at that time, which, in turn, provides a useful check on the plausibility of the model's calibration.

In the context of correlated defaults, explicit analytic solutions are often not available. While others have employed analytic approximations and bootstrapping

---

<sup>4</sup>See, for example, Christensen *et al* (2004). Pluto and Tasche (2005) also use this technique in the context of a Gaussian one-factor model.

or both (cf, Tasche (2003) and Hanson and Schuermann (2006)), our approach is to rely on numerical integration and Monte Carlo methods. For our taste, numerical integration and Monte Carlo methods allow for more direct derivations, albeit at the cost of requiring more sophisticated software implementations.

This paper is organized as follows: the next section shows how the standard one-factor framework can be employed to compute the distribution of defaults using Monte Carlo techniques; Section 3 shows how to use Bayesian approaches to compute a posterior distribution for the PD in a zero-default portfolio and to compute a posterior distribution for the aggregate shock given an observed sample default rate, the latter technique can be used to reject a model calibration even when the sample default rate falls within the 95% confidence level implied by the standard approach; Section 4 summarizes some data issues that must be addressed when applying the technique; and the implications for model building and calibration are discussed in the conclusion. There are four appendices: Appendix A presents some simple SAS code for simulating the distribution of defaults; Appendix B second discusses what constitutes a reasonable assumption for the correlation parameter within of a single-factor model; Appendix C relaxes the assumption of a uniform prior distribution for PDs in the context of zero observed defaults; and Appendix D presents an application of the technique to the default rate time series for investment and speculative grade issuers as provided by Moody's Investors Service.

## 2 THE STANDARD SINGLE-FACTOR GAUSSIAN FRAMEWORK

Consider a situation where there is a bucket of 1,000 exposures with a uniform PD of 1% (according to the output of a PD model in a rating system) and five defaults have been observed. Is the model output wrong? How does one evaluate the quality of the calibration of the model? What if one observes 15 or 25 defaults instead?

To address this issue, we seek to apply the work of Kurbat and Korablev (2002), Tasche (2005) and Stein (2006) to develop a single-factor framework that determines the distribution of defaults in a bucket.<sup>5</sup> In the later part of this section, we will present a number of analytic extensions to the framework. Appendix A provides a recipe and corresponding SAS code to simulate the distribution of defaults in a bucket given the number of exposures, the PD and the correlation assumption.

---

<sup>5</sup>This section will be written from the perspective of classical statistical inference. Specifically, we will take that  $PD = 1\%$  as the null hypothesis. Under classical statistics, one will reject the null hypothesis if the probability that an outcome as big or bigger (or alternatively as small or smaller) than the observed outcome is below some critical value such as 5% or 1% given the null hypothesis. If one observes an outcome that is not extreme given the null hypothesis, then one "fails to reject" the null hypothesis. A failure to reject a null hypothesis is not equivalent to confirming the null hypothesis with any degree of certainty – it does not show that the PD is correct.

A single-factor Gaussian model can be motivated from the perspective of a structural model.<sup>6</sup> Under a structural model, asset value follows a Brownian motion. At the horizon – in this case one year – if the asset value falls below the boundary default occurs, where the default boundary can be interpreted as a required debt payment. Therefore, the likelihood of default is determined by leverage adjusted for volatility which is summarized as “distance to default”. Given a PD, one can set the volatility of assets to one and solve for the level of leverage that produces this PD.

This single-factor framework essentially relates changes in the distance to default, ie, default probabilities, to asset return correlations. It decomposes PD into a function of a single factor that models the systemic risk, with a residual representing the idiosyncratic risk of an individual exposure. We start by assuming  $N$  exposures in a bucket. These exposures have a uniform probability of default denoted by PD. Assuming an underlying structural model, one can compute an *adjusted distance to default* (DD) as

$$DD = -\Phi^{-1}(PD) = \Phi^{-1}(1 - PD)$$

where  $\Phi^{-1}$  denotes the inverse of a cumulative normal distribution.<sup>7</sup> In order to simulate a distribution of defaults, one uses a correlation model to determine changes to DD.

Suppose that

$$\begin{aligned} r_j &= \sqrt{\rho} r_m + \sqrt{(1 - \rho)} e_j \\ r_k &= \sqrt{\rho} r_m + \sqrt{(1 - \rho)} e_k \end{aligned}$$

where  $r_j$  and  $r_k$  are the asset returns of two firms,  $j$  and  $k$ ,  $r_m$  is the return of the market portfolio,  $\rho$  is the variation of the firm’s returns explained by the market in percent, which is assumed to be constant for the two firms.  $r_m$ ,  $e_j$ , and  $e_k$  are idiosyncratic draws, or shocks, from a standard normal distribution. The correlation between the asset returns of the two firms is given by

$$\rho(r_j, r_k) \equiv \frac{\text{Cov}(r_j, r_k)}{(\text{Var}(r_j) \text{Var}(r_k))^{1/2}} = \rho$$

In a pure Merton model, the 1-year distance to default is defined as

$$DD = \frac{\log(V_A/X) + (\mu - \sigma_A^2/2)}{\sigma_A}$$

<sup>6</sup>In our view, greater realism can be achieved with a multi-factor model. For example, Moody’s KMV Portfolio Manager™ uses multiple factors representing industry, country and region risk in addition to global economic risk.

<sup>7</sup>Note that this approach differs from the approach used to map a DD to an Expected Default Frequency (EDF™) credit measure that is employed by the Moody’s KMV Public Firm Model. The Moody’s KMV Public Firm Model uses an empirical mapping.

where  $V_A$  is asset value,  $X$  is the default point,  $\mu$  is the expected return on assets (adjusted for cash outflows) and  $\sigma$  is the volatility of assets. Assuming that  $X$ ,  $\mu$  and  $\sigma$  remain constant over the period, the change in DD is given by

$$\Delta DD = \frac{\log(V_{At}) - \log(V_{At-1})}{\sigma_A} = \frac{r}{\sigma_A}$$

where  $r$  is the return on the firm's assets. Therefore, a *portfolio outcome* can be determined by drawing one aggregate shock and  $N$  idiosyncratic shocks from a standard normal distribution. For a given PD, we set the asset volatility to 1 and solve for the degree of leverage that yields the appropriate PD.<sup>8</sup> Therefore, the change in DD for exposure  $j$  is simply

$$\sqrt{\rho} r_m + \sqrt{1 - \rho} e_j$$

Default for exposure  $j$  occurs if the new DD is negative:

$$\sqrt{\rho} r_m + \sqrt{1 - \rho} e_j - \Phi^{-1}(\text{PD}) < 0$$

One can simulate a portfolio outcome by drawing one aggregate shock and an idiosyncratic shock for each exposure and compute the new DD for each exposure. One then simply counts the number of defaults for all exposures and this number forms a portfolio outcome. One then repeats this simulation many times recording the number of defaults for each simulation.<sup>9</sup> From the distribution of the number of defaults across simulations, one can compute the median number of defaults, mean defaults, the 5th percentile of defaults, 95th percentile of defaults and so forth. The respective default rates are then derived by dividing the number of defaults by  $N$ . A recipe and sample SAS code are provided in Appendix A that perform this calculation.<sup>10</sup>

Table 1 shows some results of this procedure for different numbers of exposures and assumptions of correlation, given a PD of 1%. The table presents the mean, the median, the 5th percentile (P5) and the 95th percentile (P95) of the default rate across 100,000 samples obtained from Monte Carlo simulations. We note three general observations from this set of results. First, in the absence of correlation the upper bound default rate remains close to the mean default rate for the exposures in the bucket. As gleaned from the case with 10,000 exposures and  $\rho$  equal to zero, the default rate in the 95th percentile is 1.17%. Second, upper bound default rates rise markedly with increasing correlation. Even where correlation is moderate

<sup>8</sup>Setting volatility to 1 is a scaling assumption.

<sup>9</sup>For a sufficiently large number of observations, Vasicek (1991) has derived a closed-form solution to this problem.

<sup>10</sup>For this simple problem, one could also do numerical integration. For a one-dimensional integral, numerical integration is computationally more efficient – the same degree of accuracy can be achieved with fewer calculations. Nevertheless, simple extensions to the framework (eg, a multi-period problem) will increase the dimensionality of the problem quickly. Higher-dimensional problems are often easier to accommodate with Monte Carlo methods than numerical integration.

**TABLE 1** Simulation results where the actual PD is 1%.

$\rho$	Exposures	Mean default rate (%)	Median default rate (%)	P5	P95
0.0	100	1.0	1.00	0.00	3.0
0.0	1,000	1.0	1.00	0.50	1.5
0.0	10,000	1.0	1.00	0.84	1.2
0.2	100	1.0	0.00	0.00	4.0
0.2	1,000	1.0	0.50	0.00	3.8
0.2	10,000	1.0	0.46	0.03	3.8
0.4	100	1.0	0.00	0.00	5.0
0.4	1,000	1.0	0.10	0.00	4.9
0.4	10,000	1.0	0.13	0.00	4.9

(where  $\rho$  is equal to 0.2), the 95th percentile case becomes three times larger than the predicted PD of 1%. Finally, the median default rate is considerably lower than the prediction. The median being lower than the mean implies that most of the time a “correct” model will over-predict defaults.

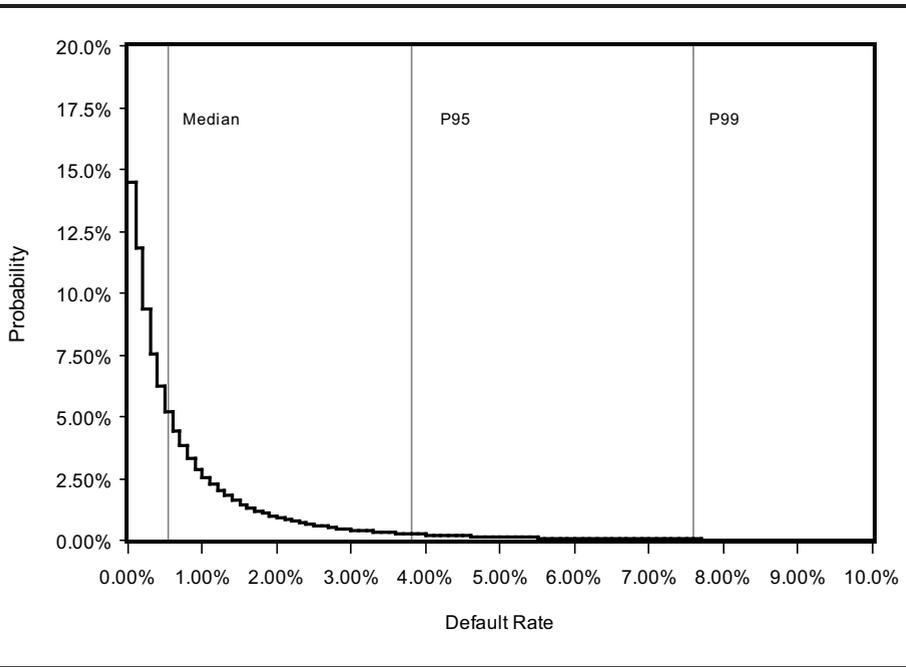
Based on the results in Table 1, our response to the questions posed at the beginning of this section is as follows. Assuming that the PD of 1% is correct and there is no correlation between exposures, there is a 5% chance of observing five defaults or fewer (a default rate of 0.50% or smaller) and a 5% chance of observing 15 defaults or more (a default rate of 1.5% or larger). With a modest amount of correlation (where  $\rho$  is assumed to be 0.2) there is a 5% chance of observing default rates that are 3.9% or higher. Therefore, observing 5, 15 or 25 defaults will not be unusual for a bucket with 1,000 exposures and a PD of 1%, given a reasonable amount of correlation (Appendix B contains a discussion as to what constitutes a reasonable amount of correlation). In the language of classical statistical inference, one would not reject the null hypothesis of a PD = 1% and  $\rho = 0.2$ .

To further illustrate the situation, Figure 2 presents the distribution of the default rate for 1,000 exposures with PDs of 1% and a  $\rho$  of 0.2. The grey lines mark the median, 95th and 99th percentiles of the distribution.

Using this framework, we extend the initial question to: What is the distribution of average default rates of a bucket over time, assuming that both the aggregate shocks and idiosyncratic shocks are both independent over time? Specifically, suppose that at the start of each year, we constructed a portfolio of 1,000 exposures all with a PD of 1%. What would be the average default rate of such a portfolio observed over several years? Table 2 presents the results from such an analysis with the time horizon varying from 4 to 12 years, for a portfolio of 1,000 exposures with a PD of 1%.<sup>11</sup>

<sup>11</sup>We performed this analysis with a straightforward extension of the recipe provided in Appendix A using 100,000 simulations.

**FIGURE 2** Distribution of defaults.



**TABLE 2** Average default rate for multi-year simulations.

$R^2$	Exposures	Years	Mean default rate (%)	Median default rate (%)	P5	P95
0	1,000	4	1.0	1.0	0.75	1.3
0	1,000	8	1.0	1.0	0.83	1.2
0	1,000	12	1.0	1.0	0.85	1.1
0.2	1,000	4	1.0	0.8	0.20	2.5
0.2	1,000	8	1.0	0.9	0.34	2.1
0.2	1,000	12	1.0	0.9	0.43	1.9
0.4	1,000	4	1.0	0.5	0.03	3.6
0.4	1,000	8	1.0	0.7	0.11	2.9
0.4	1,000	12	1.0	0.8	0.19	2.6

The results are fairly striking. For a portfolio with moderate correlation (where  $\rho$  is assumed to be 0.2) over a 12-year period, the default rate can range from 0.43% to 1.86%. Nonetheless, the impact of correlation is reduced for portfolios with longer time horizons. For samples with a longer time horizon, the gap between the mean and median default rates shrink, indicating a more symmetrical and less skewed distribution.

Furthermore, one may be interested in computing the expected number of defaults in a population with heterogeneous default probabilities. This scenario can be easily accommodated by computing the adjusted DD for each firm individually and modifying the computation in Step 4 of the recipe (Appendix A) accordingly. One needs to specify the PD and  $\rho$  for each firm.

### 3 ASSESSING PD CALIBRATION USING BAYESIAN APPROACHES

Now that we have established a framework for describing the distribution of defaults, we will show how the calibration of rating systems or PD models can be monitored using two different Bayesian approaches. Under Bayesian analysis, one defines a prior distribution characterized by one's assumption regarding the possible values for a parameter (or set of parameters) of interest. After observing an outcome, one updates the prior assumption, based on the actual outcome. In this section, we present two Bayesian approaches which can provide valuable insights in relation to two common modeling issues which occur in practice. The first approach helps determine the upper bound for a PD in relation to a low default portfolio or rating buckets with very few or zero observed defaults, while the second approach provides a means to monitoring the calibration of a rating system.

Under the first approach, we treat the PD as being unknown. We assume an uninformed prior – that we have no information regarding what the true PD would be. Consequently, our prior is that the PD is uniformly distributed between 0 and 1. We observe an outcome and update our prior on the basis of the outcome. This is particularly useful for the case of observing zero defaults.

Under the second approach, we treat the PD and asset correlation as being known and we update our expectation regarding the distribution of what the aggregate shock in the macroeconomic environment would have been in order to produce the observed sample default rate. One can then make a judgmental evaluation as to whether or not this distribution is consistent with one's knowledge of the general business conditions at the time.

The underlying analytics for both approaches are drawn from the foundation of the single-factor framework discussed above, and we will use the fact that the probability of default given the aggregate shock,  $r_m$ , is given by

$$\Pi(\text{PD}, r_m) \equiv \Phi\left(\frac{\Phi^{-1}(\text{PD}) - \sqrt{\rho} r_m}{\sqrt{1 - \rho}}\right)$$

#### 3.1 Bayesian updating with a prior that the PD has a uniform distribution

High-credit-quality firms typically form a large proportion of the large corporate segment in a bank's portfolio. However, defaults are generally rare events for such investment grade exposures. Furthermore, high-quality firms tend to be larger and more systematic (as reflected in the treatment of correlation as declining with increasing PDs in the Basel approach). Suppose that we have 10,000 of such exposures in a bucket where we have observed no defaults within one year. What

can we say about the PD for this bucket? The maximum likelihood estimate of the PD is zero. However, economic intuition and longer-term data suggest that even the PDs for high-credit-quality exposures are unlikely to be stable over time and the credit cycle. Between 1970 and 2005, realized default rates for investment grade credits ranged from 0 to 0.485%, and in 22 of these years the group experienced zero defaults (Figure 1). As such, what degree of confidence can we place in this estimate of zero derived from data that is, like any single bank's, quite limited?

To address this issue, we propose to update a prior distribution that reflects what the PD could be. The most uninformed prior as to what the PD could be is to say that it is uniformly distributed between [0, 1]. Assuming zero correlation, the probability of  $D$  defaults and that the actual probability of default is less than or equal to PD is given by

$$\text{Prob}(x \leq \text{PD}, D) = \int_0^{\text{PD}} P(D|x) f(x) dx$$

where

$$P(D|\text{PD}) = \binom{N}{D} \text{PD}^D (1 - \text{PD})^{N-D}$$

and  $f(x) = 1$ , ie, the probability density function of a uniform random variable.

By Bayes' rule the posterior cumulative distribution of PD given  $D$  is the probability of the set  $\{D, \text{pd} \in [0, \text{PD}]\}$  divided by the probability of the set  $\{D, \text{pd} \in [0, 1]\}$ . Therefore, the cumulative distribution of PD given  $D$  is given by

$$K(\text{PD}|D) = \frac{\text{Prob}(x < \text{PD}, D)}{\text{Prob}(D)} = \frac{\int_0^{\text{PD}} P(D|x) f(x) dx}{\int_0^1 P(D|x) f(x) dx}$$

For zero defaults, this distribution has the solution<sup>12</sup>

$$K(\text{PD}|D = 0) = 1 - (1 - \text{PD})^{N+1}$$

Based on this formula, we present the posterior distribution of the PD in Table 3 for a bucket with one, five, ten and 50,000 exposures, assuming zero defaults and no correlation. Where there are 1,000 exposures, one can be 95% certain that the true probability of default is less than 30 basis points. With 10,000 exposures, one can be 95% certain that the true probability of default is less than 3 basis points. For 50,000 exposures, the boundary becomes less than 1 basis point.

<sup>12</sup>We can solve for the posterior distribution for different prior distributions. For example, suppose that the prior was uniform on  $[0, \bar{P}]$  where  $\bar{P}$  is  $< 1$ . Then one can show that the solution is  $(1 - (1 - \text{PD})^{N+1}) / (1 - (1 - \bar{P})^{N+1})$  on  $[0, \bar{P}]$ . As  $(1 - (1 - \bar{P})^{N+1})$  is close to 1 for large  $N$ , this solution is very close to  $(1 - (1 - \text{PD})^{N+1})$  for large  $N$ . Therefore, the posterior distribution does not appear highly sensitive to the choice of prior for large  $N$ . Appendix C shows a similar result for another alternative prior distribution.

**TABLE 3** Posterior cumulative distribution of PD given no defaults and assuming no correlation.

PD (%)	Exposures			
	1,000	5,000	10,000	50,000
0.01	9.5%	39.3%	63.2%	99.3%
0.02	18.1%	63.2%	86.5%	100.0%
0.03	25.9%	77.7%	<b>95.0%</b>	100.0%
0.04	33.0%	86.5%	98.2%	100.0%
0.05	39.4%	91.8%	99.3%	100.0%
0.06	45.1%	<b>95.0%</b>	99.8%	100.0%
0.07	50.4%	97.0%	99.9%	100.0%
0.08	55.1%	98.2%	100.0%	100.0%
0.09	59.4%	98.9%	100.0%	100.0%
0.10	63.2%	99.3%	100.0%	100.0%
0.20	86.5%	100.0%	100.0%	100.0%
0.30	<b>95.0%</b>	100.0%	100.0%	100.0%

Allowing for correlation makes the problem somewhat more involved, because one needs to “integrate out” the aggregate shock. Note that the probability that the pd is less than or equal to PD, the aggregate shock is less than or equal to  $r_m$ , and that the number of defaults is equal to  $D$  is given by

$$\text{Prob}(x \leq \text{PD}, y \leq r_m, D) = \binom{N}{D} \int_0^{\text{PD}} \int_{-\infty}^{r_m} \Pi(x, y)^D (1 - \Pi(x, y))^{N-D} n(y) f(x) dy dx$$

where  $n(\cdot)$  is a standard normal density function and recall that  $\Pi(\text{PD}, r_m)$  is the PD given the aggregate shock. Therefore, the posterior distribution of PD is given by

$$K(\text{PD}|D) = \frac{\binom{N}{D} \int_0^{\text{PD}} \int_{-\infty}^{\infty} \Pi(x, r_m)^D (1 - \Pi(x, r_m))^{N-D} n(r_m) f(x) dr_m dx}{\binom{N}{D} \int_0^1 \int_{-\infty}^{\infty} \Pi(x, r_m)^D (1 - \Pi(x, r_m))^{N-D} n(r_m) f(x) dr_m dx}$$

Since  $\Pi(\text{PD}, r_m)$  is a nonlinear function of both PD and  $r_m$  it is unlikely that we will find an analytic solution to these double integrals. Nevertheless, they are amenable to numerical techniques.<sup>13</sup>

Table 4 presents a tabulation of the 95th percentile of this posterior distribution of the PD given zero defaults and a uniform prior for different numbers of exposures and correlations. As shown in Table 3, assuming zero correlation, for 10,000 exposures, the 95th percentile PD is 3 basis points. Therefore, from a

<sup>13</sup>As this function exhibits significant curvature when  $D = 0$  and  $N$  is large, we employ an adaptive quadrature method to evaluate the double integral.

**TABLE 4** 95th percentile of the posterior distribution of PD given zero defaults.

N	$\rho$							
	0	0.05	0.1	0.15	0.2	0.3	0.4	0.5
1,000	0.30%	0.74%	1.50%	2.65%	4.26%	8.92%	15.38%	23.27%
5,000	0.06%	0.20%	0.50%	1.06%	1.97%	5.13%	10.36%	17.58%
10,000	0.03%	0.11%	0.32%	0.72%	1.41%	4.05%	8.74%	15.59%
50,000	0.006%	0.03%	0.11%	0.29%	0.65%	2.32%	5.88%	11.78%
100,000	0.003%	0.02%	0.07%	0.19%	0.47%	1.83%	4.95%	10.44%

Bayesian perspective, if one observes 10,000 observations with no defaults one can be 95% certain that the actual PD is less than 0.03% under the assumption of no correlation. With a reasonable amount of correlation (0.2) and a large sample (100,000 exposures) a PD of 47 basis points can not be ruled out. For a high degree of correlation (0.5) a PD of 10.44% is possible even with zero defaults across 100,000 observations! This implies that even if the true PD for a bucket is relatively high, a large positive aggregate shock,  $r_m$  (ie, where we are in the good part of the credit cycle), coupled with substantial correlation can result in an observed default rate that is zero or very close to zero in a single period.

Based on this framework, institutions wishing to monitor the calibration of their rating systems on a zero-default portfolio should be estimating suitable correlation values from which a 95th percentile PD can be calculated.

### 3.2 Using Bayesian updating to size the aggregate shock

Suppose that we have a PD model that predicts a 1% default rate for a bucket with 1,000 exposures but we observed 30 defaults in a given year. According to Table 1 in Section 2, at a modest correlation of 0.2, we would expect such a default rate somewhat more often than 5% of the time. With zero correlation, such a default rate would be very unlikely. A possible inference is that the model may require recalibration. Alternatively, one may infer that there must have been a large negative shock in the macro-economic environment in order to generate such an elevated default rate. However, how do we compare this inference to our knowledge of the general business conditions and credit cycle impacting these exposures during the relevant time period? When should we consider revisiting the calibration of the model? To address this issue, we begin by noting that the distribution of the sample default rate (SDR) given  $r_m$  can be approximated by a normal distribution with a mean of

$$\Pi(r_m) = \Phi\left(\frac{\Phi^{-1}(\text{PD}) - \sqrt{\rho} r_m}{\sqrt{1 - \rho}}\right)$$

and a variance of  $\sigma_{r_m}^2 = (\Pi(r_m)(1 - \Pi(r_m)))/N$ . Note that, in contrast to Section 3.1, we have not included PD as a second argument to the function  $\Pi$ , because in this section we are treating PD as given. We have also approximated a binomial distribution by a normal distribution. Therefore, in this section we can

**TABLE 5** Aggregate shock's impact on default rates.

Aggregate shock size	Probability of aggregate shock or worse	Default rate mean $\Pi(r_m)$ (%)	Default rate std dev $\sigma_{r_m}^2$ (%)	<i>P</i> -value that default rate exceeds 3%
-2.00	2.3	5.5	0.7	1.00
-1.80	3.6	4.4	0.7	0.99
-1.60	5.5	3.6	0.6	0.84
-1.40	8.1	2.9	0.5	0.40
-1.20	11.5	2.3	0.5	0.06
-1.00	15.9	1.8	0.4	0.00
-0.80	21.2	1.4	0.4	0.00
-0.60	27.4	1.1	0.3	0.00
-0.40	34.5	0.8	0.3	0.00
-0.20	42.1	0.6	0.2	0.00
0.00	50.0	0.5	0.2	0.00

easily work with density functions as both the random variables,  $r_m$  and SDR, are now continuous variables.<sup>14</sup>

Based on this formula, Table 5 presents the probability of observing 30 or more defaults for different values of the aggregate shock,  $r_m$ , assuming a correlation of 0.2. The first column presents the aggregate shock which ranges from  $-2$  to  $0$ .<sup>15</sup> The second column presents the probability that the aggregate shock is less than or equal to this value. The third and fourth columns present the mean and standard deviation of the default rate given the aggregate shock and the fifth column presents the probability that the default rate exceeds 3% given the aggregate shock. If the aggregate shock were greater than  $-1$ , then the probability of observing the given default rate is essentially zero. Therefore, one can conclude that if the model is correct then there was a negative shock that was in excess of  $-1$ . Such an event has a *p*-value of 16% [ $\Phi(-1)$ ]. If the model is correct, then given that we observed a 3% default rate during the year, we can infer that this particular year was approximately a one-in-six-year event for companies in this portfolio. If this inference does not seem plausible according to our observations of the credit

<sup>14</sup>For a low default rate or a small sample size, one should work with the binomial distribution directly rather than resorting to the approximation of a normal distribution. When extending this analysis to the case of multiple buckets with different PDs, the normal approximation reduces the computational complexity considerably. If the distribution of the defaults in each bucket is assumed to be normally distributed given the aggregate shock, then the distribution of defaults across buckets also has a normal distribution whose mean and variance can be readily calculated from the size, mean and variance of each bucket. Hull and White (2004) work out the mathematics for determining the probability of exactly  $k$  defaults from a sample with heterogeneous PDs in the context of a CDO and discuss the corresponding numerical issues involved in the calculation.

<sup>15</sup>As these are draws from a normal distribution with a mean of 0 and a standard deviation of 1, the units are standard deviations.

cycle over the past year, then we should consider revisiting the calibration of the model.

One can take this analysis one step further, by deriving an expression for how certain we are that the aggregate shock must have been less than  $-1$  through Bayesian methods. Specifically, we derive a posterior distribution for the aggregate shock given our prior that the distribution is a standard normal combined with the observed SDR and taking as given the PD and the correlation assumption.

We can specify the probability density function of the sample default rate and the aggregate shock by

$$f(\text{SDR}, r_m) = h(\text{SDR}|r_m)g(r_m)$$

where  $g$  is the prior density function for the aggregate shock (a standard normal distribution), and  $h(\text{SDR}|r_m)$  is the conditional distribution of SDR given  $r_m$  (a normal distribution with a mean of  $\Pi(r_m)$  and a variance  $\sigma_{r_m}^2$ ). The posterior density function of  $r_m$  is then given by

$$k(r_m|\text{SDR}) = \frac{f(\text{SDR}, r_m)}{\int_{-\infty}^{\infty} h(\text{SDR}|r_m)g(r_m) dr_m}$$

The posterior cumulative distribution function (CDF) is simply

$$K(r_m|\text{SDR}) = \int_{-\infty}^{r_m} k(x|\text{SDR}) dx.$$

While both  $h(\text{SDR}|r_m)$  and  $g(r_m)$  have normal distributions, the mean and variance of  $h(\text{SDR}|r_m)$  are non-linear functions of  $r_m$ . Consequently, deriving an analytic expression for this integral does not appear to be straightforward. This integral can easily be approximated using the trapezoidal rule. The denominator can be calculated as

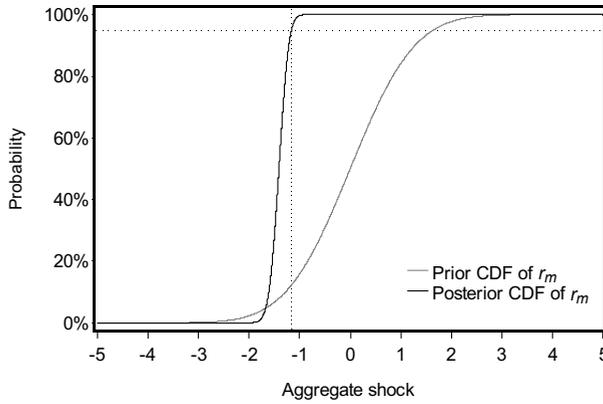
$$\int_{-\infty}^{\infty} h(\text{SDR}|r_m)g(r_m) dr_m \approx \frac{1}{\sqrt{2\pi}(N+1)} \sum_{i=0}^N \exp\left(-\frac{1}{2}(\text{zscore}(r_i)^2 + r_i^2)\right)$$

where  $\text{zscore}(r_m) = (\text{SDR} - \Pi(r_m))/\sqrt{\sigma_{r_m}^2}$  and  $r_i = -5 + i/(10N)$ , ie,  $r_m(i)$  runs from  $-5$  to  $5$ .

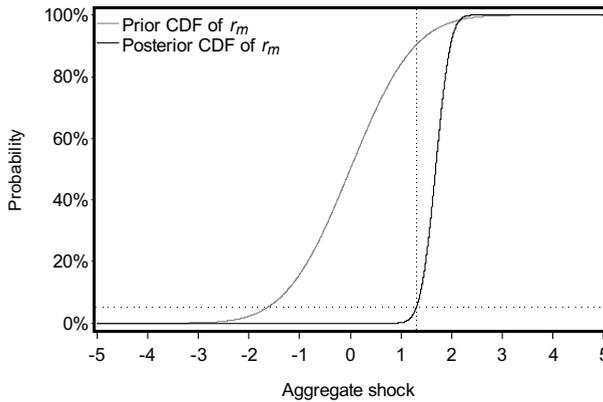
Similarly, the numerator is calculated in the same way, except that the summation runs from  $0$  to  $N(r_m)$  rather than to  $N$ , where  $N(r_m) = \max\{r_m(i)\}$  such that  $r_m(i) < r_m$ .

Based on this approach, we present two examples in Figure 3 of how we might monitor the calibration of the model. Figure 3(a) presents the prior and posterior CDF for  $r_m$  in our example. The dotted lines cross the CDF at the 95th percentile value. Given our prior, we are 95% certain that the aggregate shock is less than  $-1.16$ . The probability of the aggregate shock being less than or equal to  $-1.16$

**FIGURE 3** Prior and posterior CDF of the aggregate shock. (a) The prior and posterior distribution of the aggregate shock given a sample default rate of 3% from a bucket of 1,000 exposures with a PD of 1% and a correlation of 0.2. (b) The prior and posterior distribution of the aggregate shock given a sample default rate of 1% from a bucket of 1,000 exposures with a PD of 10% and a correlation of 0.2.



(a)



(b)

is 12% ( $\Phi(-1.16)$ ). Therefore, observing a default rate of 3% when the model predicts a default rate of 1% is only consistent with the default experience being the worst of one out of eight years, ie, at or close to the bottom of the credit cycle.

Figure 3(b) presents the result of how we can interpret a particularly low default rate. If one takes as given that  $PD = 10\%$  and  $\rho = 0.2$  and one encounters a sample default rate of 1%, then one could be 95% certain that the aggregate shock was larger than 1.3. Such an aggregate shock would occur approximately one out of 10 years ( $1 - \Phi(1.3)$ ). Therefore, unless the macroeconomic environment was

particularly favorable, such a low sample default rate would suggest that the model was overly conservative.

Note that both of these examples are fairly extreme. In the first, the sample default rate was three times the predicted and in the second the sample default rate was one-tenth. This does reflect the nature of credit as a unique asset class. Most outcomes in any one year for any one bucket will not provide enough information to conclude that the model produces PDs that are either too high or are too low with any degree of certainty (cf, Stein (2006), Tasche (2005)). Nevertheless, a large departure from the predicted value is informative regarding the aggregate shock that must have occurred. If this information is inconsistent with one's knowledge of the credit environment, one can reject a model calibration even though the observed default rate was within the 95% confidence level.

This analysis can easily be extended to the case of several buckets, because given the aggregate shock the distribution of defaults in each bucket is a binomial distribution that can be approximated by a normal distribution for a large sample and the distribution of the sum of normal distributions is a normal distribution.

Institutions that seek to calibrate their rating models correctly need to incorporate a methodology that accounts for the credit cycle. As macroeconomic shocks cause spikes in default rates unevenly across the portfolios of financial institutions, these institutions need to be able to quantify whether their rating systems are accurately capturing changes in credit quality with appropriate distributional shifts of defaults across rating buckets.

This method will allow one to determine that a model calibration needs to be revisited where the standard approach of computing a confidence level for a given correlation assumption (cf, Table 1, or Tasche (2005)) will not. Consider the previously mentioned example of observing 30 defaults in a bucket of a 1,000 exposures with a PD of 1%. Under a standard approach, one would not reject the hypothesis of a PD of 1% given 20% correlation. If 30 defaults were observed in an expansionary period in the economy, however, one would conclude that the model calibration needed to be revisited.

#### **4 DATA CONSIDERATIONS**

The above methodology is developed under the assumption that the data is clean – that a default is a clearly defined and meaningful concept, all defaults are captured in the data, and each default can be properly linked to the relevant information regarding the firm (eg, financial statements). Data samples used to develop, validate and calibrate private firm default risk models typically do not have all of these characteristics. Consequently, the interpretation of level validation studies must acknowledge and adjust for the data deficiencies. These issues include sample selection biases, data collection issues and differences in the definition of default. There are similar issues associated with level validation of ratings (cf, Hamilton and Varma (2006)) and EDF measures for publicly traded firms (cf, Kurbat and Korablev (2002)). Arguably, these issues are larger and more difficult to manage for the validation of private firm models. This section

provides an overview of the data considerations encountered in validating private firm models.

#### 4.1 Sample selection issues

When banks originate a new loan, they will ask for the two most recent financial statements in the loan application and enter both into their systems.<sup>16</sup> In such cases, it will be impossible for a firm to default within 15 months of the first financial statement, because the originating bank does not yet have a loan outstanding for the obligor to default on. In addition, the possibility of a firm defaulting within 15 months of the second financial statement will often be limited. For example, suppose one applied for a loan at the end of September of 2003 with financial statements dated 31 December for both 2001 and 2002. In order to default within 15 months of the second financial statement, the borrower would have to default before April of 2004, which is within six months of applying for the loan. The process of a loan becoming approved, the funds being disbursed, the ceasing of payments and finally becoming classified as a defaulter can easily use up more than these six months. We often find significantly lower default rates on the first and second statements for a firm in our own modeling on the data from bank databases (Dwyer 2005). When part of the sample of historical financial statements lacks any chance of defaulting, the result is a sample default rate lower than the population default rate.

The window for default chosen for a study may also omit many defaults. For example, suppose one defined a one-year default as a default within 3–15 months of a financial statement. This definition omits defaults that did not occur within 15 months of the last financial statement. We find that many private firm defaulters do not deliver their last financial statement. For example, suppose a firm's fiscal year ends in December and that they defaulted in May of 2005. They probably would have delivered their December 2003 financial statement in February or March of 2004. There is a good chance, however, that they would not have delivered their December 2004 financial statement in February or March of 2005 because this would have been about the time that they stopped making payments on their debt. In the Moody's KMV Credit Research Database (CRD), 25% of the defaults do not have financial statements within 15 months of their default date in North America. This type of sample selection bias will also result in the sample default rate being lower than the population default rate.

A third form of sample selection bias can result from including observations for which the default window extends past the end of the data collection period. For example, consider a financial statement dated 31 December 2004. For such a statement, a 15-month default window would end on 31 March 2006. Suppose that the data was gathered in September of 2005. Then firms that defaulted between October of 2005 and March of 2006 would not yet have been captured as having

---

<sup>16</sup>Of course, the extent to which this practice is implemented will vary both across financial institutions and between business units within each financial institution.

defaulted in the database. The technical term for this issue is *right-hand censored* data (cf, Greene (2000, Ch. 20.3)). This type of sample selection bias will also result in the sample default rate being lower than the population default rate.

## 4.2 Data collection issues

All financial institutions are in the process of implementing systems to capture defaults events. Default events are typically captured in a different system than financial statements, there are often problems in linking defaults to financial statements. Active, performing obligors are more likely to have historical data maintained in bank systems than inactive, defaulted borrowers. Post-merger systems integration will often lead to database problems. The net effect is often an under-representation of defaults (cf, Dwyer and Stein (2005; 2006)).

## 4.3 Default definition issues

Basel II provides a single default definition that is intended to be the world-wide standard for default. This standard can be characterized as 90 days past due and everything else that could be considered a default or a bankruptcy. Achieving uniform application of this standard is taking some time. Many countries have only recently begun collecting 90 days past due information. Consequently, most data sets that collect defaults over a long-time period do not consistently apply the Basel definition of default throughout the data collection process. This will tend to bias downward the default rates observed in a sample relative to the Basel definition.

## 4.4 Implications of data issues for model building and validation

Private firm default risk models are often not calibrated to exactly match the average default rate in the development data set. In fact, many modelers chose to set the average default rate in the model above the actual default rate in the model development sample. This difference is intended to adjust for the extent to which the data set does not capture all default events and the extent to which the data set does not reflect a full credit cycle. Furthermore, adjustments may be made to account for differences in how the definition of default is defined in the database and the targeted definition of default. Often the target definition is either the expectation of a real credit loss or the Basel Definition of default or both. The chosen average default rate for a model, often referred to as the central default tendency, is validated relative to a wide variety of available sources for overall reasonableness. Basel II explicitly recognizes these issues (cf, paragraphs 417, 448 and 462 of Basel Committee on Banking Supervision (2004)).

In validating a model, one needs to take into account the same issues to the extent possible when comparing the realized default rates to the actual default rate: Is the coverage of default events consistent and complete? What is the definition of default? What is the stage of the credit cycle?

## 5 CONCLUSION: IMPLICATIONS FOR MODEL BUILDING, CALIBRATING AND VALIDATING

A correct model for single-firm default risk will over-predict defaults most of the time. There are three reasons for this phenomenon. First, low PDs have skewed distributions in small samples. Second, correlation of negative events will skew the distribution of defaults as well. Third, data issues will generally understate the realized default rate on many samples.

These observations have a number of implications for building and calibrating models. For example, the fact that a correct model should over-predict defaults in most years needs to be taken into account during the model building process. By maximizing the log-likelihood of a model on a particular sample, one fits the average PD of the model to that of the development sample.<sup>17</sup> If one does so, on a sample that does not contain a recession, then the average PD produced by the model will be too low. Table 2 can be interpreted to give a sense of how large this effect could be. If one were to use 4,000 observations across four years to build a model with a correlation of 0.2, then the median default rate on such a sample will be 20% below the true default rate and the 5th percentile is 80% smaller than the true default rate.<sup>18</sup> Therefore, 50% of the time one would understate default risk by more than 20% and 5% of the time one would understate default risk by 80%.

If one observes a model that consistently under-predicts defaults, then one can conclude that the PDs are too low with some degree of confidence. Except for times during a recession, a standalone PD model will over-predict defaults most of the time. Under certain circumstances one can show that the PDs produced by a model are too high. For example, a model that consistently over-predicts defaults during recessions is likely to be too conservative. Because of the skewness of the default distribution, however, it is easier to show that the PDs are too low than too high. By bringing knowledge of the economic conditions facing the portfolio, this approach allows one to determine that a calibration needs to be revisited even when the default rate is within the 95% confidence level computed under the standard approach that allows for correlation.

When applying the framework to actual data, one must be cognizant of the data issues involved and consider the following questions: How complete is the data set? Does it capture all defaults? Has the default definition changed over time? What is the stage of the credit cycle? One rarely has access to a large data set that

---

<sup>17</sup>For example, one can show that in a logistic model an implication of the first-order conditions is that the average default rate will equal the average of the predicted probability of default (cf, Maddala (1983, p. 26)).

<sup>18</sup>This interpretation should be thought of as follows. Suppose one had many samples that are statistically independent in terms of both the aggregate and idiosyncratic shocks. If one were to build many different models across these samples by maximizing a likelihood function that does not allow for correlation, then 50% of the time the average PD produced by the model would be 20% below the true value and 5% of the time the average PD produced by the model would be 80% below the true value.

spans more than one credit cycle, that relies on a consistent definition of default, contains uniform accounting practices and utilizes consistent and complete data collection mechanisms. Owing to the scarcity of such data, there is significant intrinsic uncertainty regarding the actual level of risk associated with borrowers. Consequently, level validation of PD models will continue to be a challenging and interesting exercise for some time to come.

Appendix D presents an application of the technique using two long time series of default rates (speculative grade and investment grade issuers as rated by Moody's Investors Service). With a long time series, this framework can be extended to actually calibrate the correlation assumption to be consistent with the variability of default rates observed in the time series. For example, if one observes a relatively constant default rate across a large number of homogenous exposures over time then such data is consistent with a low degree of correlation. In contrast, if the default rate changes markedly over time, this is evidence of substantial correlation.<sup>19</sup> In Appendix D, we estimate that speculative grade bonds exhibit a correlation of 11% while investment grade bonds exhibit a correlation of 10%. Furthermore, we show that this technique could be used to reject the hypothesis of the PD of speculative grade defaults being 1.4% using only 2004 or 2005 data even though the  $p$ -value of the sample default rate observed in these years given this hypothesis is less than 90%. The reason is that the observed default rates in 2004 and 2005 would have implied a negative aggregate shock which is inconsistent with economic conditions during these years.

## APPENDIX A RECIPE FOR COMPUTING THE DISTRIBUTION OF DEFAULTS

Given  $N$ , PD,  $\rho$ , NumSims:

1. draw  $r_m$  from a standard normal distribution;
2. draw  $e_j$  from a standard normal distribution for each exposure;
3. compute  $\sqrt{\rho} r_m + \sqrt{1-\rho} e_j - \Phi^{-1}(\text{PD})$  for each exposure and flag the negative values as defaulters;
4. sum up the number of defaulters and record;
5. repeat NumSims times;
6. divide the number of defaults for each simulation by  $N$  to convert to a default rate (if desired); and
7. compute the P5, P10, median, and mean of the number of defaults or the default rate as desired across the simulations.

---

<sup>19</sup>Moody's KMV has used this approach to estimate correlations for retail exposures. Other examples include the work of Cantor and Falkenstein (2001) and Duffie *et al* (2006). While neither of these two explicitly uses a Gaussian one-factor framework, both estimate parameters associated with an underlying latent variable. Examples that do use a Gaussian one-factor model include the work of Aguais *et al* (2007), Jakubík (2006) and Rösch (2003).

**A.1 SAS code**

```

%let rho = 0.2;
%let pd = .01;
%let Exposures = 1000;
%let NumSims = 10000;

data test;
  do Sims =1 to &NumSims;
    default=0;
    rm = normal(0);
    do n = 1 to &exposures;
      ei = normal(0);
      dd= &rho**.5*rm + (1-&rho)**.5*ei;
      default = default + (dd<probit(&PD))/&exposures;
    end;
  output;
end;

proc univariate data=test ; var default ; run;

```

**APPENDIX B WHAT REPRESENTS A REASONABLE VALUE OF CORRELATION?**

In the Moody’s KMV Portfolio Manager Product, correlations range from 0.1 to 0.65. For private firms, we provide a model to predict the correlation of private firms on the basis of their country, industry, and size. Nevertheless, these correlations are designed to work in the context of a multi-factor model. The relationship between the correlation that is appropriate for a multi-factor model and that of a single-factor model is unclear. We do find that correlation increases with the size of the firm.

Figure B.1 presents the relationship between correlation and probability of default as prescribed by paragraphs 272 and 273 of Basel Committee on Banking Supervision (2004) for different firm sizes.

Basel II requires a formula for correlation in computing regulatory capital that is based on industry practices. This formula could be employed as a reasonable starting point. Paragraphs 272 and 273 of Basel Committee on Banking Supervision (2004) provide the following formula:

if  $S > 50$  then

$$\text{Correlation}(R) = 0.12(\text{weight}) + 0.24(1 - \text{weight})$$

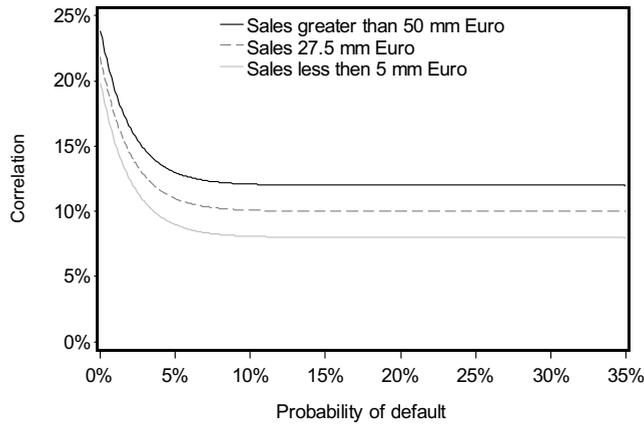
if  $S \in [5, 50]$

$$\text{Correlation}(R) = 0.12(\text{weight}) + 0.24(1 - \text{weight}) - 0.04(1 - (S - 5)/45)$$

if  $S \leq 5$

$$\text{Correlation}(R) = 0.12(\text{weight}) + 0.24(1 - \text{weight}) - 0.04$$

where  $\text{weight} = (1 - \exp(-50*PD))/(1 - \exp(-50))$  and  $S$  is total annual sales in millions of Euros.

**FIGURE B.1** Correlation and PD from Basel II.

This formula implies that a large firm's correlation is bounded between 0.12 and 0.24 and that the smallest firm is bounded between 0.08 and 0.20 (see Figure B.1).

### APPENDIX C AN ALTERNATIVE TO THE UNIFORM DISTRIBUTION ASSUMPTION IN THE CASE OF NO DEFAULTS WITH ZERO CORRELATION

In this section, we show that, for a prior distribution on the PD that is very different from the uniform, the posterior distribution is very similar for a large sample size given no correlation.

One method for establishing an upper bound for a PD in the case of zero defaults and no correlation is based on an answer to the following question: What is the smallest PD such that the probability of observing zero defaults is 5%?<sup>20</sup> The answer to this question is solving  $0.05 = (1 - \text{PD})^N$  for  $N$ .<sup>21</sup> Note the similarity between this solution and the solution to the 95th percentile of the posterior distribution given a uniform distribution:  $0.05 = (1 - \text{PD})^{N+1}$ . We can give the answer to this question a Bayesian interpretation, by finding a prior distribution that would yield a posterior distribution consistent with this answer.

<sup>20</sup>Pluto and Tasche (2005) take such an approach.

<sup>21</sup>Solving this equation yields one form of the so-called *rule of threes*. Taking logs of both sides and using the facts that  $\log(0.05) \approx -3$  and  $\log(1 - \text{PD}) \approx -\text{PD}$  yields  $\text{PD} \approx 3/N$ , ie, that the minimum PD is three divided by the sample size. Put differently, suppose a portfolio has  $3/\text{PD}$  exposures and that each have a probability of default of PD. Then one can be 95% certain that the portfolio will have at least one default under the assumption of independence.

Let the prior distribution of PD be given by

$$f(\text{PD}) \begin{cases} \frac{\phi}{(1 - \text{PD})} & \text{on } [0, \phi) \\ 0 & \text{otherwise} \end{cases}$$

where  $\phi < 1$  and is  $\phi$  is  $-(\log(1 - \phi))^{-1}$ . This distribution places a larger weight on the larger PDs. In fact, the distribution is only well defined if  $\phi < 1$ , as otherwise the density would approach infinity as PD approaches 1. The constant,  $\phi$ , simply ensures that the density function integrates to one.

For this prior, the posterior distribution of PD can be derived as

$$k(\text{PD}|\text{SDR} = 0) = \frac{\int_0^{\text{PD}} (1 - w)^{N-1} dw}{\int_0^{\phi} (1 - w)^{N-1} dw} = \frac{1 - (1 - \text{PD})^N}{1 - (1 - \phi)^N}$$

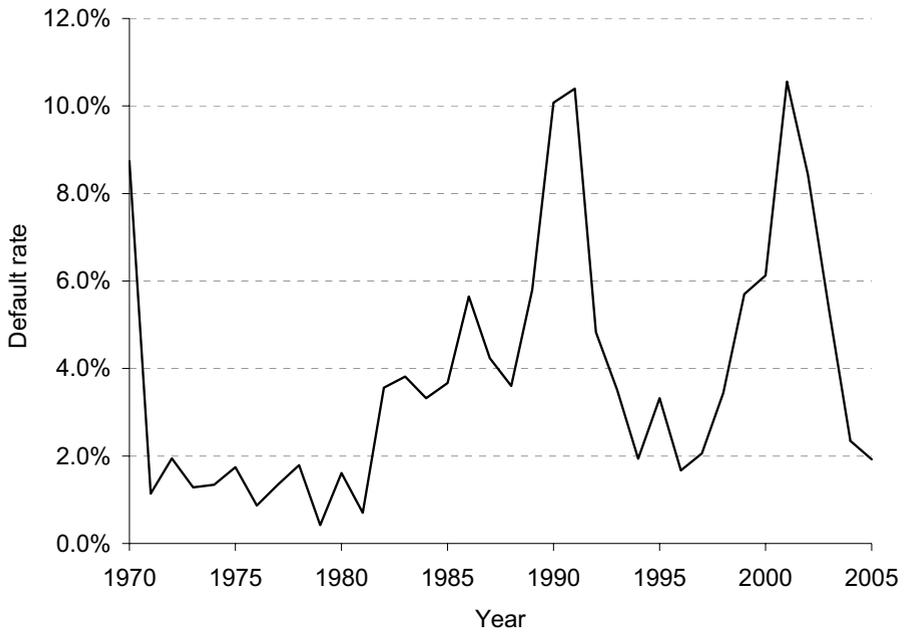
Clearly, the limit of the posterior as  $\phi$  approaches 1 is  $1 - (1 - \text{PD})^N$  and the 95th percentile of this distribution is determined by the solution to  $0.05 = (1 - \text{PD})^N$ . Therefore, a rather significant change in the prior distribution has only a modest change in the posterior for a large  $N$ .

The Bayesian approach also allows one to state that we are 95% certain that the true value is in the range approximated by  $[0, 3/N]$  given the prior. Under the classical approach, one can state that any null hypothesis that the PD was not in this region would have been rejected with 95% certainty (cf, footnote 1 of Pluto and Tasche (2005)). Nevertheless, this region is not a confidence interval in the sense of classical statistics. Under classical statistics, the confidence interval is random while the true value is fixed. A 95% confidence interval has the interpretation that the probability that the confidence interval contains the true value is 0.95. The difficulty of constructing such an interval in the case of a binomial distribution results from the variance of the distribution changing with the mean of the distribution (cf, Section 6.3 of Hogg and Craig (1978)).

### APPENDIX D AN APPLICATION TO INVESTMENT AND SPECULATIVE GRADE DEFAULT RATES

The default studies of Moody’s Investors Service provide a long time series that one can use to test the model. For simplicity, we will break the series into two categories: investment and speculative grade issuers. The long time series enables us to set the PD equal to the average of the default rate time series and to calibrate the correlation assumption to the observed standard deviation of the default rate time series.

Figure 1 and Figure D.1 present the time series of default rates for investment and speculative-grade issuers as rated by Moody’s Investors Service for a long

**FIGURE D.1** The default rate of speculative grade issuers 1970–2005.

Source: Exhibit 30 of Hamilton and Varma (2006).

time series (1970–2005).<sup>22</sup> The second and fifth columns of Table D.1 provide statistics and quantiles regarding these two time series. Given a PD and a correlation assumption one can simulate this time series by drawing an aggregate shock for each year and an idiosyncratic shock for each issuer in each year. One then computes the time series of the default rate as well as the average, standard deviation, 25th percentile, median, 75th percentile, 90th percentile and maximum for this time series and records these values. One can repeat this exercise 10,000 times and take the average of each statistic and quantile across the simulations.

One can also take the standard deviation of these values to determine the extent of sampling variability that is in the data. For these simulations, we chose the PD to match the time series average of the realized default rate and chose the correlation parameter to match the standard deviation of the realized default rate. Our algorithm is to perform the simulations for each value

<sup>22</sup>For both of these time series the number of rated issuers is growing. Investment grade issuers grew from 738 to 3,390 and speculative grade issuers grew from 286 to 1,559. The counts for the speculative grade and investment issuers are computed from Exhibit 27 and Exhibit 30 of Hamilton and Varma (2006).

**TABLE D.1** Distribution of observed versus modeled default rates for investment and speculative grade issuers. A comparison of the distribution of the observed time series of default rates from 1970–2005 to the distribution of default rates observed across 10,000 simulations.

	Investment grade			Speculative grade		
	Realized	Average across sims	Std dev across sims	Realized	Average across sims	Std dev across sims
Average	0.068%	0.068%	0.021%	3.8%	3.8%	0.49%
Std dev	0.118%	0.118%	0.049%	2.8%	2.9%	0.61%
P25	0.000%	0.000%	0.000%	1.7%	1.8%	0.36%
Median	0.000%	0.011%	0.019%	3.3%	3.1%	0.51%
P75	0.102%	0.097%	0.028%	5.4%	5.2%	0.80%
P90	0.271%	0.204%	0.068%	8.7%	7.8%	1.41%
P95	0.318%	0.331%	0.134%	10.4%	10.1%	2.19%
Max	0.485%	0.526%	0.287%	10.5%	12.6%	3.40%

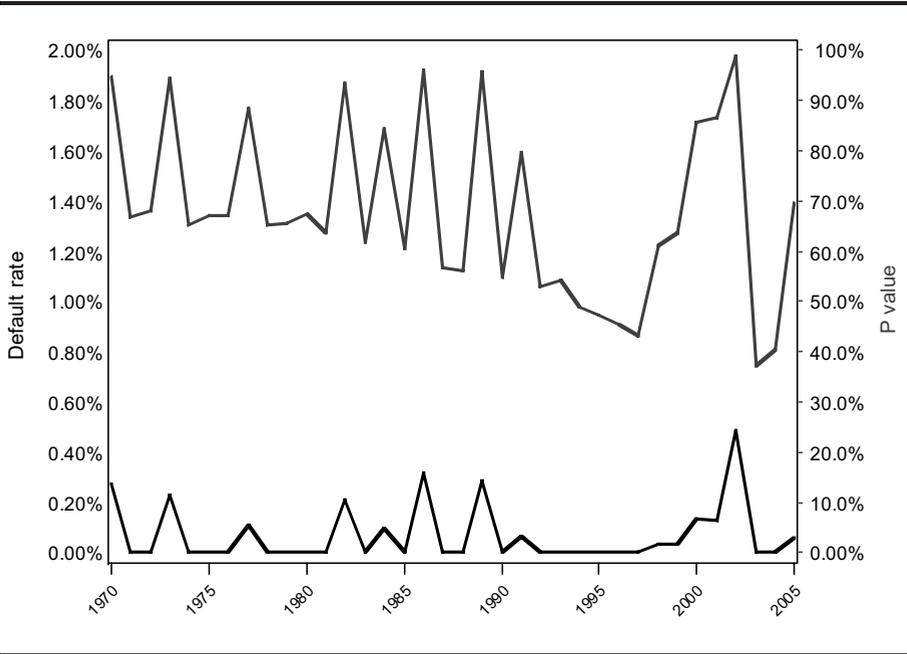
of  $\rho \in \{0, 0.01, 0.02, \dots, 0.2\}$  and to then chose the  $\rho$  that minimizes the square of the difference between the average of the standard deviation of the time series default rate and the standard deviation of the realized time series default rate. We present the results of such an analysis for a PD and correlation assumption of  $\{0.068\%, 0.11\}$  and  $\{3.839\%, 0.10\}$  for investment and speculative-grade issuers, respectively.

This very simple model fits the general properties of the time series rather well.<sup>23</sup> The second and fifth columns present the realized statistics and quantiles for the actual time series. The third and sixth columns present the average of these values across the 10,000 simulations. The fourth and seventh columns present the standard deviation of the values across the simulations. The realized 25th, median, 75th, 90th and 95th percentiles are within the sampling variability that one would expect given the model. The largest departure is the realized 90th percentile for investment grade which is approximately one standard deviation away from the average of the simulations. The worst default rate observations in the realized 36-year time period are lower than the average value across the simulations.

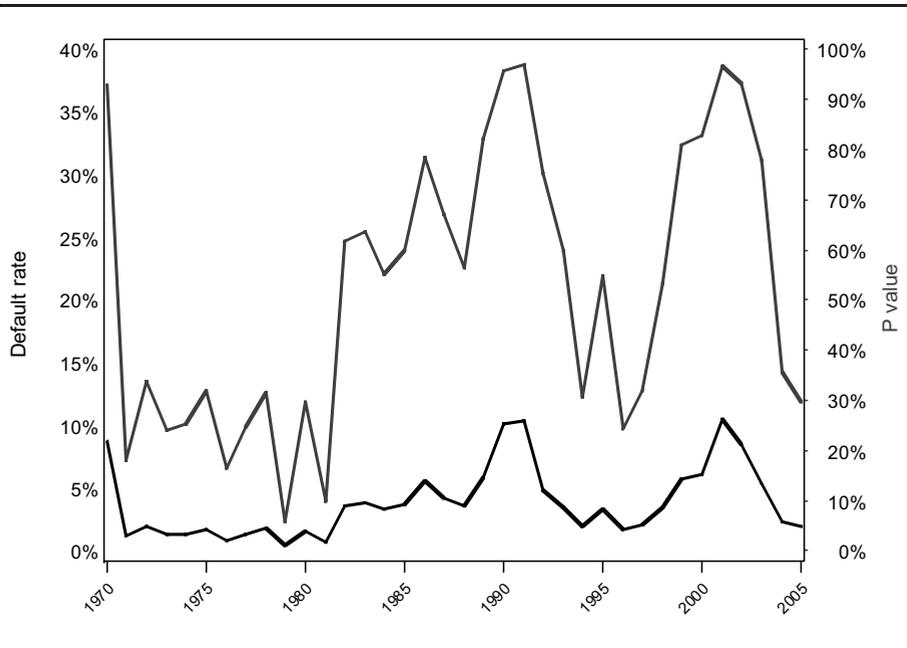
Figures D.2 and D.3 present the time series of the default rates and the corresponding  $p$ -values of the observed default rates for these series using the same PD and correlation assumptions that were used in Table D.1 and the

<sup>23</sup>For purposes of this simple analysis, we assume that the aggregate shocks are independent between time periods. This assumption implies zero correlation between the current and prior year default rate. In the case of the investment grade default rate series we can not reject the hypothesis of zero correlation. For the speculative default rate series, however, there is substantial positive correlation (0.65). The model could be extended to reflect this feature of the data by incorporating some autocorrelation into the aggregate shock (eg, a first order autoregressive process).

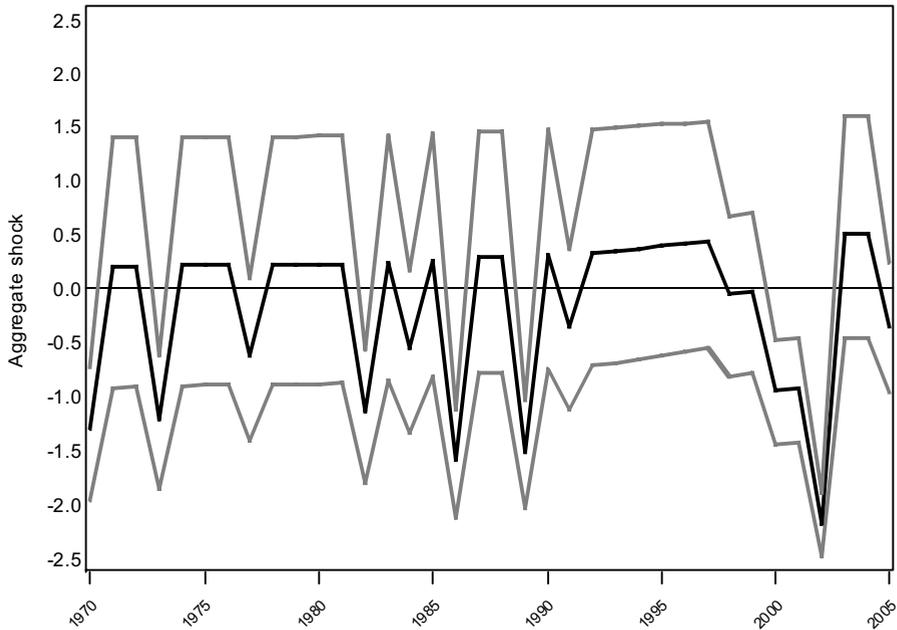
**FIGURE D.2** Investment grade default rate and corresponding  $p$ -value.



**FIGURE D.3** Speculative grade default rate and corresponding  $p$ -value.



**FIGURE D.4** Posterior distribution of the aggregate shock for investment grade credits. The 10th, 50th and 90th percentiles of the posterior distribution for the aggregate shock for investment grade issuers given the observed default rate and the assumptions that the PD and correlation are 0.068% and 11%, respectively.

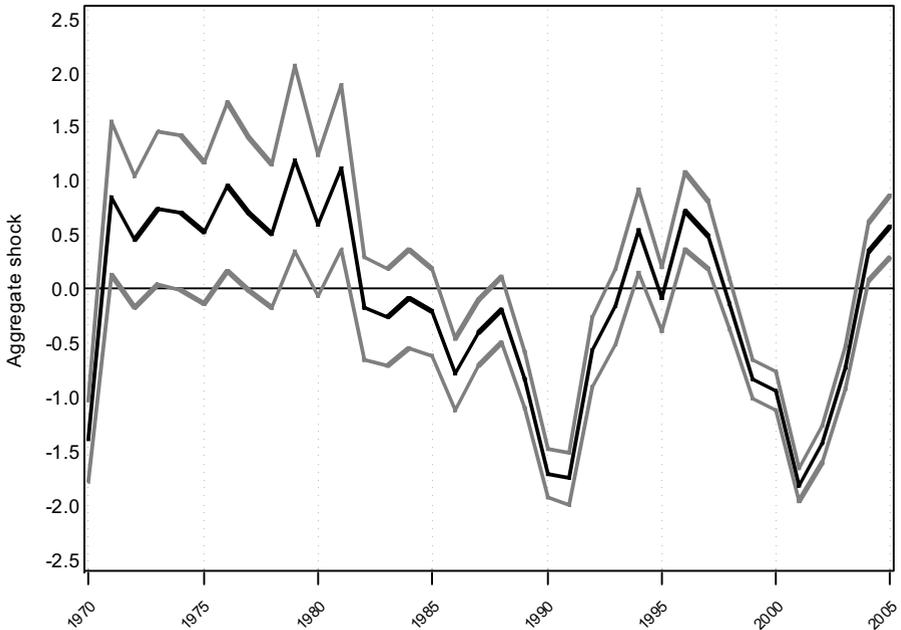


methodology in Section 2. For the investment grade series the  $p$ -values exceed the 95th percentile in 1986, 1989 and 2002. The years 1989 and 2002 correspond to recessionary periods in the economy. Similarly, for the speculative grade series the  $p$ -value exceeds the 95th percentile in 1990, 1991 and 2001, which also correspond to recessionary periods.

Figures D.4 and D.5 present the 10th, 50th and 90th percentiles for the posterior distribution of the aggregate shock for the investment grade and speculative grade default rates computed using the methodology in Section 3.2. For the speculative grade series there were large negative shocks in 1970, 1990, 1991, 2000 and 2001. These time periods correspond to recessions with the exception of 1970. During 1970, Penn Central Railroad and 25 of its affiliates defaulted, which explains the spike. For the investment grade series, the negative aggregate shocks are likely to have occurred in 1970, 1973, 1982, 1986, 1989, 2000, 2001 and 2002. All of these years correspond to recessionary periods with the exception of 1970 and 1986, with 1970 being explained by Penn Central Railroad.

With the benefit of a 36-year time series, one can calibrate the model well and the data appears consistent with the implications of this simple model. Often banks

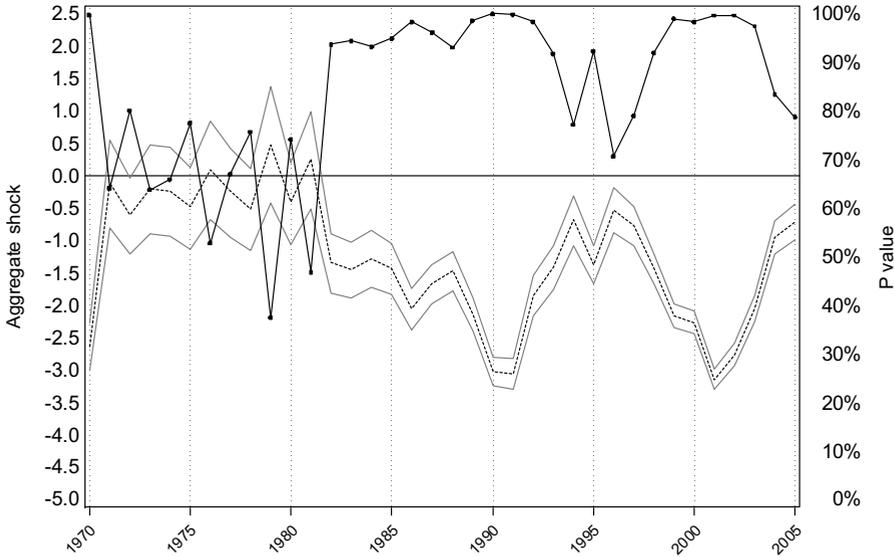
**FIGURE D.5** Posterior distribution of aggregate shock for speculative grade credits. The 10th, 50th and 90th percentiles of the posterior distribution for the aggregate shock for speculative grade issuers given the observed default rate and the assumptions that the PD and correlation are 3.8% and 10%, respectively.



will have a very limited time series of default rates to evaluate their model on due to changing business practices. Can this model show that a calibration is wrong? For example, suppose that one claimed that the default rate for speculative grade bonds was 1.4% and one only had access to 2004 and 2005 data to test this hypothesis. In these years the actual default rate was 2.3% and 1.9%, respectively. Figure D.6 presents the  $p$ -value and the 10th, 50th and 90th percentiles for the distribution of the aggregate shock for the entire 36-year time series given a PD of 1.4% and a correlation of 10%.

Looking only at 2004 and 2005, the  $p$ -values for a PD of 1.4% are 84% and 79%, respectively. Therefore, the standard methodology would not allow you to reject the hypothesis of a 1.4% PD. Nevertheless, if the PD really were 1.4% then there would have been a negative aggregate shock in both of these years – the 90th percentile for the distribution of the aggregate shock is less than zero in both cases. This implication is counterfactual; these years have been characterized as expansionary in the world economies. Therefore, one could conclude that the calibration needed to be revisited even though the  $p$ -value of the realized default rate was less than 95% in both years according to the traditional approach.

**FIGURE D.6** The  $p$ -value and posterior distribution for speculative grade default rates given PD = 1.4%. The grey lines present the 10th, 50th and 90th percentiles of the posterior distribution for the aggregate shock for speculative grade issuers given the observed default rate and the assumptions that the PD and correlation are 1.4% and 10%, respectively. The black line at the top, which is read off the right axis, provides the  $p$ -value of the observed default rate in each year.



**REFERENCES**

Aguais, S., Forest, L., Wong, E., and Diaz-Ledezma, D. (2007). Point in time versus through-the-cycle ratings. *The Basel Handbook: A Guide for Financial Practitioners*, Ong, M. (ed). Riskbooks, London, pp. 183–210.

Basel Committee on Banking Supervision (2004). International convergence of capital measurement and capital standards ('A Revised Framework'). Bank for International Settlements.

Basel Committee on Banking Supervision (2005). An Explanatory Note on the Basel II IRB Risk Weight Functions. Bank for International Settlements.

Cantor, R., and Falkenstein, E. (2001). Testing for Rating Consistency in Annual Default Rates. Moody's Investors Service.

Christensen, J. H. E., Hanson, E., and Lando, D. (2004). Confidence sets for continuous-time rating transition probabilities. *Journal of Banking and Finance* **28**, 2575–2602.

Duffie, D., Eckner, A., Horel, G., and Saita, L. (2006). Frailty correlated default. Stanford University.

Dwyer, D., and Stein, R. (2005). Moody's KMV RiskCalc™ 3.1 Technical Document. Moody's KMV.

- Dwyer, D. (2005). Examples of overfitting encountered when building private firm default prediction models. Moody's KMV.
- Dwyer, D., and Stein, R. M. (2006). Inferring the default rate in a population by comparing two incomplete default databases. *Journal of Banking and Finance* **30**, 797–810.
- Embrechts, P., McNeil, A., and Straumann, D. (2001). Correlation and dependence in risk management: properties and pitfalls. *Risk Management: Value at Risk and Beyond*, Dempster, M. (ed). Cambridge University Press, pp. 76–233.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). Bayesian data analysis. Chapman & Hall/CRC, New York.
- Greene, W. H. (2000). *Econometric Analysis*, 4th edn. Prentice-Hall, Upper Saddle River, NJ.
- Hamilton, D., and Varma, P. (2006). Default and recovery rates of corporate bond issuers, 1920–2005. Moody's Investors Service.
- Hanson, S., and Schuermann, T. (2006). Confidence intervals for probabilities of default. *Journal of Banking & Finance* **30**(8), 2281–2301.
- Hogg, R. V., and Craig, A. T. (1978). *Introduction to Mathematical Statistics*, 4th edn. Macmillan.
- Hull, J., and White, A. (2004). Valuation of a CDO and an  $n$ th to default CDS without Monte Carlo simulation. *Journal of Derivatives* **12**(2), 8–23.
- Jakubík, P. (2006). Does credit risk vary with economic cycles? The Case of Finland. Charles University, Prague.
- Kurbat, M., Korablev, I. (2002). Methodology for testing the level of the EDF<sup>TM</sup> credit measure. Moody's KMV.
- Maddala, G. S. (1983). *Limited Dependent and Qualitative Variables in Econometrics (Econometric Society Monographs)*. Cambridge University Press, Cambridge.
- Pluto, K., and Tasche, D. (2005). Estimating probabilities for low default portfolios. Deutsche Bundesbank.
- Rösch, D. (2003). Correlations and business cycles of credit risk: evidence from bankruptcies in Germany. University of Regensburg.
- Stein, R. (2006). Are the probabilities right? A first approximation to the lower bound on the number of observations required to test for default rate accuracy. *Journal of Investment Management* **4**(2), 61–71.
- Tasche, D. (2003). A traffic lights approach to PD validation. Deutsche Bundesbank.
- Tasche, D. (2005). Rating and probability of default validation. Studies on the Validation of Internal Rating Systems, Working Paper No. 14, Basel Committee on Banking Supervision, Bank for International Settlements.
- Vasicek, O. (1987). Probability of loss on loan portfolio. Moody's KMV.
- Vasicek, O. (1991). Limiting loan loss probability distribution. Moody's KMV.