View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Research Papers in Economics

ISSN 1936-5098

CAE Working Paper #09-01

The Maximum Entropy Distribution for Stochastically Ordered Random Variables with Fixed Marginals

by

Nicholas M. Kiefer

January 2009

The Maximum Entropy Distribution for Stochastically Ordered Random Variables with Fixed Marginals

Nicholas M. Kiefer*

January 23, 2009

Abstract

Stochastically ordered random variables with given marginal distributions are combined into a joint distribution preserving the ordering and the marginals using a maximum entropy formulation. A closed-form expression is obtained. An application is in default estimation for different portfolio segments, where priors on the individual default probabilities are available and the stochastic ordering is agreeable to separate experts. The ME formulation allows an efficiency improvement over separate analyses.

Keywords: Bayesian analysis, stochastic dominance, joint distributions, risk management, defaults, small probability estimation

1 Introduction

Consider random variables X and Y with marginal distributions $F(x)$ and $G(y)$ and with the stochastic ordering restriction $X \geq Y$ $a.s.$ We seek a joint distribution $H(x, y)$ with marginals $F(x)$ and $G(y)$, with $X \geq Y$ $a.s.$, and with minimal additional information represented. The problem arises in risk management, where X and Y might be random default rates for different portfolio segments with Y

*Cornell University Departments of Economics and Statistical Sciences, 490 Uris Hall, Ithaca, NY 14853-7601, US. email:nmk1@cornell.edu; US Department of the Treasury Office of the Comptroller of the Currency, Risk Analysis Division, 250 E. Street, SW, DC 20219; and CREATES, Aarhus University, Building 1322, DK-8000 Aarhus C, Denmark.

consisting of safer assets. In one application the distributions F and G are priors for default rates for different portfolio segments. These are assessed from different experts, but the experts can agree on the stochastic ordering. A maximum entropy approach to obtaining a joint distribution is proposed. The motivation is that information on default rates in one of the portfolio segments is relevant for default rates in the other, even if the defaults are independent, due to dependence in the joint prior induced by the stochastic ordering. The experts are willing to accept each other's assessment of the marginal prior for the relevant portfolio segment - perhaps due to an Aumann argument (each knows the other is an expert and is willing to accept his processing of information), neither has good information about the dependence, but they are in agreement on the ordering. The maximum entropy approach specifies the joint distribution with the required margins and the stochastic ordering and as little additional structure as possible.

2 Bivariate Distributions with Given Margins

A necessary and sufficient condition for the existence of a bivariate distribution $H(X,Y)$ with marginal distributions $F(X), G(Y)$ and with $X \geq Y$ $a.s.$ is that the distribution function G first-order stochastically dominates F, i.e. $G(z) \geq F(z)\ \forall z$. The necessity is clear. For sufficiency, let $y^*(x) = \max_y\{y|G(y) = F(x)\}$ and let

$$H(X,Y) = \left\{ \begin{array}{ll} G(Y) & for Y < y^*(X) \\ G(y^*(X)) & otherwise \end{array} \right\}.$$

This is not a very interesting distribution but it establishes existence of the required joint distribution. This condition is well known, see Strassen (1965), Theorem 11, or Kamae, Krengel, and O'Brien (1977), Theorem 1. We now assume strict first-order stochastic dominance, i.e. $G(z) > F(z)$ for $z \in (0,1)$. If equality holds in an interval then the variables are functionally related in an interval and this is not an interesting case. If necessary, the analysis could be carried out separately for intervals in which strict dominance holds.

3 The Maximum Entropy Joint Distribution

Let $h(x,y)$ be the desired maximum-entropy (ME) joint density and let K constraints indexed by k be given by $Ec_k(x,y) = 0$. Entropy is a measure of uncertainty, so we seek the distribution that satisfies the constraints but otherwise has minimal information. See Jaynes (2003) for a discussion by an advocate of the ME approach in many settings. The (differential) entropy constrained maximization problem is

$$\max_h \{-\int h(x,y)\ln(h(x,y))dxdy\}$$

$$s.t. \int\int h(x,y)c_k(x,y)dxdy = 0 \; for \; k = 1, ..., K$$
$$and \int\int h(x,y)dxdy = 1$$

With Lagrange multipliers λ_k and μ the FOC is (differentiating purely formally with respect to $h(x,y)$)

$$-\ln(h(x,y)) - 1 + \sum_k \lambda_k c_k(x,y) + \mu = 0$$

and solving for the density at (x,y)

$$h(x,y) = \exp\{-1 + \sum_k \lambda_k c_k(x,y) + \mu\}$$

This is a result associated with Boltzmann. In our setting the constraints on the marginals are:

$$\int_0^1 \int_0^x (I(x \le \alpha) - F(\alpha))h(x,y)dydx = 0$$
$$\int_0^1 \int_0^x (I(y \le \beta) - G(\beta))h(x,y)dydx = 0$$

With Lagrange multipliers (functions) $\lambda(\alpha)$ and $\eta(\beta)$ the density takes the form

$$h(x,y) = c \times I(x \ge y)\exp(\int_0^1 \lambda(\alpha)I(x \le \alpha)d\alpha + \int_0^1 \mu(\beta)I(y \le \beta)d\beta) \quad (1)$$
$$= c \times I(x \ge y)a(x)b(y)$$

where

$$
\begin{aligned}
c^{-1} &= \int_0^1 \int_0^x \exp(\int_0^1 \lambda(\alpha) I(x \leq \alpha) d\alpha + \int_0^1 \mu(\beta) I(y \leq \beta) d\beta) dx dy \\
&= \int_0^1 \int_0^x a(x) b(y) dy dx
\end{aligned}
$$

incorporates the adding up constraint and absorbs the constants $\int_0^1 \lambda(\alpha)F(\alpha)d\alpha$ and $\int_0^1 \mu(\beta)G(\beta)d\beta$. The key is that the density factors into a function of x and a function of y. These functions must be such that the constraints are satisfied. Let $f(x)$, and $g(y)$ be the required marginal densities. Rewrite the constraints, dropping c from the notation - i.e., normalizing $a(x)b(y)$,

$$
\begin{aligned}
\int_0^1 I(x \geq y) a(x) b(y) dy &= f(x) \\
\int_0^1 I(x \geq y) a(x) b(y) dx &= g(y)
\end{aligned}
$$

we see that the number of unknowns (the function $a(x)b(y)$) is equal to the number of nonredundant constraints due to the factorization of $h(x, y)$ arising from the ME specification. The functions $a(x)$ and $b(y)$ are only identified separately up to scaling. From the discussion above we see that the constraints are inconsistent if $G(z) < F(z)$ for any z. The first constraint can be rewritten

$$
a(x) \int_0^x b(y) dy = a(x) B(x) = f(x)
$$

defining the function $B(x)$, and similarly $b(y) \int_y^1 a(x)dx = g(y) = b(y)A(y)$. Note that $a(x) = -A'(x)$ and $b(x) = B'(x)$. Thus consider the differential equations

$$
\begin{aligned}
-A'(x)B(x) &= f(x) \\
B'(x)A(x) &= g(x)
\end{aligned}
$$

Subtracting: $A'(x)B(x) + B'(x)A(x) = g(x) - f(x)$. Integrating: $A(x)B(x) = \int_0^x (g(u) - f(u))du = G(x) - F(x) + k$. Since $A(x)B(x) = 0$ for $x \in \{0, 1\}$ it is clear that $k = 0$. From $B'(x)A(x) = g(x)$, divide through by $A(x)B(x)$ for $x \in (0, 1)$ to obtain

$$
B'(x)/B(x) = d \ln B/dx = g(x)/(G(x) - F(x)).
$$

thus

$$\ln B = \int_{1/2}^{x} g(u)/(G(u) - F(u))du + k$$

Here the (different) integrating constant k will be determined by the normalization and we are free to choose a convenient point from which to calculate the antiderivative. Hence

$$B'(x) = g(x)/(G(x) - F(x)) \exp(\int_{1/2}^{x} g(u)/(G(u) - F(u))du + k)$$

and

$$A'(x) = -f(x)/(G(x) - F(x)) \exp(-\int_{1/2}^{x} f(u)/(G(u) - F(u))du + k)$$

Here k is a different integrating constant - both constants appear only in the normalizing constant c and can be set to zero here. Applying equation 1

$$\begin{aligned} h(x,y) &= c \times I(x \geq y) \times a(x)b(y) \\ &= \frac{c \times I(x \geq y) \times f(x)g(y)}{(G(x) - F(x))(G(y) - F(y))} \\ &\exp(\int_{1/2}^{y} g(u)/(G(u) - F(u))du \\ &- \int_{1/2}^{x} f(u)/(G(u) - F(u))du) \end{aligned} \tag{2}$$

and $c^{-1} = \int_0^1 \int_0^x a(x)b(y)dydx$. Thus we have a closed-form solution for the maximum-entropy distribution with given marginal distributions and stochastically ordered random variables.

4 Examples

We first consider a simple example admitting explicit calculation: $G(y) = y$; $F(x) = x^2$. Applying equation (2)

$$\begin{aligned} h(x,y) = c \times I(x \geq y)(2x/((x-x^2)(y-y^2)) \\ \exp(\int_{1/2}^{y} 1/(u-u^2)du \\ - \int_{1/2}^{x} 2u/(u-u^2)du. \end{aligned}$$

Integrating and simplifying

$$h(x,y) = c \times I(x \geq y) \times 8 \times (x-1)/(y-1)^2$$

and the normalization constant is

$$c^{-1} = \int_0^1 \int_0^x 8(1-x)/(1-y)^2 dxdy = 4$$

giving finally

$$h(x,y) = 2I(x \geq y)(1-x)/(1-y)^2$$

It is easy to verify that this joint density has the appropriate marginal densities. The density has differential entropy (using nats) $-1/2 - \ln 2$. In contrast the entropy of the maximum-entropy joint distribution without the ordering constraint is $1/2 - \ln 2$, so there is a substantial increase in information by using the constraint. The correlation is $\rho_{xy} = 0.82$. Of course the correlation is 0 without the ordering constraint. The probability that the ordering constraint is violated in the product joint (corresponding to separate analyses) is 1/3.

The second and perhaps more relevant example uses Beta distributions. Let $G(y)$ be the $Beta(1,3)$ distribution with pdf $g(y) = 3(1-y)^2$ and $F(x)$ be the $Beta(3,3)$ distribution with pdf $f(y) = 30(1-x)^2x^2$. This specification satisfies the necessary and sufficient condition for existence of a joint with the stochastic ordering $Y \leq X$ $a.s.$ Applying equation 2,

$$h(x,y) = \frac{c \times I(x \geq y)(90(1-y)^2(1-x)^2x^2)}{(3x-3x^2-9x^3+15x^4-6x^5)(3y-3y^2-9y^3+15y^4-6y^5)}$$
$$\exp(\int_{1/2}^{y} 3(1-u)^2/(3u-3u^2-9u^3+15u^4-6u^5)du$$
$$-\int_{1/2}^{x} 30(1-u)^2u^2/(3u-3u^2-9u^3+15u^4-6u^5)du.$$

After considerable calculation and using $\int_{1/2}^{x} p(u)du = -\int_{x}^{1/2} p(u)du$ to simplify calculations for $x < 1/2$ and computing the integrating constant we have

$$h(x,y) = \frac{I(x \geq y)(30(x-1)x((x-1)^2(1+2x))^{2/3}(1-y)^{-1/3})}{(y-1)(1+2y)^{5/3}}.$$

The correlation in the maximum entropy joint distribution is 0.55. The probability that the constraint is violated when the product distribution is used for the joint is 0.18. The entropy in the ME distribution is -1.06; in the product distribution is -0.70, so the increase in information by imposing the stochastic ordering is again substantial though less than with the previous example. Here, there is perhaps less to gain from a joint analysis than in the previous example, because the marginals are widely separated.

5 Application to Default Rates

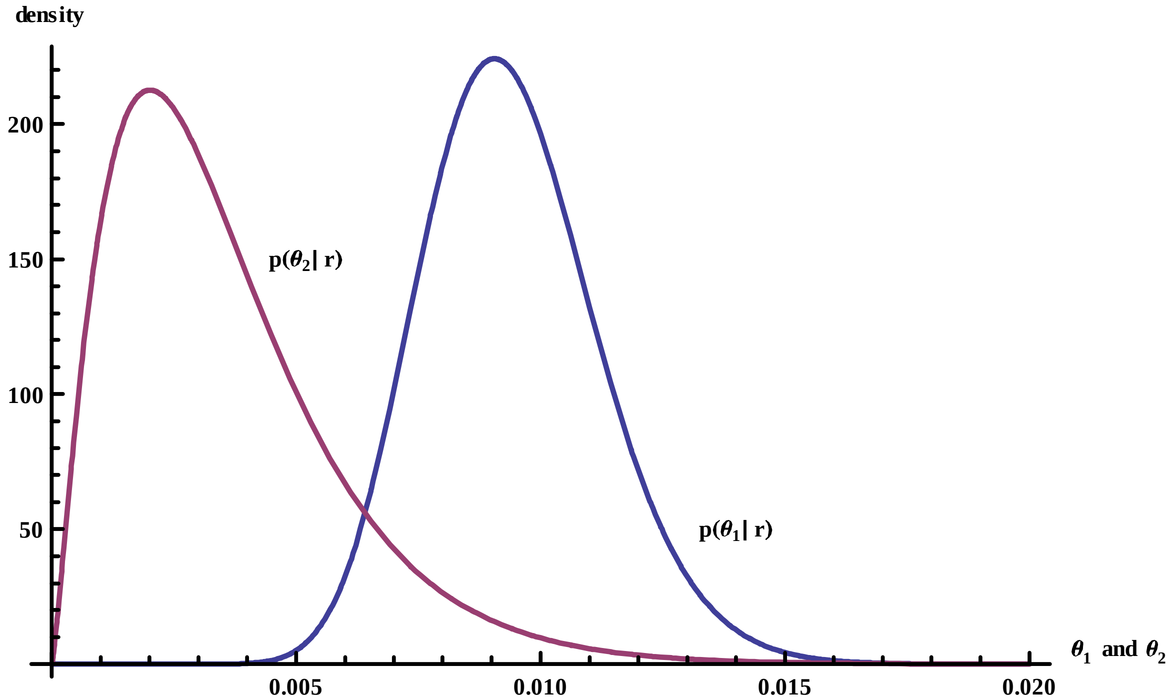
Estimation of default probabilities (PD) and other parameters - but perhaps default rates are the most crucial - for portfolio segments consisting of reasonably homogeneous assets is essential to prudent risk management. It is also crucial for compliance with Basel II (B2) rules for banks using the Internal Ratings Based approach to determine capital requirements (Basel Committee on Banking Supervision (2004)). This is the only approach approved in the US and is the approach expected of large banks in countries adopting the B2 rules. Typically default rates are estimated separately for different portfolio segments. The requirements demand an annual default probability, estimated over a sample long enough to cover a full cycle of economic conditions. It has been noted that for very safe assets, or for assets new to the market, calculations based on historical data may "not be sufficiently reliable" (Basel Committee on Banking Supervision (2005)) to form a probability of default estimate, since so few defaults are observed. Kiefer (2008) has proposed a Bayesian approach, relying on informative priors elicited from industry experts. That paper considered a portfolio of loans to highly-rated, large, internationally active and complex banks. The typical data set here would have perhaps 50 to 100 loans and no defaults. The approach was applied separately (Kiefer (2009)) to a segment the middle of the risk profile of the portfolio. Although the risk is in the middle of the asset mix, the probability of default is still "small." It is in fact likely to be about 0.01; defaults, though seen, are rare. The bulk of a typical bank's commercial loans are concentrated in these segments (segments differ across banks). Very low risk institutions are relatively few in number and they have access to capital through many avenues in addition to commercial loans. Very high risk loans are largely avoided and when present are often due to the reclassification of a safer loan as conditions change. To put this in perspective, the middle-quality loans are approximately S&P Baa or Moody's BBB. In practice the bulk of these loans are to unrated companies and the bank has done its own rating to assign the loans to risk "buckets." This assignment already indicates the availability of nontrivial prior or expert information. The elicitation method included a specification of the problem and some specific questions over e-mail followed by a discussion including feedback and revision. General discussions of the elicitation of prior distributions are given by O'Hagan, Buck, Daneshkhah, Eiser, Garthwaite, Jenkinson, Oakley, and Rakow (2006) and Kadane and Wolfson (1998); details on the elicitations underlying the present application are given in the cited papers.

Typical data consist of a number of asset/years for a group of similar assets. In each year there is either a default or not. This is a clear simplification of the actual problem in which asset quality can improve or deteriorate and assets are not completely homogeneous. Nevertheless, it is useful to model the problem as one of independent Bernoulli sampling with unknown parameter θ. Let d_i indicate whether the ith observation out of n was a default ($d_i = 1$) or not ($d_i = 0$), let $D = \{d_i, i = 1, ..., n\}$ denote the whole data set and $r = r(D) = \sum_i d_i$ the count of defaults. The statistic $r(D)$ is sufficient and has distribution

$$p(r|\theta) = \binom{n}{r}\theta^r(1-\theta)^{n-r} \tag{3}$$

The model can be elaborated but for illustration of the gains from a joint analysis we will use the simple specification, noting that it is widely used in practice. The

Figure 1: Posterior densities from separate analysis

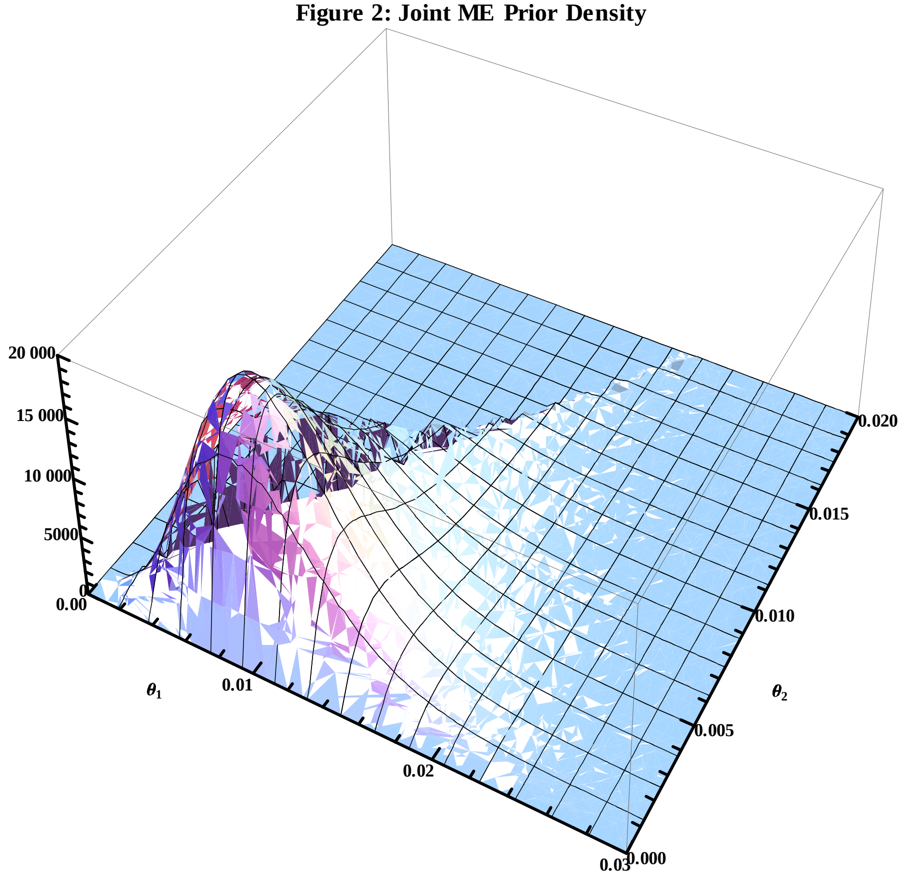


prior information can also be refined, but for illustration we fit Beta distributions to assessments given by the experts. For the mid-portfolio example with default rate θ_1 we have $\alpha = 6.8$ and $\beta = 647$ corresponding to a prior mean of 0.010 and a prior standard deviation of 0.004. For the lowest-risk portfolio with default rate θ_2 we have .$\alpha = 2.3$ and $\beta = 545$ corresponding to a prior mean of 0.0042 and a prior standard deviation of 0.0028. These distributions satisfy the necessary and sufficient conditions for the existence of a joint distribution with the stochastic ordering (stroke of luck here).

First, we report the posterior analysis for the two segments separately. For the mid-portfolio segment we use a bucket of mid-portfolio corporate bonds of S&P-rated firms in the KMV North American Non-Financial Dataset. Default rates were computed for 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. Details on the data are given in Kiefer (2009). The posterior distributions are shown in Figure 1.

The posterior summary statistics for the mid-portfolio segment are $E\theta_1$ =0.0094 and $sd(\theta_1)$ =0.0018 . For the low-default portfolio we report results with a typical

Figure 2: Joint ME Prior Density



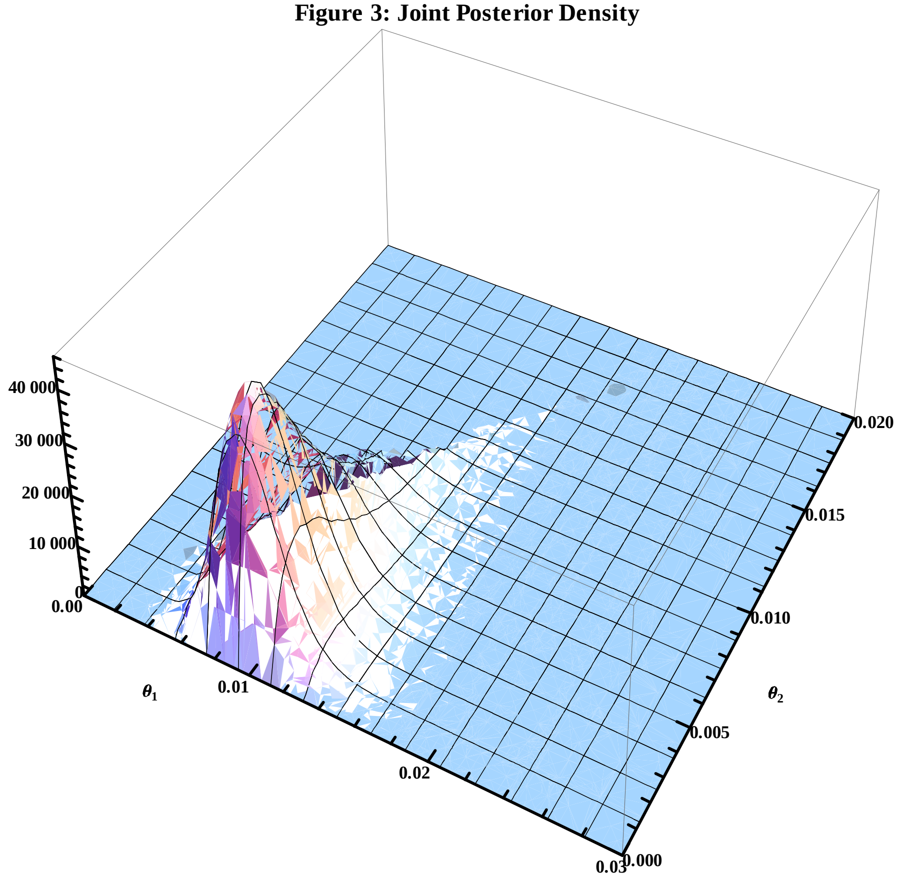
dataset consisting of 100 asset/years and zero defaults. The posterior summary statistics are $E\theta_2 = 0.0035$ and $sd(\theta_2) = 0.0023$. This analysis is the same as a joint analysis based on independent priors, so the joint is the product of these marginals. The entropy of that joint prior distribution is -8.79. The posterior entropy is -9.70.

The prior distribution given by the elicited marginal distributions combined with the stochastic ordering is shown in Figure 2.

Here the correlation is now 0.31 and the entropy is -8.93. The posterior density is given in Figure 4.

The posterior entropy is -9.78. The moments are $E\theta_1 = 0.0093$, $sd(\theta_1) = 0.0018$ and $E\theta_2 = 0.0036$, $sd(\theta_2) = 0.0023$. The posterior correlation is 0.13. Here there is very little gain in terms of location or precision from imposing the stochastic ordering of the default rates, due probably to the rather high precision of the marginal prior distributions assessed from the experts. There is information on correlation, not available in the standard analysis, which is a part of the characterization of the posterior uncertainty about the default rates not available from the separate anal-

Figure 3: Joint Posterior Density

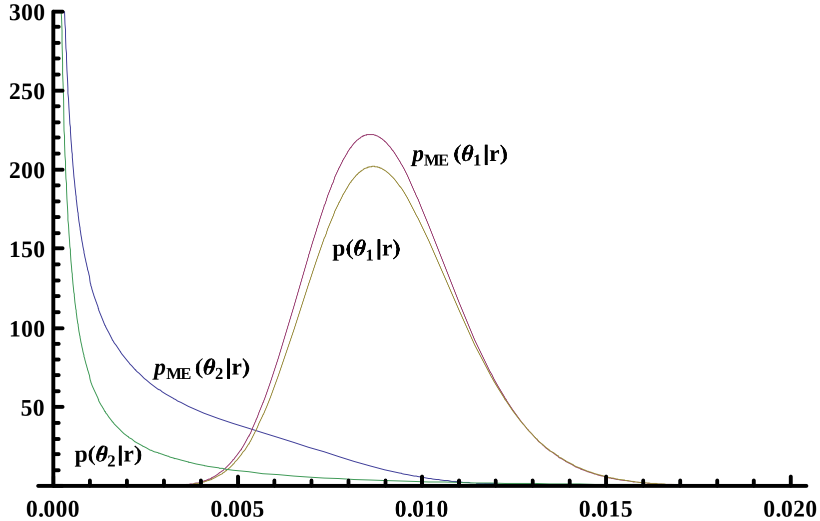


yses. This could be quite important for managing the risk in the overall portfolio.

6 Robustness: Less Confident Experts

In our application the priors are well separated and so there was little gain from imposing the stochastic ordering. The posterior correlation may be important for stress-testing as discussed in Basel Committee on Banking Supervision (2009). To illustrate a case with more substantial gains from imposing the restriction, we consider less-confident experts. That is, consider priors with the same means as the assessed priors, but with larger variances. This can be considered a robustness or sensitivity analysis. In the language of risk management, this might be a part of model validation. To fix ideas, we consider here increasing the prior standard deviations by a factor of three. With this change the probability that the ordering is violated in the product prior is 0.24. The correlation in the ME prior imposing the ordering is 0.77. Using the same data, the posterior moments under the product prior corresponding to separate analyses are $E\theta_1$ =0.0091, $sd(\theta_1)$ =0.0020 and $E\theta_2$ =0.0016, $sd(\theta_2)$ =0.0031. With the ME prior we have $E\theta_1$ =0.0091, $sd(\theta_1)$ =0.0018 and $E\theta_2$ =0.0021, $sd(\theta_2)$ =0.0026 and the correlation is 0.15. This exercise is informative about the value of the joint analysis. In the higher-default portfolio segment with the larger sample size the prior and the data are in agreement and there is little change in location under any analysis. For the low-default portfolio segment, where a typical sample will have zero defaults, the data and the prior disagree (to some extent). In the separate analysis, the data on the related segment (related through the prior alone in this simple specification) do not affect the inference on the default rate in the low-default portfolio segment. In the joint analysis, the data are informative and the resulting posterior mean for the low-default portfolio is substantially increased, from 0.0016 to 0.0021. For both default rates, the posterior precision is increased by moving to the joint analysis. The marginal posterior densities from both the separate and the ME analyis are reported in Figure 4 (in the analysis above these are visually identical). This figure clearly shows the substantial effect of the joint analysis on the uncertainty about the lower default rate. As noted, the higher-default segment, with more data and data in agreement with the prior, is little affected by the joint analysis.

Figure 4: Posterior densities – less certain experts



7 Conclusion

The maximum-entropy joint distribution for two stochastically-ordered random variables with given marginal distributions is obtained in closed form. An interpretation is that this distribution could be used as a sensible joint prior distribution when information is obtained from separate experts on related random variables, experts who can agree on the stochastic ordering but have no further information on the dependence. An application to default rates for different portfolio segments is given. The method is clearly feasible and can be valuable in providing a way to combine data information across portfolio segments, exploiting dependence introduced through the prior due to the stochastic ordering restriction. The Bayesian approach in the application is not necessary for the approach to be useful. For example, the expert information could alternatively be based on previous analyses, separately by segment, and the marginal distributions involved could potentially be sampling distributions of previous estimators. This case is applicable to the credit-scoring problem, which typically considers 20 segments separately. The potential gain from a joint analysis here is clear and will be pursued in ongoing research.

References

BASEL COMMITTEE ON BANKING SUPERVISION (2004): "International Convergence of Capital Measurement and Capital Standards: A Revised Framework," Bank for International Settlements.

——— (2005): "Basel Committee Newsletter No. 6: Validation of Low-Default Portfolios in the Basel II Framework," Discussion paper, Bank for International Settlements.

——— (2009): "Principles for sound stress testing practices and supervision," Consultative Document 147, Bank for International Settlements.

JAYNES, E. T. (2003): *Probability Theory: The Logic of Science*. New York: Cambridge University Press.

KADANE, J. B., AND L. J. WOLFSON (1998): "Experiences in Elicitation," *The Statistician*, 47(1), 3–19.

KAMAE, T., U. KRENGEL, AND G. L. O'BRIEN (1977): "Stochastic Inequalities on Partially Ordered Spaces," *The Annals of Probability*, 5(6), 899–912.

KIEFER, N. M. (2008): "Default Estimation for Low Default Portfolios," *Journal of Empirical Finance*, 16.

KIEFER, N. M. (2009): "Default Estimation and Expert Information," *Journal of Business and Economic Statistics*, (forthcoming).

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

STRASSEN, V. (1965): "The Existence of Probability Measures with Given Marginals," *Annals of Mathematical Statistics*, 36(2), 423–439.