Updating Beliefs about Variables Given New Information on How Those Variables Relate
Robert Bordley & Vicki Bier
October 28, 2008

Abstract

Bayesian techniques specify how to update beliefs about a variable given information on that variable or related variables. In many cases, statistical analyses also provide information about the relationship between variables, but the Borel Paradox prohibits many natural ways of updating beliefs conditioned on information about a relationship. This paper presents a method by which beliefs can be updated without violating the Borel Paradox under certain circumstances. We apply our approach to relationships specified by a statistical model (i.e., regression), and relationships described by statistical simulation.

1 Introduction

Bayesian techniques specify how to update prior beliefs about Y given information on some variable X (where X is a column vector and Y is typically a scalar). In many cases, however, statistical analysis directly provides information on the relationship between Y and X; for example:

- a decision maker is presented with a statistical model (i.e., a regression) relating Y and X
- a decision maker is presented with the results of a simulation that specifies the values of Y resulting from various choices of X

In this case, applying Bayes Rule is more complicated than when simply updating beliefs given information on a variable. As a result, the authors have observed practical contexts where Bayes Rule is misapplied.

As an illustration, let ϕ denote the new relational information arising from a statistical analysis. One common approach for estimating $\Pr.\{Y|\phi\}$ combines the analysis-based probability $\Pr.\{Y|X,\phi\}$ with the decision maker's prior probability $\Pr.\{X\}$ to derive the inferred estimate, $\int_x \Pr.\{Y|X,\phi\}\Pr.\{X\}$. This approach is only consistent with Bayes Rule when $\Pr.\{X|\phi\} = \Pr.\{X\}$. Thus, this approach presumes that the information ϕ does not change the decision maker's prior probability on X.

Suppose, however, that the decision maker is already fairly comfortable with his prior probability distribution over Y. If the analysis implies that this prior distribution should change significantly, the decision maker will often question the inputs to the analysis. Even if those inputs came from the same decision maker, he might be more willing to view them as wrong than to change his beliefs about Y. If he has little confidence in his beliefs about

X, the decision maker might even invoke the celebrated *Garbage-in/Garbage-out* maxim and reject whatever the analysis says about Y.

Since the variable of interest is typically Y (rather than X), this effectively reduces the value of doing the analysis (and may even get the analyst fired). While it may be embarrassing for the analyst's model to be rejected, there is clearly no reason to believe that the client should change his beliefs only about outputs and not about inputs. Nonetheless, dismissing the analysis by changing only beliefs about inputs, rather than outputs, is just as simplistic. An alternative, less simplistic, approach first updates priors on X, by using the model to make predictions about values of Y for which the decision maker has good intuitions. Then, the updated priors on X are held fixed and the model is used to make inferences about the posterior value of Y. Although this second approach is better, this still does not represent an exact Bayesian approach to this problem — and hence does not fully utilize all available information.

A fully Bayesian distribution on Y, given the new information ϕ, updates beliefs about both Y and X, using the information ϕ about how Y and X are related. In other words, we would like to be able to use ϕ to update the decision maker's priors on both X and Y to get:

$$\Pr.\{XY|\phi\} \propto \Pr.\{\phi|XY\} \Pr.\{XY\} \tag{1}$$

from which $\Pr.\{Y|\phi\}$ can be computed. In some cases, the new information ϕ can be interpreted as specifying the exact functional relationship between X and Y (for example, ϕ might specify $Y = X$.) When ϕ does specify the exact functional relationship between Y and X, equation(1) is equivalent to an existing technique known as Bayesian synthesis (Raferty, Givens & Zeh, 1992). In the language of Bayesian synthesis, X, Y are parameters, $\Pr.\{X, Y\}$ is the pre-model distribution and $\Pr.\{XY|\phi\}$ is the post-model distribution arising from using the deterministic model ϕ to restrict the parameters to some subspace.

The functional relationship, ϕ, specifies a value of y for each possible value of x, or, conversely, it specifies the impossibility of observing an infinite set of possible combinations of (x, y). As a result, the decision maker's prior beliefs often assign a probability of zero to the set of possible (x, y) pairs that are perfectly consistent with the model. This makes conditioning on ϕ ill-defined so when probabilities over a lower-dimensional space are inferred from probabilities over a higher-dimensional space, the resulting conditional probability densities need not be invariant with respect to variable transformations. This failure of invariance, known as the Borel Paradox, means that simple reparameterizations of the model that, in principle, should not matter (like transforming X from length to area) lead to different results. Proschan & Presnell(1998) demonstrate this problem in computing the probability of Y conditioned on $Y = X$ when prior beliefs have X, Y independent Gaussian random variables. Schweder & Hort (1996) and Wolpert (1995) verified the seriouness of the Borel Paradox for Bayesian synthesis by showing how the results of Bayesian synthesis varied arbitrarily with different variable transformations. These concerns ultimately led to the abandonment of Bayesian synthesis (Poole & Raftery, 2000).

However, we would argue that there are many realistic contexts in which the decision maker's prior distribution does assign nonzero measure to the set of (x, y) pairs that are consistent with the model ϕ. For example, the error term in stochastic models typically allows a wide range of y values for each x value because of randomness. Specifically consider the case where a decision maker is offered the linear model

$$Y = \beta_0 + \beta' X + \epsilon$$

where β is a column vector with β' denoting its transpose and ϵ is a Gaussian random variable. Suppose the decision maker's prior beliefs about X and Y are Gaussian with means m_X, m_Y and with $\Sigma_{XX}, \Sigma_{XY}, \Sigma_{YY}$ being the variance-covariance matrix among the X's, the covariance between the X's and Y, and the variance of Y. As Appendix I shows, these Gaussian prior beliefs over X, Y are equivalent to prior beliefs over X and a new uncorrelated Gaussian variable, ϵ_0, defined by $\epsilon_0 = Y - \alpha_0 - \alpha' X$ (where the scalar α_0, the column vector α, and the variance of ϵ_0 are all computable from from the variance of Y, its covariance with X, and the mean of Y.) Thus, an individual with a Gaussian prior over X and Y can be viewed as implicitly relating Y to X via a linear model. So the individual, faced with the linear model $Y = \beta_0 + \beta' X + \epsilon$, will assign the model nonzero measure and interpret the model as providing information on the parameters α_0, α and the variance of ϵ_0. Under these circumstances, the Borel Paradox is avoided, and the individual's prior beliefs about X and Y can be updated in the conventional Bayesian fashion.

More generally, note that any infinitely differentiable functional relationship between Y and X can be represented by a Taylor series, with coefficients describing all possible derivatives of Y with respect to X. Hence, both the decision maker's prior beliefs about X and Y and the proposed new model can be rewritten in terms of the Taylor series although the Taylor coefficients of the model will typically differ from the coefficients implicit in the decision maker's prior beliefs. Thus, the model provides new information on the coefficients in the decision maker's prior model. This new information allows the decision maker to update his prior beliefs about X, Y.

The next section illustrates this approach when the decision maker is given a linear model relating Y to X. As we show, the same result can be derived using a procedure called Bayesian Melding by combining the decision maker's prior on Y with the model prediction (based on the decision maker's prior on X) as if they were two correlated forecasts. The third section illustrates our approach when the decision maker is instead given the results of a simulation or other experiment relating Y to X. In this case, we find that beliefs about both Y and X change (contrary to some arguments that simulation results should not change beliefs about inputs). However, we find that beliefs about inputs are updated somewhat differently than beliefs about outputs. In both cases, the decision maker is assumed to have a Gaussian prior over X and Y.

2 Updating One's Beliefs Given an Analytical Model

2.1 Bayesian Solution

Suppose the individual's new information, ϕ, is an analytical model of the form

$$Y = \beta_0 + \beta' X + \epsilon$$

where v is the (known) variance of ϵ. Then if the decision maker knows X and Y, the only way the model, ϕ, can be valid is if $\epsilon = Y - \beta_0 - \beta' X$. Thus we write $\Pr.\{\phi|X,Y\}$, which describes the likelihood of the model being valid, as proportional to the probability density of $\epsilon = Y - \beta_0 - \beta' X$. Since ϵ is assumed Gaussian with known variance v, the probability density of ϕ given Y and X is proportional to $\exp(-\frac{1}{2v}(Y - \beta_0 - \beta' X)^2)$. As Appendix II shows, this implies:

Proposition 1: Suppose X, Y are Gaussian with means m_X, m_Y and variance-covariance matrix Σ. Let Z be a $K+1$ column vector whose first K components equal X and whose last component is $Y - \beta_0$ (with m_Z denoting the mean of Z). Let B be a $K+1$ column vector whose first components equal $-\beta$, and whose last component equals 1. Define the relative variances of the variables (X, Y) by the elements of the following column vector:

$$w = \frac{B'\Sigma}{v + B'\Sigma B}$$

Note that $\beta_0 + \beta' m_X - m_Y$ measures the degree to which the model is inconsistent with the individual's prior beliefs. Then the posterior density of Z is Gaussian with mean

$$m_{Z|\phi} = m_Z(1 - B'w) = m_Z + (\beta_0 + \beta' m_X - m_Y)w$$

and variance $\Sigma(1 - B'w)$.
Proof : See Appendix II.

Thus, the change in each variable's mean value equals the product of the expected error in the model — given the prior beliefs — and the proportion of variance associated with each variable. If w_Y is the last element of w, then we have

$$w_Y = \frac{\Sigma_{YY} - \beta' \Sigma_{XY}}{v + \Sigma_{YY} - 2\beta' \Sigma_{XY} + \beta' \Sigma_{XX} \beta} \tag{2}$$

so that Y has posterior mean

$$m_{Y|\phi} = m_Y(1 - w_Y) + (\beta_0 + \beta' m_X) w_Y \tag{3}$$

and posterior variance

$$\Sigma_{YY} - w_Y \beta' \Sigma_{XY} \tag{4}$$

As a result, the percentage reduction in variance arising from the model ϕ is

$$w_Y[1 - \beta' \frac{\Sigma_{X,Y}}{\Sigma_{YY}}]$$

We now compare this solution with a proposed heuristic solution to the problem.

2.2 Bayesian Melding

After Bayesian synthesis was abandoned because of the Borel Paradox, Roback (1998), Roback & Givens(1999) and Roback & Givens (2001) proposed a supra-Bayesian approach. In this approach, an inferred distribution on Y is constructed by coupling the decision maker's prior distribution over X with the newly proposed statistical relationship between X and Y. The inferred distribution on Y is then combined with the prior distribution on Y to get a posterior distribution using 'supra-Bayesian' techniques (Morris, 1977; Winkler, 1981; Bordley, 1982; Lindley, 1983; French, 1985; Genest & Zidek, 1986). These techniques are called supra-Bayesian because they require additional criteria beyond what the Bayesian formalism requires (e.g., Madansky, 1978).

In our particular problem, a supra-Bayesian approach might view the decision maker as combining two forecasts of Y:

- a forecast Y_0, based on his prior beliefs about Y, which treats Y as Gaussian with mean m_Y
- a second forecast $Y_1 = \beta_0 + \beta' X + \epsilon$ based on the prior distribution of X, so that the mean of Y_1 is $m_Y^* = \beta_0 + \beta' m_X$

Then, as Appendix III notes, these forecasts are inversely weighted by their correlation-adjusted variances and normalized (Bordley, 1986) to produce the combined forecast:

$$m_Y^* = m_Y(1 - w_Y^*) + w_Y^* m_{Y_1} = m_Y(1 - w_Y^*) + (\beta_0 + \beta' m_X) w_Y^*$$

where

$$w_Y^* = \frac{\Sigma_{Y_0Y_0} - \beta'\Sigma_{Y_0X}}{\Sigma_{Y_0Y_0} - 2\beta'\Sigma_{Y_0X} + \beta'\Sigma_{XX}\beta + v}$$

Thus, the solution arising from interpreting this problem as a problem in combining forecasts is, in fact, consistent with the fully Bayesian approach. Note, however, that achieving this consistency required careful consideration of the correlation between the two forecasts. Neglecting the correlation would have erroneously estimated w_Y^* as $\frac{\Sigma_{Y_0Y_0}}{\Sigma_{Y_0Y_0} + \beta'\Sigma_{XX}\beta + v}$ which would have overestimated the amount of weight to assign to the new forecast. Clemen & Winkler (1995) rightly stress the importance of adjusting for the correlation between forecasts.

These results are consistent with Roback and Givens(2001)'s proposed use of forecast combination techniques to update a decision maker's prior beliefs about Y and X in light of analytical information on how Y and X relate. In other words, what they termed a supra-Bayesian approach coincides — in the Gaussian case — with the conventional Bayesian approach. However, as we now show, forecast combination does not seem as applicable for integrating information from simulations or other experiments.

3 Updating One's Beliefs Given Experimental Results

The previous section focused on the case in which the individual learned of a model specifying the exact relationship between the scalar Y and the vector X. In the case of an experiment, the experimenter prespecifies a vector of input values $X(1), ...X(m)$, and observes the corresponding output values $Y(1), ...Y(m)$. Hence, the information provided by the experiment is the sequence:

$$(X(1), Y(1)), ..., (X(m), Y(m))$$

If X is a vector of length K, this sequence can be written as

$$(X_1(1), ...X_K(1), Y(1)), ..., (X_1(m), ...X_K(m), Y(m))$$

Even though the model does not explicitly specify the parameters β_0 and β relating X to Y, the decision maker's prior over those parameters will still be relevant. Since the decision maker is familiar only with X and Y, it seems reasonable to assume that any information the decision maker has about β_0 and β can be completely derived from his beliefs about X and Y in the model $Y = \beta_0 + \beta' X + \epsilon$. Thus the decision maker will not be able to discriminate between different values of β_0 and β that are all equally consistent with the values of X and Y (in the sense of having equally large or small values of ϵ). More specifically, we will assume that the probability distribution over all values of β_0, β consistent with any particular values of X and Y is uniform. As we show, given this assumption, we have the following result:

Proposition 2: Suppose we run T trials of the experiment in which the vector of input values has mean $\hat{X}$ and variance-covariance V. Suppose that the resulting output has mean $\hat{Y}$ with variance s^2, and that the observed correlation between inputs and outputs is r. Let $x = X - \hat{X}$ and $y = Y - \hat{Y}$. If the variance in the simulation noise is given by a nonzero value v, then defining the factor $v(x) = \frac{2v}{T}(1 + x'V^{-1}x)$ implies that

$$\Pr.\{x, y|\hat{X}, \hat{Y}, r\} \propto |v(x)|^{-1} \exp(-\frac{(y - r'x)^2}{2v(x)}) \Pr(x, y)$$

Proof: See Appendix IV.

Note that the likelihood function in the proof of Proposition 1, $\exp(-\frac{1}{2v}(Y - \beta_0 - \beta' X)^2)$, corresponds to the likelihood function in Proposition 2 with $\beta_0 = \hat{Y} - r'\hat{X}, \beta = r$, and $v = v(x)$. Thus, Proposition 2 differs from Proposition 1 only in that the variance $v(x)$ increases for experimental results associated with inputs that the decision maker considers unlikely a priori. By discounting outliers in this manner, the results in Proposition 2 effectively reduce the variability among the experiment's inputs. The resulting joint distribution is, of course, not Gaussian.

However, the distribution of Y (conditional on x) will be Gaussian with posterior conditional variance

$$\sigma^2_{x|\phi} = \frac{1}{\frac{1}{v(x)} + \frac{1}{\Sigma_{YY|x}}}$$

and posterior conditional mean

$$m_{Y|x,\phi} = \frac{\frac{r'x}{v(x)} + \frac{m_{Y|x}}{\Sigma_{YY|x}}}{\frac{1}{v(x)} + \frac{1}{\Sigma_{YY|x}}}$$

Define the adjusted difference between the prior mean of Y and the simulated mean by

$$\delta(x) = \frac{\frac{m_{Y|x}}{\Sigma_{YY|x}} - \frac{r'x}{v(x)}}{\frac{1}{\Sigma_{YY|x}} + \frac{1}{v(x)}}$$

Then, as Appendix V shows, the marginal posterior distribution of x has the form

$$\frac{\Sigma_{YY|x}}{v(x) + \Sigma_{YY|x}} \exp(-\frac{1}{2}\delta(x)) \Pr.\{x\}$$

which is a non-Gaussian adjustment of the decision maker's Gaussian prior on x that essentially reduces the likelihood of input values leading to implausible values of y. If $\Sigma_{YY|x}$ is large (i.e., if the decision maker doesn't consider y to be highly predictable from x), then the resulting change in the prior probability of x will be small. Also, note that the degree of adjustment will be smaller for input values that the decision maker considers unlikely (i.e., for cases where $v(x)$ is large).

These results are directly relevant to a debate (Andradottir & Bier, 2000) in the literature over whether the results of a simulation provide information on the inputs to that simulation. Chick (1997) argued that the results of a simulation — which rely on inputs specified by the decision maker — cannot provide any information about the reasonableness of the inputs. An opposite perspective is suggested based on the 'reverse Monte Carlo' method sometimes used in the physical sciences (Lenoble, 1985; Petty, 1994), which "exploits obvious symmetries with respect to the direction of time " (Petty, 1994). This perspective suggests that the decision maker's beliefs about both inputs and outputs are symmetrically affected by the results of the simulation. The results of this paper present an intermediate view: The results of a simulation do affect the decision maker's beliefs about the reasonableness of the inputs, although not in the same way as they affect beliefs about the outputs.

4 Conclusions

Bayesian methodology clearly specifies how to update beliefs about a variable on the basis of information about that variable, or other variables. However, many statistical analyses, instead of providing information about variables, provide information mainly about the relationship between variables. As we show, an exact Bayesian approach to this problem can be defined when the decision maker's prior implicitly presumes the same kind of relationship as specified in the statistical analysis. When the prior is Gaussian (and the statistical model is linear), deriving the Bayesian posterior distribution is straightforward. As we show, the resulting posterior distribution coincides with the result arising from an approach based on combining forecasts.

We then examine the Bayesian solution when the relationship is specified through simulation or the results of experiments. While the formulas for the mean and variance are somewhat similar, the posterior distribution of X is no longer Gaussian, even though the conditional distribution of Y given X remains Gaussian.

Our results show that current practices in modeling and simulation generally understate the degree of uncertainty associated with their results. This arises because priors on both inputs and outputs should be updated based on the results of the model and simulation results. Since current practices do not revise priors on inputs, they typically adjust priors on outputs more than is appropriate. The resulting overestimation of the informativeness of model and simulation results can have serious practical implications.

The current paper provides a solution which is valid under certain restrictions (e.g., the assumption of linearity). While satisfied in certain social science applications, further work is needed to relax these restrictive assumptions, This could eventually enable the development of user-friendly software (similar to the updating software that already exists) for updating both model inputs and outputs simultaneously. This would lead to a more rigorous approach to output analysis both for simulation in particular as well as for more general contexts.

Acknowledgements

The authors would like to thank Prof. Nilay Argon of the University of North Carolina for her patient readings of numerous early drafts of this paper, for numerous useful comments, and for probing questions that greatly improved the clarity of the final manuscript.

This research was supported in part by the United States Department of Homeland Security through the Center for Risk and Economic Analysis of Terrorist Events (CREATE) under grant number EMW-2004-GR-0112. However, any opinions, findings, and conclusions or recommendations in this document are those of the authors and do not necessarily reflect views of the United States Department of Homeland Security.

References

[1] Andradottir, S., and V. M. Bier. (2000) "Applying Bayesian Ideas in Simulation," *Simulation Practice and Theory,*Vol.8, 253-280.

[2] Berger, James. *Statistical Decision Theory and Bayesian Analysis.* New York: Springer-Verlag, 1995.

[3] Bordley, R. (1982) "A Multiplicative Formula for Aggregating Probability Estimates." *Management Science.10*

[4] Bordley, R. (1986). "Linear Combination of Forecasts with an Intercept: A Bayesian Approach." *Journal of Forecasting.*5,4.

[5] Chick, S. E. “Bayesian Analysis for Simulation Input and Output.” in S. Andradttir, K. J. Healy, D. H. Withers and B. L. Nelson, eds., *Proceedings of the 1997 Winter Simulation Conference,* 253-260, 1997.

[6] Clemen, R. & R. Winkler. (1995). “Limits for the Precision & Value of Information from Dependent Sources.” *Operations Research.* 33,427-442.

[7] French, S. (1985). “Group consensus probability distributions: A critical survey.” in Bernardo, J., DeGroot,M., Lindley, D. & Smith A. (ed.) *Bayesian Statistics 2* ,pp.183-202, Elsevier Science Publishers B.V., Amsterdam.

[8] Gelman, A., J. Carlin, H. Stern & D.Rubin(1995). *Bayesian Data Analysis.*Chapman & Hall, London.

[9] Genest,C, Weerahandi, S & Zidek, J. (1984). “Aggregating Opinions through Logarithmic Pooling.” *Theory and Decision.* 17:61-70.

[10] Genest, C. & Zidek, J. (1986) “Combining Probability Distributions: A Critique & an Annotated Bibliography.” *Statistical Science.* 1:114-148.

[11] Givens, G. & M. Bravington(1996). “Confirmation, revision and sensitivity analysis of the 1994 Scientific Committee Assessment of the Bering-Chukfor-Beaufort Seas Stock of Bowhead Whales.” *Report of the International Whaling Commission.* 46,501-508.

[12] Glynn, P. W. “Problems in Bayesian Analysis of Stochastic Simulation.” J. R. Wilson, J. O. Henriksen, and S. D. Roberts, eds., *Proceedings of the 1986 Winter Simulation Conference*

[13] Lindley, D., Tversky A. & Brown, R. (1979). “On the Reconciliation of Probability Assessments.” *Journal of the Royal Statistical Society, Series A*: 142:146-180.

[14] Madansky, A. (1978). “Externally Bayesian Groups.” *Unpublished Manuscript.* University of Chicago.

[15] Martz, H., and R. Waller. *Bayesian Reliability Analysis.* New York: John Wiley, 1982.

[16] Poole, D. & A. Raftery.(2000). “Inference for Deterministic Simulation Models: The Bayesian Melding Approach.” *Journal of the American Statistical Association,*95,1222-1255.

[17] Proschan, M. & B. Presnell(1998). “Expect the Unexpected from Conditional Expectation.” *The American Statistician*, 52,3,248-252.

[18] Raftery, A., Givens, G. & J. Zeh(1998). “Inference from a Deterministic Population Dynamics Model about Bowhead Whale. (with discussion) *Journal of the American Statistical Association.*93,451-463.

[19] Raiffa, H. & R. Schlaiffer(1961). *Applied Statistical Decision Theory.* New York: Wiley.

[20] Roback, P. (1998). "The pooling of prior distributions via logarithmic and supra-Bayesian methods with application to Bayesian inference in deterministic simulation." *PhD Thesis.*Colorado State University.

[21] Roback, P. & G. Givens.(2001). "Supra-Bayesian Pooling of Priors Linked by a Deterministic Simulation Model." *Communications in Statistics: Simulation and Computing.* 30,447-476.

[22] Roback, P. & G. Givens.(1999) "Logarithmic pooling of priors linked by a deterministic simulation model" , *J. Computational and Graphical Statistics,* 8: 452-478.

[23] Schweder, T. & N. Hjort(1996). "Bayesian Synthesis or Likelihood Synthesis—What does Borel's Paradox Say?" *Report of the International Whaling Commission,*46,475-480.

[24] Smith, A.F. & A.E. Gelfand(1992), "Bayesian statistics without tears: A sampling/resampling perspective," *American statistican* 46, no. 2, pp. 84-88.

[25] Winkler, R. (1981). "Combining Probability Distributions from Dependent Information Sources." *Management Science.* 27:479-488.

[26] Wolpert, R. (1995). "Comment on 'Inference from a Deterministic Population Dynamics Model for Bowhead Whales'" by A. Raftery, G.. Givens & J. Zeh. *Journal of the American Statistical Association,*90,427-427.

5 Appendices

5.1 Appendix I

To determine the scalar α_0, the column vector α and the variance of ϵ_0, define the variance of Y, its covariance with X, and the mean of Y, respectively, by

$$\begin{aligned} \Sigma_{YY} &= \alpha'\Sigma_{XX}\alpha + Var(\epsilon_0) \\ \Sigma_{XY} &= \alpha'\Sigma_{XX} \\ m_Y &= \alpha_0 + \alpha' m_X \end{aligned}$$

Then the parameters of the linear equation, $Y = \alpha_0 + \alpha' X + \epsilon_0$, can be inferred from $\Sigma_{YY}, \Sigma XY, m_Y$, using

$$\begin{aligned} \alpha' &= \Sigma_{XY}\Sigma_{XX}^{-1} \\ \alpha_0 &= m_Y - \alpha' m_X = m_Y - \Sigma_{XY}\Sigma_{XX}^{-1} m_X \\ Var(\epsilon_0) &= \Sigma_{YY} - \alpha'\Sigma_{XX}\alpha = \Sigma_{YY} - \Sigma_{XY}\Sigma_{XX}^{-1}\Sigma_{XX}\Sigma_{XX}^{-1}\Sigma_{XY} = \Sigma_{YY} - \Sigma'_{XY}\Sigma_{XX}^{-1}\Sigma_{XY} \end{aligned}$$

5.2 Appendix II

Let Z be the vector whose first components are X and whose last component is $Y - \beta_0$. Define the row vector B whose first elements are $-\beta$ and whose last element is 1. If X is a K by 1 vector, then

$$(\beta_0 + \beta' X - Y)^2 = X'\beta\beta' X + (Y - \beta_0)^2 - 2X'\beta(Y - \beta_0)$$

Hence the posterior probability is

$$\Pr.\{XY|\beta_0, \beta\} \propto \exp(-Z'BB'Z)/2v) \Pr.\{XY\} \tag{5}$$

Unlike the traditional inverse variance-covariance matrix, BB' is singular.

If X, Y are Gaussian with means m_X, m_Y and variance-covariance matrix Σ, then Z is Gaussian with its first components having mean m_x and its last component having mean $m_Y - \beta_0$. If the variance-covariance of Z is Σ, then

$$\Pr.\{X, Y\} \propto \exp(-\frac{1}{2}(Z' - m_Z')\Sigma^{-1}(Z - m_Z))$$

so that

$$\Pr.\{X, Y|\beta_0, \beta\} \propto \exp(-\frac{1}{2}[Z'[\frac{BB'}{v} + \Sigma^{-1}]Z - 2m_Z'\Sigma^{-1}Z + m_Z'\Sigma^{-1}m_Z])$$

where all values of X, Y are possible because of the normality of the error term. Following conventional arguments in Raiffa and Schlaiffer, define

$$W = (\frac{BB'}{v} + \Sigma^{-1})^{-1} = \Sigma(I + \frac{BB'}{v}\Sigma)^{-1}$$

so that

$$\Pr.\{X, Y|\beta_0, \beta\} \propto \exp(-\frac{1}{2}Z'W^{-1}Z + m_Z'\Sigma^{-1}Z)$$

Defining

$$m_{Z|\phi}' = m_Z'\Sigma^{-1}W = m_Z'(I + BB'\Sigma/v)^{-1}$$

implies

$$\Pr.\{X, Y|\beta_0, \beta\} \propto \exp(-\frac{1}{2}[Z' - m_{Z|\phi}']W^{-1}[Z - m_{Z|\phi}])$$

so that Z is Gaussian with mean $m_{Z|\phi}$ and variance-covariance W. Cross-multiplying

$$m_{Z|\phi}' = m_Z'(I + BB'\Sigma/v)^{-1}$$

by $I + BB'\Sigma/v$ gives

$$m_{Z|\phi}' + m_{Z|\phi}'BB'\Sigma/v = m_Z'$$

Cross-multiplying by B gives

$$m_{Z|\phi}'B + m_{Z|\phi}'BB'\Sigma B/v = m_Z'B$$

Dividing by $1 + B'\Sigma B/v$ gives

$$m'_{Z|\phi}B = \frac{m'_Z B}{1 + B'\Sigma B/v}$$

Substituting this expression into the previous equation:

$$m'_{Z|\phi} + m'_{Z|\phi}BB'\Sigma/v = m'_Z$$

gives

$$m'_{Z|\phi} + \frac{m'_Z B}{v + (B'\Sigma B)}B'\Sigma = m_Z$$

which can be rewritten as

$$m_{Z|\phi} = m_Z[I - \frac{BB'\Sigma}{v + B'\Sigma B}]$$

Note that $-m_Z B = \beta_0 + m_X\beta' - m_Y$ is the expected error in the model given the decision maker's priors. The term $w = \frac{B'\Sigma}{v+B'\Sigma B}$ reflects the proportion of the overall variance — attributable to the variables and the model — attributable to a specific variable. Hence, we have

$$m_{Z|\phi} - m_Z = -m_Z Bw$$

Also note that since

$$W = (\Sigma^{-1} + \frac{BB'}{v})^{-1} = (I + \frac{BB'}{v}\Sigma)^{-1}\Sigma$$

we have

$$W + W\frac{BB'}{v}\Sigma = \Sigma$$

and

$$WB + W\frac{B}{v}\Sigma B = \Sigma B$$

so that

$$WB = \frac{\Sigma B/v}{1 + B'\Sigma B}$$

Substituting into $W + W\frac{BB'}{v}\Sigma = \Sigma$ gives

$$W + \frac{\Sigma B/v}{1 + B'\Sigma B}B'\Sigma/v = \Sigma$$

Hence

$$W = \Sigma[I - \frac{BB'\Sigma}{v + B'\Sigma B}] = \Sigma(I - Bw)$$

5.3 Appendix III

To compute the combined 'forecast' (Bordley, 1986), we need to apply the following steps:

1. compute the variance-covariance matrix between the two forecasts

$$\begin{pmatrix} \Sigma_{Y_0Y_0} & \Sigma_{Y_0Y_1} \\ \Sigma_{Y_0Y_1} & \Sigma_{Y_1Y_1} \end{pmatrix}$$

2. construct the inverse variance-covariance matrix

$$\begin{pmatrix} \frac{\Sigma_{Y_1Y_1}}{(\Sigma_{Y_1Y_1}\Sigma_{Y_0Y_0}-\Sigma^2_{Y_0,Y_1}} & -\frac{\Sigma_{Y_0Y_1}}{(\Sigma_{Y_1Y_1}\Sigma_{Y_0Y_0}-\Sigma^2_{Y_0,Y_1}} \\ -\frac{\Sigma_{Y_0Y_1}}{(\Sigma_{Y_1Y_1}\Sigma_{Y_0Y_0}-\Sigma^2_{Y_0,Y_1}} & \frac{\Sigma_{Y_0,Y_0}}{(\Sigma_{Y_1Y_1}\Sigma_{Y_0Y_0}-\Sigma^2_{Y_0,Y_1}} \end{pmatrix}$$

3. assign m_{Y_1} a weight inversely proportional to $\Sigma_{Y_1Y_1} - \Sigma_{Y_0Y_1}$, and m_Y a weight inversely proportional to $\Sigma_{Y_0Y_0} - \Sigma_{Y_0Y_1}$.

Since $\Sigma_{Y_1Y_1} = \beta'\Sigma_{XX}\beta + v$ and $\Sigma_{Y_0Y_1} = \beta'\Sigma_{XY}$, the combined forecast will be

$$m^*_Y = m_Y(1 - w^*_Y) + w^*_Y m_{Y_1} = m_Y(1 - w^*_Y) + (\beta_0 + \beta' m_X)w^*_Y$$

where

$$w^*_Y = \frac{\Sigma_{Y_0Y_0} - \beta'\Sigma_{Y_0X}}{\Sigma_{Y_0Y_0} - 2\beta'\Sigma_{Y_0X} + \beta'\Sigma_{XX}\beta + v}$$

5.4 Appendix IV

We can write

$$\Pr.\{XY|X(t), Y(t), t \in T\} = \Pr.\{Y(t)|X, Y, X(t), t \in T|XY\} \Pr.\{X(t)|X, Y, t \in T\} \Pr.\{XY\}$$

Since $X(t)$, by itself, provides no information about X, Y, this becomes

$$\Pr.\{X, Y|X(t), Y(t), t \in T\} = \Pr.\{X, Y\} \Pr.\{Y(t)|X, Y, X(t), t \in T\}$$

We can condition on β and write

$$\Pr.\{Y(t)|X, Y, X(t), t \in T\} = \int_\beta \Pr.\{Y(t)|X, Y, X(t), \beta, t \in T\} \Pr.\{\beta|X, Y, X(t), t \in T\}$$

Now $Y = \beta_0 + \beta'X + \epsilon$ and $Y(t) = \beta_0 + \beta'X(t) + \epsilon(t)$. Thus observing $Y(t)$ is the same as observing $\epsilon(t) = Y(t) - \beta_0 - \beta'X(t)$. In addition, knowing Y and X tells us that β_0 must equal $Y - \beta'X - \epsilon$. Thus, the probability density of $Y(t)$ can be determined from the probability density of

$$\epsilon(t) = Y(t) - Y + \beta'(X - X(t)) + \epsilon$$

Defining $\epsilon^*(t) = \epsilon(t) - \epsilon$ implies that

$$\Pr.\{Y(t)|X, Y, X(t), \beta, t \in T\} = \Pr.\{\epsilon^*(t) = Y(t) - Y + \beta'(X - X(t))|X, Y, X(t), \beta, t \in T\}$$

If we assume $\epsilon(t)$ and ϵ are Gaussian and independent, then this is proportional to

$$\prod_{t \in T} \exp(-(Y(t) - Y + \beta'(X - X(t)))'(Y(t) - Y + \beta'(X - X(t))))$$

which equals

$$\exp(-\sum_{t \in T}(Y(t) - Y + \beta'(X - X(t)))'(Y(t) - Y + \beta'(X - X(t))))$$

Define $y(t) = Y(t) - Y$, $x(t) = X(t) - X$, and $\epsilon^*(t) = \epsilon(t) - \epsilon$ so that $y(t) = \beta' x(t) + \epsilon^*(t)$. Since ϵ is Gaussian with variance v and independent of $\epsilon(t)$, $\epsilon(t) - \epsilon$ has variance $2v$. As a result, we have

$$\begin{aligned} \Pr.\{Y(t)|XY, X(t), t \in T\} &\propto \int_\beta \prod_{t \in T} \Pr.\{\epsilon(t)\} f(\beta|X, Y) d\beta \\ &= \int_\beta \exp(-\sum_{t \in T} \frac{(y(t) - \beta' x(t))^2}{4v}) f(\beta|X, Y) d\beta \end{aligned}$$

Define $EQ = \sum_{t \in T} \frac{Q(t)}{|T|}$. Then the above equation becomes

$$\begin{aligned} \Pr.\{Y(t)|X, Y, X(t), \quad t \quad &\in T\} \propto \int_\beta \exp(-\frac{E(y(t) - \beta' x(t))^2}{4Tv}) f(\beta|X, Y) d\beta \\ &= \int_\beta \exp(-\frac{T}{4v}(E(y^2(t)) - 2\beta' E(y(t)x(t)) + \beta' E(x(t)x'(t))\beta)) f(\beta|X, Y) d\beta \end{aligned}$$

Letting $V_x = E(x(t)x'(t))$ and $r = E(y(t)x(t))$ gives

$$\int_\beta \exp(-\frac{T}{2v}[(\beta - r'V_x^{-1})V_x(\beta - V_x^{-1}r) + E(y^2(t)) - r'V_x^{-1}V_xV_x^{-1}r]) f(\beta|X, Y) d\beta$$

Assuming β conditionally uniform given X, Y and integrating gives

$$\Pr.\{Y(t)|X, Y, X(t), t \in T\} \propto \frac{T}{4v}|V_x| \exp(-\frac{T}{4v}(E(y^2(t)) - r'V_x^{-1}r))$$

Rewriting $E(y^2(t)), V_x$, and r explicitly in terms of Y and X gives

$$E(y^2) = E((Y(t) - Y)^2) = (Y - E(Y))^2 + E((Y(t) - E(Y))^2)$$

$$V_x = (X - E(X))'(X - E(X)) + E[(X(t) - E(X))'(X(t) - E(X)))]$$

$$E(y(t)x(t)) = (Y - E(Y))(X - E(X)) + E((X(t) - E(X(t)))(Y(t) - E(Y(t))))$$

Let $V = [(X(t) - E(X))\hat{}(X(t) - E(X))]$ describe the variation in the simulation inputs. Using a Cholesky decomposition of V, define lower triangular matrix S such that $S'S = V$, and define s such that $s^2 = E[(Y(t) - Y)^2]$. Define renormalized variables $y = \frac{Y - E(Y)}{s}$ and $x = S^{-1}(X - E(X))$. In terms of these variables, we have

$$E(y^2(t)) = s^2(y^2 + 1)$$

$$V_x = S(x'x + I)S'$$

Defining $C = s^{-1}rS^{-1}$ gives

$$r = s[yx + C]S$$

Also define $V = [x'x + I]$. Substituting into the expression for $\Pr.\{Y(t)|X, Y, X(t), t \in T\}$ gives

$$\begin{aligned}\frac{T}{4v}|V_x| \exp(-\frac{T}{4v}(s^2(y^2+1) - s[yx + C]S[S(x'x + I)S']^{-1}S[xy + C]s)) &\propto \\ |V_x| \exp(-\frac{Ts^2}{4v}(y^2 - [yx + C][I + x'x]^{-1}[xy + C])) &= \\ |V_x| \exp(-\frac{Ts^2}{4v}(y(1 - x[I + x'x]^{-1}x)y) - 2C[I + x'x]^{-1}xy - C[I + x'x]^{-1}C)\end{aligned}$$

To evaluate this expression, note the following three tautologies

$$\begin{aligned}1 - x[x'x + I]^{-1}x &= 1 - x[1 - x'x + x'xx'x - x'xx'xx'x + ...]x' \\ &= 1 - xx' + x'xx'x - x'xx'x + ... \\ &= 1/(1 + x'x)\end{aligned}$$

$$\begin{aligned}C[I + x'x]^{-1}xy &= C[I - x'x + x'xx'' - ...]xy \\ &= [Cx - Cx'x'x + Cx'(x'xx'x) - ...]y \\ &= Cxy[1 - x'x + x'x - ...] \\ &= Cxy/[1 + x'x]\end{aligned}$$

$$\begin{aligned}C[I + x'x]^{-1}C &= C[I - x'x + x'xx'x - ...]C \\ &= C'C - Cx'xC' + Cx'(xx')xC' - ... \\ &= (C'C) - (Cx')^2[1 - (xx') + (xx'xx') - ...] \\ &= C'C - \frac{(Cx')^2}{1 + xx'}\end{aligned}$$

Substituting these three tautologies into the expression gives

$$\begin{aligned}\Pr.\{Y(t)|X, Y, X(t), t \in T\} &= |V_x| \exp(-\frac{Ts^2}{4v}(\frac{y^2}{1 + x'x} - \frac{2Cxy}{1 + x'x} + \frac{(Cx')^2}{1 + xx'} - C'C)) \\ &= |V_x| \exp(-\frac{Ts^2}{4v}(\frac{(y - Cx)^2}{1 + x'x}))\end{aligned}$$

Since $y = \frac{Y - E(Y)}{s}$, $S^{-1}(X - E(X)) = x$ and $r = sCS$, this becomes

$$|V_x| \exp(-\frac{Ts^2((Y - E(Y))/s - (r/s)(X - E(X))^2)}{4v[1 + x'x]}$$

$$= |I + (X - E(X))S^{-1}S^{-1}(X - E(X))|^{-1} \exp(-\frac{T(Y - E(Y) - r(X - E(X)))^2}{4v[1 + (X - E(X))S^{-1}S^{-1}(X - E(X))]})$$

Since $V^{-1} = S^{-1}S^{-1}$, our final result for $\Pr.\{Y(t)|X, Y, X(t), t \in T\}$ becomes

$$|I + (X - E(X))V^{-1}(X - E(X))|^{-1} \exp(-\frac{T(Y - \hat{Y} - r(X - E(X)))^2}{4v[1 + (X - E(X))V^{-1}(X - E(X))]})$$

which gives the result in the text.

5.5 Appendix V

To compute the marginal distribution of x, we will integrate y out of the joint distribution of y and x.

$$\begin{aligned} \Pr.\{x, y|\hat{X}, \hat{Y}\} &\propto \frac{1}{v(x)} \exp(-\frac{1}{2}[\frac{y^2}{v(x)} - \frac{2y(rx)}{v(x)} + \frac{(rx)^2}{v(x)}] \exp(-\frac{(y - m_{y|x})^2}{2\Sigma_{yy|x}}) \Pr.\{x\} \\ &\propto \frac{1}{v(x)} \exp(-\frac{1}{2}[(\frac{1}{v(x)} + \frac{1}{\Sigma_{yy|x}})y^2 - 2(\frac{rx}{v(x)} + \frac{m_{y|x}}{\Sigma_{yy|x}}y + \frac{(r'x)^2}{v(x)} + \frac{m_{y|x}^2}{\Sigma_{yy|x}}]) \end{aligned}$$

We now rearrange terms to create a quadratic term depending on y and a separate term independent of y, in order to integrate out y.

$$\Pr.\{x, y|\hat{X}, \hat{Y}\} \propto \frac{1}{v(x)} \exp(-\frac{1}{2}(\frac{1}{v(x)} + \frac{1}{\Sigma_{yy|x}})(y - \frac{\frac{r'x}{v(x)} + \frac{m_{y|x}}{\Sigma_{yy|x}}}{\frac{1}{v(x)} + \frac{1}{\Sigma_{yy|x}}})^2 - \frac{1}{2}\frac{(r'x)^2}{v(x)} - \frac{1}{2}\frac{m_{y|x}^2}{\Sigma_{yy|x}} + \frac{(\frac{(r'x)}{v(x)} + \frac{m_{y|x}}{\Sigma_{yy|x}})^2}{\frac{1}{v(x)} + \frac{1}{\Sigma_{yy|x}}})$$

Integrating out y gives

$$\Pr.\{x|\hat{X}, \hat{Y}\} \propto \frac{\frac{1}{v(x)}}{\frac{1}{v(x)} + \frac{1}{\Sigma_{yy|x}}} \exp(-\frac{1}{2}\frac{(r'x)^2}{v(x)} - \frac{1}{2}\frac{m_{y|x}^2}{\Sigma_{yy|x}} + \frac{(\frac{(r'x)}{v(x)} + \frac{m_{y|x}}{\Sigma_{yy|x}})^2}{\frac{1}{v(x)} + \frac{1}{\Sigma_{yy|x}}})$$

Let $Q = -\frac{1}{2}\frac{(r'x)^2}{v(x)} - \frac{1}{2}\frac{m_{y|x}^2}{\Sigma_{yy|x}} + \frac{(\frac{(r'x)}{v(x)} + \frac{m_{y|x}}{\Sigma_{yy|x}})^2}{\frac{1}{v(x)} + \frac{1}{\Sigma_{yy|x}}})$ so that $\Pr.\{x|\hat{X}, \hat{Y}\} = \exp(Q)$. We now rearrange the expression for Q to get

$$\begin{aligned} Q &= -\frac{1}{2}\frac{(r'x)^2}{v(x)} - \frac{1}{2}\frac{m_{y|x}^2}{\Sigma_{yy|x}} + \frac{(\frac{(r'x)^2}{v^2(x)} + 2\frac{(r'x)m_{y|x}}{v(x)\Sigma_{yy|x}} + \frac{m_{y|x}^2}{\Sigma_{yy|x}^2})^2}{\frac{1}{v(x)} + \frac{1}{\Sigma_{yy|x}}} \\ &= -\frac{1}{2(\frac{1}{v(x)} + \frac{1}{\Sigma_{yy|x}})}[-\frac{(r'x)^2}{v^2(x)} - \frac{(r'x)}{v(x)\Sigma_{yy|x}} - \frac{m_{y|x}^2}{\Sigma_{yy|x}^2} - \frac{m_{y|x}^2}{v(x)\Sigma_{yy|x}} + (\frac{(r'x)^2}{v^2(x)} + 2\frac{(r'x)m_{y|x}}{v(x)\Sigma_{yy|x}} + \frac{m_{y|x}^2}{\Sigma_{yy|x}^2})^2] \\ &= -\frac{1}{2(\frac{1}{v(x)} + \frac{1}{\Sigma_{yy|x}})}[-\frac{(r'x)}{v(x)\Sigma_{yy|x}} - \frac{m_{y|x}^2}{v(x)\Sigma_{yy|x}} + 2\frac{(r'x)m_{y|x}}{v(x)\Sigma_{yy|x}}] \\ &= -\frac{1}{2(\frac{1}{v(x)} + \frac{1}{\Sigma_{yy|x}})}[\frac{(r'x)}{v(x)} - \frac{m_{y|x}^2}{\Sigma_{yy|x}}]^2 \end{aligned}$$

which, with minor rearranging, becomes the result in the text.