EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 1409 / DECEMBER 2011

ANALYSIS OF VARIANCE FOR BAYESIAN INFERENCE

by John Geweke

and Gianni Amisano

EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 1409 / DECEMBER 2011

ANALYSIS OF VARIANCE FOR BAYESIAN INFERENCE[1]

by John Geweke[2]
and Gianni Amisano[3]

In 2011 all ECB publications feature a motif taken from the €100 banknote.

1 Support from Australian Research Council grant 110104732 is gratefully acknowledged.
2 University of Technology Sydney, P.O. Box 123, Broadway, NSW 2007, Australia, Erasmus University, Netherlands and University of Colorado, USA; e-mail: John.Geweke@uts.edu.au
3 European Central Bank, Kaiserstrasse 29, 60311 Frankfurt am Main, Germany; e-mail: gianni.amisano@ecb.europa.eu

Address
Kaiserstrasse 29
60311 Frankfurt am Main, Germany

Postal address
Postfach 16 03 19
60066 Frankfurt am Main, Germany

Telephone
+49 69 1344 0

Internet
http://www.ecb.europa.eu

Fax
+49 69 1344 6000

Information on all of the papers published in the ECB Working Paper Series can be found on the ECB's website, http://www.ecb.europa.eu/pub/scientific/wps/date/html/index.en.html

ISSN 1725-2806 (online)

CONTENTS

Abstract

This paper develops a multi-way analysis of variance for non-Gaussian multivariate distributions and provides a practical simulation algorithm to estimate the corresponding components of variance. It specifically addresses variance in Bayesian predictive distributions, showing that it may be decomposed into the sum of extrinsic variance, arising from posterior uncertainty about parameters, and intrinsic variance, which would exist even if parameters were known. Depending on the application at hand, further decomposition of extrinsic or intrinsic variance

(or both) may be useful. The paper shows how to produce simulation-consistent estimates of all of these components, and the method demands little additional effort or computing time beyond that already invested in the posterior simulator. It illustrates the methods using a dynamic stochastic general equilibrium model of the US economy, both before and during the global financial crisis.

Keywords: analysis of variance, Bayesian inference, predictive distributions, posterior simulation

JEL codes: C11, C53

Non-technical summary

This paper follows the Bayesian paradigm of integrating information. In this context it provides a new decomposition of variance for predictive distributions. Here are some of the questions that motivate this research.

1. In prediction and other decision-making situations, econometricians sometimes replace parameters with point estimates rather than using full posterior or predictive distributions. This eliminates the contribution of parameter uncertainty to the distribution relevant to the decision at hand. The impact could be anywhere from an academic footnote to a disastrous outcome in the real world. Can Bayesian analysis provide systematic guidance on this point?

2. Understanding complex interactions in large models and their impact on predictive distributions relevant for decision-making is an essential component in the improvement of decision support. Are there tools that Bayesians could employ on a regular basis to identify links between model components and features of predictive distributions?

3. Emphasis on economic prediction over longer horizons has never been greater, due to pressing problems such as structural financial problems in many countries. In models that draw on historical time series, predictions naturally tend to be driven more by actual behaviour in the near term and more by aspects of model specification in the longer term. Can the structure of the impact of alternative information sources be decomposed systematically over a prediction horizon?

We believe that the answers to all three of these, and similar, questions are yes,.and this paper provides additions to the Bayesian econometrician's set of tools to address such questions. The basic approach is to use the law of total variance iteratively to identify multiple sources of variance.

In the paper we take up details pertinent to Bayesian analysis, making two specific contributions. The .first is the decomposition of the predictive distribution into extrinsic variance, i.e. due to parameter uncertainty, and intrinsic variance, which would exist even if parameters were known.

The second contribution is to show that given a posterior simulator, very little additional effort or computing time is required to produce simulation-consistent estimates of these variance components.

The paper concludes with an illustrative application of the techniques developed in the paper, to predictive distributions from a widely used dynamic stochastic general equilibrium (DSGE) model just before and then during the recent financial crisis in the U.S. This illustration shows, among other things, that the understatement of predictive variance inherent in replacing parameters with known values is systematically greater in volatile than quiescent times.

1 Introduction

Bayesian inference is a remarkable intellectual tool that can integrate information from widely different sources and draw out its implications for specific decisions under consideration. That many economists recognize this fact is due in no small part to the work and unrelenting efforts of Arnold Zellner over his long and rich career. While these activities took many forms, three of the most important were his 1971 book *Bayesian Inference in Econometrics* (Zellner, 1971), the Seminar on Bayesian Inference in Econometrics and Statistics that convened regularly in the following quarter-century, and the International Society for Bayesian Analysis which he was instrumental in founding in the early 1990's.

This paper follows the Bayesian paradigm of integrating information. In this context it provides a new decomposition of variance for predictive distributions. Here are some of the questions that motivate this research.

1. In prediction and other decision-making situations, econometricians sometimes replace parameters with point estimates rather than using full posterior or predictive distributions. This eliminates the contribution of parameter uncertainty to the distribution relevant to the decision at hand. The impact could be anywhere from an academic footnote to a disastrous outcome in the real world. Can Bayesian analysis provide systematic guidance on this point?

2. Understanding complex interactions in large models and their impact on predictive distributions relevant for decision-making is an essential component in the improvement of decision support. Are there tools that Bayesians could employ on a regular basis to identify links between model components and features of predictive distributions?

3. Emphasis on economic prediction over longer horizons has never been greater, due to pressing problems including climate change, aging population and structural financial problems in many countries. In models that draw on historical time series, predictions naturally tend to be driven more by actual behavior in the near term and more by aspects of model specification in the longer term. Can the structure of the impact of alternative information sources be decomposed systematically over a prediction horizon?

We believe that the answers to all three of these, and similar, questions are "yes," and this paper provides additions to the Bayesian econometrician's set of tools to address such questions. The basic approach is to use the law of total variance iteratively to identify multiple sources of variance, and these ideas are developed in Section 2. If the relevant statistical structure of information and the problem at hand is Gaussian this amounts to no more than analysis of variance with multiple factors, well understood for a century and standard training in statistics. But contemporary econometric models are fundamentally nonlinear, and our approach accounts for that. To the best of our knowledge, this is a new contribution.

This decomposition is a feature of the population. Simulations from the population provide estimates of the decomposition. Section 3 lays out the details, and shows that the estimates are simulation consistent: i.e., they converge almost surely to population counterparts as the size of the simulation sample increases.

Section 4 takes up details pertinent to Bayesian analysis, making two specific contributions. The first is the decomposition of the predictive distribution: first, into extrinsic variance, that which is due to parameter uncertainty, and intrinsic variance, which would exist even if parameters were known; and, second, the further decomposition of extrinsic and intrinsic variance, either (or both) of which may be relevant depending on the application at hand. The second contribution in this section is to show that given a posterior simulator, very little additional effort or computing time is required to produce simulation-consistent estimates of these variance components.

The paper concludes with an illustrative application of the techniques developed in the paper, to predictive distributions from a widely used dynamic stochastic general equilibrium model just before and then during the recent financial crisis in the U.S. This illustration shows, among other things, that the understatement of predictive variance inherent in replacing parameters with known values is systematically greater in volatile than quiescent times. We believe that this finding reflects a general principle that should emerge in other applications as well.

The ideas in this paper have more than one intellectual heritage. From the Bayesian perspective of Section 4, which motivates the work, the decomposition of predictive variance into extrinsic and intrinsic components is a natural outgrowth of prior and posterior predictive analysis, which can be traced to Good (1956), Roberts (1965) and Box (1980) and is fully described by Lancaster (2004), Sections 2.4-2.5, and Geweke (2005), Section 8.3. An immediate precursor of the work in Section 4 is the idea of prior predictive analysis of variance mentioned in passing in Geweke (2010), Section 4.2.3. From the distribution theory perspective of Section 2, all of the decompositions of variance here are repeated applications of the law of total variance, which is stated at the start of that section. Virtually all of the applications of this idea in traditional analysis of variance have been in the context of linear (Gaussian) models, and that treatment is insufficient for the general case as well as for the application of the law of total variance in the context of posterior distributions. We are not aware of expositions of these ideas at this level of generality, or exploitation of the modification of simulation with balanced design and multi-factor analysis with no interactions to access these decompositions of variance as developed in Section 3.

2 Population

Let U be a random vector defined in the usual way, partitioned as $U' = (Y', Z')$, where Y and Z are vectors. The following result is well known (e.g. Weiss, 2005, pp 385-386).

Proposition 1 *(Law of Total Variance) Suppose that* $U' = (Y', Z')$ *is an* $(E, \mathcal{E})$*-valued random vector defined on a probability space* $(\Omega, \mathcal{F}, P)$*, and that* $\mathrm{var}(U)$ *exists. Then*

$$\mathrm{var}(Y) = \mathrm{E}_Z[\mathrm{var}(Y|Z)] + \mathrm{var}_Z[\mathrm{E}(Y|Z)]. \tag{1}$$

The next three results are contingent on:

Condition 1 $V = (W', Y', Z')'$ *is an* $(E, \mathcal{E})$*-valued random vector defined on a probability space* $(\Omega, \mathcal{F}, P)$ *and* $\mathrm{var}(V)$ *exists.*

The following two extensions of Proposition 1 are immediate.

Proposition 2 *Given Condition 1, denote by* $\mathbf{w}$ *any point in the support of the marginal distribution of* W*. Then*

$$\mathrm{var}(Y \mid W = \mathbf{w}) = \mathrm{E}_Z[\mathrm{var}(Y|Z, W = \mathbf{w})] + \mathrm{var}_Z[\mathrm{E}(Y|Z, W = \mathbf{w})]. \tag{2}$$

Proof. In Proposition 1 replace the distribution of U with the distribution of V conditional on $W = \mathbf{w}$. ■

Proposition 3 *Given Condition 1,*

$$\mathrm{E}_W[\mathrm{var}(Y \mid W)] = \mathrm{E}_{W,Z}[\mathrm{var}(Y \mid W, Z)] + \mathrm{E}_W\{\mathrm{var}_Z[\mathrm{E}(Y \mid W, Z)]\}. \tag{3}$$

Proof. Integrating both sides of (2) with respect to the measure dP_W yields the result. ■

Proposition 3 leads directly to the following result, which in turn provides the foundation for the rest of the methods discussed in this paper.

Proposition 4 *Given Condition 1,*

$$\mathrm{var}(Y) = \mathrm{E}_{W,Z}[\mathrm{var}(Y \mid W, Z)] + \mathrm{E}_Z\{\mathrm{var}_W[\mathrm{E}(Y \mid W, Z)]\} + \mathrm{var}_Z[\mathrm{E}(Y \mid Z)]. \tag{4}$$

Proof. In (3) reverse the roles of W and Z to write

$$\mathrm{E}_Z[\mathrm{var}(Y \mid Z)] = \mathrm{E}_{W,Z}[\mathrm{var}(Y \mid W, Z)] + \mathrm{E}_Z\{\mathrm{var}_W[\mathrm{E}(Y \mid W, Z)]\}. \tag{5}$$

Then substitute (5) in (1) to obtain (4). ■

Next, suppose:

Condition 2 X *is an* $(E, \mathcal{E})$*-valued random vector defined on a probability space* $(\Omega, \mathcal{F}, P)$ *and* $\mathrm{var}(X)$ *exists.*

Let X be partitioned

$$X' = (X_1', \ldots, X_n'). \tag{6}$$

For $n = 2$, taking $Y = X_1$ and $Z = X_2$ in (1) leads to

$$\mathrm{var}(X_1) = \mathrm{E}_{X_2}[\mathrm{var}(X_1|X_2)] + \mathrm{var}_{X_2}[\mathrm{E}(X_1|X_2)] \tag{7}$$

For the case $n = 3$, taking $Y = X_1$, $W = X_3$ and $Z = X_2$ in (4) yields

$$\begin{aligned} \mathrm{var}(X_1) &= \mathrm{E}_{X_2,X_3}[\mathrm{var}(X_1 \mid X_2, X_3)] + \mathrm{E}_{X_2}\{\mathrm{var}_{X_3}[\mathrm{E}(X_1 \mid X_2, X_3)]\} \\ &\quad + \mathrm{var}_{X_2}[\mathrm{E}(X_1 \mid X_2)]. \end{aligned} \tag{8}$$

Now apply Proposition 3 to the leading term of (8), with $Y = X_1$, $W = (X_2, X_3)$ and $Z = X_4$, to obtain

$$\begin{aligned} \mathrm{var}(X_1) &= \mathrm{E}_{X_2,X_3,X_4}[\mathrm{var}(X_1 \mid X_2, X_3, X_4)] \\ &\quad + \mathrm{E}_{X_2,X_3}\{\mathrm{var}_{X_4}[\mathrm{E}(X_1 \mid X_2, X_3, X_4)]\} \\ &\quad + \mathrm{E}_{X_2}\{\mathrm{var}_{X_3}[\mathrm{E}(X_1 \mid X_2, X_3)]\} + \mathrm{var}_{X_2}[\mathrm{E}(X_1 \mid X_2)]. \end{aligned} \tag{9}$$

This procedure can be iterated, as follows.

Proposition 5 *Given Condition 2 and the partition (6),*

$$\begin{aligned} \mathrm{var}(X_1) &= \mathrm{E}_{X_2,\ldots,X_n}[\mathrm{var}(X_1 \mid X_2, \ldots, X_n)] \\ &\quad + \mathrm{E}_{X_2,\ldots,X_{n-1}}\{\mathrm{var}_{X_n}[\mathrm{E}(X_1 \mid X_2, \ldots, X_n)]\} \\ &\quad + \ldots + \mathrm{E}_{X_2,\ldots,X_j}\left\{\mathrm{var}_{X_{j+1}}[\mathrm{E}(X_1 \mid X_2, \ldots, X_{j+1})]\right\} \\ &\quad + \ldots + \mathrm{E}_{X_2}\{\mathrm{var}_{X_3}[\mathrm{E}(X_1 \mid X_2, X_3)]\} + \mathrm{var}_{X_2}[\mathrm{E}(X_1 \mid X_2)]. \\ &= \mathrm{E}_{X_2,\ldots,X_n}[\mathrm{var}(X_1 \mid X_2, \ldots, X_n)] \qquad (10) \\ &\quad + \sum_{j=1}^{n-1} \mathrm{E}_{X_2,\ldots,X_j}\left\{\mathrm{var}_{X_{j+1}}[\mathrm{E}(X_1 \mid X_2, \ldots, X_{j+1})]\right\}. \qquad (11) \end{aligned}$$

Proof. The case $n = 4$ is (9). Apply Proposition 3 to the leading term (10), with $W = (X_2, \ldots, X_n)$, $Z = X_{n+1}$ and $Y = X_1$. This leads to the substitution

$$\begin{aligned} &\mathrm{E}_{X_2,\ldots,X_{n+1}}[\mathrm{var}(X_1 \mid X_2, \ldots, X_{n+1})] \\ &+ \mathrm{E}_{X_2,\ldots,X_n}\left\{\mathrm{var}_{X_{n+1}}[\mathrm{E}(X_1 \mid X_2, \ldots, X_{n+1}) \mid X_2, \ldots, X_{n+1}]\right\}. \end{aligned}$$

for (10). By induction, the proposition is true. ■

The foregoing results are quite general, depending only on Conditions 1 and 2. In the special case that P in Condition 2 is Gaussian, we have

$$x_1 = \delta + \sum_{j=2}^{n} \beta_j x_j + \varepsilon$$

where ε is independent of $(x_2,\ldots,x_n)$, $\mathrm{E}(x_1 \mid x_2,\ldots,x_n) = \delta + \sum_{j=2}^n \beta_j x_j$, $\mathrm{E}(\varepsilon) = \mathrm{E}(\varepsilon \mid x_2,\ldots,x_n) = 0$, and $\mathrm{var}(\varepsilon) = \mathrm{var}(x_1 \mid x_2,\ldots,x_n)$. Let

$$\eta_j = x_j - \mathrm{E}(x_j \mid x_1,\ldots,x_{j-1}) \ \ (j = 2,\ldots,n).$$

Then there is a one-to-one linear transformation between $(x_2,\ldots,x_j)$ and $\left(\eta_2,\ldots,\eta_j\right)$. In particular,

$$x_1 = \sum_{j=2}^{n} \alpha_j \eta_j + \varepsilon,$$

where $\alpha_j = cov\left(x_1,\eta_j\right)/\mathrm{var}\left(\eta_j\right)$ and $\mathrm{var}(x_1) = \sum_{j=2}^n \alpha_j^2 \mathrm{var}\left(\eta_j\right) + \mathrm{var}(\varepsilon)$. Moreover,

$$\mathrm{E}(x_1 \mid x_2,\ldots,x_{j+1}) = \mathrm{E}(x_1) + \sum_{i=1}^{j+1} \alpha_i \eta_i$$

and

$$\begin{aligned} & \mathrm{var}_{x_{j+1}}\left[\mathrm{E}(x_1 \mid x_2,\ldots,x_{j+1}) \mid x_2,\ldots,x_j\right] \\ = \ & \mathrm{var}_{x_{j+1}}\left(\sum_{i=2}^{j+1} \alpha_i \eta_i \mid \eta_2,\ldots,\eta_j\right) = \alpha_{j+1}^2 \mathrm{var}\left(\eta_{j+1}\right). \end{aligned}$$

None of these terms involve x_j or η_j, because in the Gaussian distribution conditional variances do not depend on the values of the variables conditioned upon.

Thus the leading term (10) in Proposition 5 reduces to $\mathrm{var}(\varepsilon)$ in the Gaussian case, and term j in the sum (11) is $\alpha_{j+1}^2 \mathrm{var}\left(\eta_{j+1}\right)$. Were we to divide the equation in Proposition 5 by $\mathrm{var}(x_1)$, then (10) would be $1 - R^2$, where R^2 is the coefficient of multiple correlation between x_1 and $(x_2,\ldots,x_n)$; the term j in (11) would be the increment to population R^2 when x_{j+1} is introduced into the set of regressors that already contains $x_1,\ldots,x_j$.

3 Simulation

All of these population decompositions have analogs in simulation. This is important for the Bayesian applications that motivate this work. What follows takes the components $X_1,\ldots,X_n$ to be scalars. The vector case exactly parallels the scalar case but entails more awkward and space-consuming notation.

Suppose that it is feasible to simulate

$$x_j^{(m)} \sim p(x_j \mid x_2,\ldots,x_{j-1}) \quad (m = 1,2,\ldots; j = 2,\ldots,n), \tag{12}$$

where the case $j = 2$ is unconditional simulation, as well as

$$x_1^{(m)} \sim p(x_1 \mid x_2,\ldots,x_n) \quad (m = 1,2,\ldots). \tag{13}$$

We shall suppose that the random sequences $\left\{x_j^{(m)}\right\}$ are ergodic. Then there are simulations that parallel the decompositions in the previous section, which provide method of moments estimates of the terms in the population decomposition given in Proposition 5.

For the simplest case $n = 2$, this involves simulating $x_2^{(m_2)}$ $(m_2 = 1, \ldots, M_2)$, and conditional on each $x_2^{(m_2)}$ simulating $x_1^{(m_1,m_2)}$ $(m_1 = 1, \ldots, M_1)$ from the distribution of x_1 conditional on $x_2^{(m_2)}$. Denote

$$\overline{x}_1^{(m_2)} = M_1^{-1} \sum_{m_1=1}^{M_1} x_1^{(m_1,m_2)} \text{ and } \overline{x}_1 = M_2^{-1} \sum_{m_2=1}^{M_2} \overline{x}_1^{(m_2)} = (M_1 M_2)^{-1} \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} x_1^{(m_1,m_2)}.$$

Then

$$\begin{aligned}
(M_1 M_2)^{-1} \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \left(x_1^{(m_1,m_2)} - \overline{x}_1\right)^2 &= (M_1 M_2)^{-1} \sum_{m_2=1}^{M_2} \sum_{m_1=1}^{M_1} \left(x_1^{(m_1,m_2)} - \overline{x}_1^{(m_2)}\right)^2 \\
&\quad + M_2^{-1} \sum_{m_2=1}^{M_2} \left(\overline{x}_1^{(m_2)} - \overline{x}_1\right)^2. \qquad (14)
\end{aligned}$$

The terms in this relationship constitute consistent estimates of the terms of the equation in Proposition 5 for the case $n = 2$, equivalently for the terms of the equation in Proposition 1 with $Y = x_1$ and $Z = x_2$.

When the distribution of (x_1, x_2) is Gaussian, or more generally when the distribution of $x_1 \mid x_2$ is conditionally homoscedastic, we need only a single simulation $\left(x_1^{(m)}, x_2^{(m)}\right)$ from the joint distribution of x_1 and x_2 to consistently estimate the terms in the decomposition. The fact that in the more general case the conditional variance of x_1 depends in a nontrivial way on the value of x_2 necessitates the double simulation in m_1 and m_2.

Next consider the case $n = 3$. This case turns out to be significant, because the calculations here are essentially those that are required for the general case. Corresponding to each simulation $x_2^{(m_2)}$ $(m_2 = 1, \ldots, M_2)$ from the unconditional distribution of x_2 there are simulations $x_3^{(m_2,m_3)}$ $(m_3 = 1, \ldots, M_3)$ from the distribution of x_3 conditional on $x_2^{(m_2)}$. Then corresponding to each simulation $x_2^{(m_2,m_2)}$ there are simulations $x_1^{(m_1,m_2,m_3)}$ $(m_1 = 1, \ldots, M_1)$ from the distribution of x_1 conditional on $\left(x_2^{(m_2,m_3)}, x_3^{(m_3)}\right)$.

Corresponding to these simulations define the (conditional) sample means

$$\begin{aligned}
\overline{x}_1^{(m_2,m_3)} &= M_1^{-1} \sum_{m_1=1}^{M_1} x_1^{(m_1,m_2,m_3)} \quad (m_2 = 1, \ldots, M_2; m_3 = 1 \ldots, M_3), \\
\overline{x}_1^{(m_2)} &= M_3^{-1} \sum_{m_3=1}^{M_3} \overline{x}_1^{(m_2,m_3)} \quad (m_2 = 1, \ldots, M_2), \\
\overline{x}_1 &= M_2^{-1} \sum_{m_2=1}^{M_3} \overline{x}_2^{(m_3)}.
\end{aligned}$$

These are simulation-consistent approximations of $\mathrm{E}\left(x_1 \mid x_2^{(m_2)}, x_3^{(m_2,m_3)}\right)$, $\mathrm{E}\left(x_1 \mid x_2^{(m_2)}\right)$ and $\mathrm{E}(x_1)$, respectively. By "simulation-consistent" we refer to almost sure limits as $M_j \to \infty$ $(j = 1, 2, 3)$. Relative rates of divergence of the terms M_j do not matter.

The simulation-consistent method of moments approximation of $\mathrm{var}(x_1)$, the left side of (8), is

$$(M_1 M_2 M_3)^{-1} \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \sum_{m_3=1}^{M_3} \left(x_1^{(m_1,m_2,m_3)} - \overline{x}\right)^2. \tag{15}$$

Corresponding to the three terms on the right side of (8), the simulation-consistent approximation of $\mathrm{E}_{X_2,X_3}[\mathrm{var}(X_1 \mid X_2, X_3)]$ is

$$(M_2 M_3)^{-1} \sum_{m_2=1}^{M_2} \sum_{m_3=1}^{M_3} \left[M_1^{-1} \sum_{m_1=1}^{M_1} \left(x_1^{(m_1,m_2,m_3)} - \overline{x}_1^{(m_2,m_3)}\right)^2 \right]; \tag{16}$$

the simulation-consistent approximation of $\mathrm{E}_{X_2}\{\mathrm{var}_{X_3}[\mathrm{E}(X_1 \mid X_2, X_3)]\}$ is

$$M_2^{-1} \sum_{m_2=1}^{M_2} \left[M_3^{-1} \sum_{m_3=1}^{M_3} \left(\overline{x}_1^{(m_2,m_3)} - \overline{x}_1^{(m_3)}\right)^2 \right]; \tag{17}$$

and the simulation consistent approximation of $\mathrm{var}_{X_2}[\mathrm{E}(X_1 \mid X_2)]$ is

$$M_2^{-1} \sum_{m_2=1}^{M_2} \left(\overline{x}_1^{(m_2)} - \overline{x}_1\right)^2. \tag{18}$$

Consistent with the identity in (8), the sum of (16) – (18) is (15). The computations (16)-(18) are implemented with three lines of code in the Matlab function provided in the Appendix of this paper. Taking $M_1 = M_2 = M_3 = 100$ is adequate in our experience, as indicated by the finding that computations starting with different seeds of the random number generator produce quite similar results. Even in the case in which x_1 is replaced by a vector of modest size (e.g., 10 or 12 components) the computations require well under one second.

This process could be iterated to the general case of Proposition 2. The outcome would be a simulation sample ideally suited to multi-factor analysis of variance with a balanced design, but the simulation requirements are overwhelming e.g. the number of simulations required is $\left(M^{(n+1)} - M^2\right) / (M - 1)$ in the case where $M_1 = \ldots, M_n = M$. This is impractical for n much larger than 3 and M sufficiently large to provide reliable approximations. Instead consider term j of (11),

$$\mathrm{E}_{X_2,\ldots,X_j}\left\{\mathrm{var}_{X_{j+1}}[\mathrm{E}(X_1 \mid X_2, \ldots, X_{j+1})]\right\}$$

and note that we can use the approach in the previous paragraph, replacing x_2 with $(x_2, \ldots, x_j)$ and x_3 with x_{j+1}. The requisite simulations are all possible by virtue of the assumption that simulations are possible from the distributions (12) – (13), made at beginning of this section. In the case $M_1 = \ldots, M_n = M$, this entails nM^3 simulations, a number that is linear in n and is reasonable in the applications that motivate this work.

4 Application to Bayesian inference

These decompositions are useful tools in the interpretation of posterior distributions that are accessed by means of simulation, as is the case in most Bayesian work. Let $\mathbf{Y}_T = [\mathbf{y}_1, \ldots, \mathbf{y}_T]$ denote the observables (random *ex ante*) where T is the size of the sample, $\mathbf{Y}_T^o$ their observed values (fixed *ex post*), A the model, $\boldsymbol{\theta}_A \in \Theta_A$ the parameter vector, $p(\boldsymbol{\theta}_A \mid A)$ the prior density, and $p(\mathbf{Y}_T \mid \boldsymbol{\theta}_A, A)$ the distribution of observables conditional on the parameters. The posterior density of the parameters is $p(\boldsymbol{\theta}_A \mid \mathbf{Y}_T^o, A) \propto p(\boldsymbol{\theta}_A \mid A) \cdot p(\mathbf{Y}_T^o \mid \boldsymbol{\theta}_A, A)$, and we assume that a posterior simulator is available that generates an identically distributed ergodic process $\boldsymbol{\theta}_A^{(m)} \sim p(\boldsymbol{\theta}_A \mid \mathbf{Y}_T^o, A)$ $(m = 1, 2, 3, \ldots)$. For importance sampling, there are obvious modifications to the computation of simulation moments involving the weighting function, and these will also be consistent if the usual regularity conditions (Geweke, 2005, Theorem 4.2.2) are satisfied.

From a formal perspective, Bayesian inference is always undertaken to inform a policy decision. Let $\boldsymbol{\omega}$ denote the random vector pertinent to the loss function $L(\boldsymbol{\omega})$ governing the decision: e.g., for a central bank, $\boldsymbol{\omega}$ could consist of measures of output and inflation in some future quarters; for a retailer, $\boldsymbol{\omega}$ could be sales of specified products in specified markets; for a government agency seeking to adjust census count, $\boldsymbol{\omega}$ might contain measures of characteristics of the actual population; in a pure prediction problem $\boldsymbol{\omega}$ consists of future values $(\mathbf{y}_{T+1}, \ldots, \mathbf{y}_{T+H})$ over a specified horizon H. The model informs the policy decision if and only if it specifies the conditional distribution $p(\boldsymbol{\omega} \mid \mathbf{Y}_T, \boldsymbol{\theta}_A, A)$, and we assume that is the case. This conditional distribution could be degenerate: for example, if the decision involves testing the hypothesis $\boldsymbol{\theta}_A \in \Theta_{A1}$ then $\omega = I_{\Theta_{A1}}(\boldsymbol{\theta}_{A1})$. Simulation from $p(\boldsymbol{\omega} \mid \mathbf{Y}_T^o, \boldsymbol{\theta}_A, A)$ is generally straightforward and less demanding than simulating from the posterior distribution. In many of these specific instances, expected loss conditions on the prospective action taken by the decision-maker. For our purposes one of a number of alternative actions is then subsumed in A.

4.1 Extrinsic and intrinsic variance

The distribution relevant for decision-making conditions on the model A and data $\mathbf{Y}_T^o$,

$$p(\boldsymbol{\omega} \mid \mathbf{Y}_T^o, A) \propto \int_{\Theta_A} p(\boldsymbol{\theta}_A \mid \mathbf{Y}_T^o, A)\, p(\boldsymbol{\omega} \mid \mathbf{Y}_T^o, \boldsymbol{\theta}_A, A)\, d\boldsymbol{\theta}_A.$$

Given a posterior sample $\boldsymbol{\theta}_A^{(m_2)}$ $(m_2 = 1, \ldots, M_2)$, this distribution can be accessed by means of auxiliary simulations from the model for $\boldsymbol{\omega}$,

$$\boldsymbol{\omega}^{(m_1,m_2)} \sim p\left(\boldsymbol{\omega} \mid \mathbf{Y}_T^o, \boldsymbol{\theta}_A^{(m_2)}, A\right) \ \ (m_1 = 1, \ldots, M_1).$$

The corresponding sample moments of $\boldsymbol{\omega}$, and in particular the approximation

$$(M_1 M_2)^{-1} \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} L\left(\boldsymbol{\omega}^{(m_1,m_2)}\right)$$

of $\mathrm{E}\left[L\left(\boldsymbol{\omega}\right) \mid \mathbf{Y}_T^o, A\right]$ are simulation-consistent if and only if $M_2 \to \infty$; it is not necessary that M_1 increase at all, for this purpose, and indeed $M_1 = 1$ is sufficient.

From (7) with $X_1 = \boldsymbol{\omega}$ and $X_2 = \boldsymbol{\theta}_A$, and recognizing that the relevant distribution conditions on the data $\mathbf{Y}_T^o$ and the model specification A,

$$\operatorname{var}\left(\boldsymbol{\omega} \mid \mathbf{Y}_T^o, A\right) = \mathrm{E}_{\boldsymbol{\theta}_A}\left[\operatorname{var}\left(\boldsymbol{\omega} \mid \boldsymbol{\theta}_A, \mathbf{Y}_T^o, A\right)\right] + \operatorname{var}_{\boldsymbol{\theta}_A}\left[\mathrm{E}\left(\boldsymbol{\omega} \mid \boldsymbol{\theta}_A, \mathbf{Y}_T, A\right)\right]. \tag{19}$$

We refer to the first component on the right-hand side as the *intrinsic variance* of $\boldsymbol{\omega}$: it is the variation in $\boldsymbol{\omega}$ that would exist if one knew the parameter vector $\boldsymbol{\theta}_A$, averaged using the posterior distribution of $\boldsymbol{\theta}_A$. We refer to the second component as the *extrinsic variance* of $\boldsymbol{\omega}$: it is the variance in $\boldsymbol{\omega}$ that is due to not knowing $\boldsymbol{\theta}_A$. If the distribution of $\boldsymbol{\omega}$ conditional on $\mathbf{Y}_T$ and $\boldsymbol{\theta}_A$ is degenerate, as is the case with conventional tests of hypotheses about $\boldsymbol{\theta}_A$, then there is no intrinsic variance. If the prior distribution is dogmatic then there is no extrinsic variance. In most realistic cases both intrinsic and extrinsic variance are positive.

Making the corresponding substitution in the simulation (14)

$$\begin{aligned}
& (M_1M_2)^{-1}\left(\boldsymbol{\omega}^{(m_1,m_2)} - \overline{\boldsymbol{\varpi}}\right)\left(\boldsymbol{\omega}^{(m_1,m_2)} - \overline{\boldsymbol{\varpi}}\right)' \\
= \; & (M_1M_2)^{-1}\sum_{m_2=1}^{M_2}\sum_{m_1=1}^{M_1}\left(\boldsymbol{\omega}^{(m_1,m_2)} - \overline{\boldsymbol{\varpi}}^{(m_2)}\right)\left(\boldsymbol{\omega}^{(m_1,m_2)} - \overline{\boldsymbol{\varpi}}^{(m_2)}\right)' && (20) \\
& + M_1^{-1}\sum_{m_1=1}^{M_1}\left(\overline{\boldsymbol{\varpi}}^{(m_2)} - \overline{\boldsymbol{\varpi}}\right)\left(\overline{\boldsymbol{\varpi}}^{(m_2)} - \overline{\boldsymbol{\varpi}}\right)' && (21)
\end{aligned}$$

where

$$\overline{\boldsymbol{\varpi}}^{(m_2)} = M_1^{-1}\sum_{m_1=1}^{M_1}\boldsymbol{\omega}^{(m_1,m_2)} \quad \text{and} \quad \overline{\boldsymbol{\varpi}} = M_2^{-1}\sum_{m_2=1}^{M_2}\overline{\boldsymbol{\varpi}}^{(m_2)}.$$

As $M_1 \to \infty$ and $M_2 \to \infty$ (20) converges to $\mathrm{E}_{\boldsymbol{\theta}_A}\left[\operatorname{var}\left(\boldsymbol{\omega} \mid \boldsymbol{\theta}_A, A\right)\right]$ in (19) and (21) to $\operatorname{var}_{\boldsymbol{\theta}_A}\left[\mathrm{E}\left(\boldsymbol{\omega} \mid \boldsymbol{\theta}_A, \mathbf{Y}_T, A\right)\right]$.

4.2 Decomposition of intrinsic variance

Let the vector of interest be partitioned $\boldsymbol{\omega}' = (\boldsymbol{\omega}_1', \boldsymbol{\omega}_2')$, and suppose that it is feasible to simulate

$$\begin{aligned}
\boldsymbol{\omega}_1^{(m)} &\sim p\left(\boldsymbol{\omega}_1 \mid \mathbf{Y}_T^o, \boldsymbol{\theta}_A, A\right), && (22) \\
\boldsymbol{\omega}_2^{(m)} &\sim p\left(\boldsymbol{\omega}_2 \mid \boldsymbol{\omega}_1, \mathbf{Y}_T^o, \boldsymbol{\theta}_A, A\right). && (23)
\end{aligned}$$

For example in a pure prediction problem we might have $\boldsymbol{\omega}_1 = \mathbf{y}_{T+1}$ and $\boldsymbol{\omega}_2 = \mathbf{y}_{T+2}$ or $\boldsymbol{\omega}_2 = (1/4)\sum_{s=1}^{4}\mathbf{y}_{T+s}$; or, ω_1 could be a monetary policy instrument and $\boldsymbol{\omega}_2$ would consist of the remaining variables in a macroeconomic model. In the notation of the previous section, take $X_1 = \boldsymbol{\omega}_2$, $X_2 = \boldsymbol{\theta}_A$ and $X_3 = \boldsymbol{\omega}_1$. Then from the particular case (8) of Proposition 5,

$$
\begin{aligned}
\operatorname{var}(\boldsymbol{\omega}_2) &= \mathrm{E}_{\boldsymbol{\theta}_A,\boldsymbol{\omega}_1}\left[\operatorname{var}(\boldsymbol{\omega}_2 \mid \boldsymbol{\omega}_1, \boldsymbol{\theta}_A)\right] && (24)\\
&\quad + \mathrm{E}_{\boldsymbol{\theta}_A}\left\{\operatorname{var}_{\boldsymbol{\omega}_1}\left[\mathrm{E}(\boldsymbol{\omega}_2 \mid \boldsymbol{\theta}_A, \boldsymbol{\omega}_1)\right]\right\} && (25)\\
&\quad + \operatorname{var}_{\boldsymbol{\theta}_A}\left[\mathrm{E}(\boldsymbol{\omega}_2 \mid \boldsymbol{\theta}_A)\right]. && (26)
\end{aligned}
$$

All of the moments in this decomposition condition on the data $\mathbf{Y}_T^o$ and model specification A as well as the vectors explicitly indicated; we omit those terms to keep the expressions from being unduly cluttered. The term (26) is the extrinsic variance of $\boldsymbol{\omega}_2$, and therefore (24)-(25) provides a decomposition of the intrinsic variance of $\boldsymbol{\omega}_2$. The bracketed term in (24) is the variance in $\boldsymbol{\omega}_2$ that would remain even if one knew both $\boldsymbol{\theta}_A$ and $\boldsymbol{\omega}_1$, and (24) averages this with respect to the posterior distribution of $\boldsymbol{\theta}_A$ and predictive distribution of $\boldsymbol{\omega}_1$. The term in braces in (25) is the variance in $\boldsymbol{\omega}_2$ attributable to not knowing $\boldsymbol{\omega}_1$, and (25) averages this with respect to the posterior distribution of $\boldsymbol{\theta}_A$. More loosely speaking, (25) is the portion of intrinsic variance that is resolved (disappears) once $\boldsymbol{\omega}_1$ becomes known. As a specific instance, if T is the fourth quarter of 2011, $\mathbf{y}_t$ is a vector of growth rates, $\boldsymbol{\omega}_1 = \mathbf{y}_{T+1}$ and $\boldsymbol{\omega}_2 = (1/4)\sum_{s=1}^{4}\mathbf{y}_{T+s}$, then (25) is the variance in the annual growth rate for 2012 that will be resolved at the close of the first quarter of 2012.

Following the methods of Section 3, it is straightforward to compute a simulation-consistent approximation of (24)-(26). Conditional on a simulation sample $\boldsymbol{\theta}_A^{(m_2)}$ of size M_2 from the posterior distribution of $\boldsymbol{\theta}_A$, generate M_3 values $\boldsymbol{\omega}_1^{(m_2,m_3)}$ from (22) and then conditional on each of these M_2M_3 draws generate M_1 values $\boldsymbol{\omega}_2^{(m_1,m_2,m_3)}$ from (23). Then substituting $\boldsymbol{\theta}_A^{(m_2)}$ for $x_2^{(m_2)}$, $\boldsymbol{\omega}_1^{(m_2,m_3)}$ for $x_3^{(m_2,m_3)}$ and $\boldsymbol{\omega}_2^{(m_1,m_2,m_3)}$ for $x_1^{(m_1,m_2,m_3)}$, compute the moments as indicated in (16)-(18).

This process can be iterated For example, continuing with the specific case of predicting four successive quarters of growth rates, intrinsic variance can be decomposed into four rather than two components, yielding the variance in the 2012 annual growth rate that will be resolved following quarters 2 and 3 as well as quarter 1 of 2012.

4.3 Decomposition of extrinsic variance

Let the parameter vector be decomposed $\boldsymbol{\theta}_A' = (\boldsymbol{\theta}_{1A}', \boldsymbol{\theta}_{2A}')$, and suppose that it is feasible to simulate

$$
\begin{aligned}
\boldsymbol{\theta}_{A1}^{(m)} &\sim p\left(\boldsymbol{\theta}_{A1} \mid \mathbf{Y}_T^0, A\right), && (27)\\
\boldsymbol{\theta}_{A2}^{(m)} &\sim p\left(\boldsymbol{\theta}_{A2} \mid \boldsymbol{\theta}_{A1}, \mathbf{Y}_T^0, A\right). && (28)
\end{aligned}
$$

The simulation (27) is from the marginal distribution of $\boldsymbol{\theta}_{A1}$, so these simulations can be taken as the corresponding subvector of the posterior simulation sequence itself. The second simulation need not be straightforward or even feasible, though it usually is. If the posterior simulator is a pure Metropolis algorithm then the same algorithm can be used in (28), and indeed this simulation should be less challenging due to the diminished order of the parameter vector. If the posterior simulator is a Gibbs sampling algorithm and none of the blocks include components of both $\boldsymbol{\theta}_{A1}$ and $\boldsymbol{\theta}_{A2}$ then a subset of the conditional distributions required for the full posterior simulator provides (28).

Turning first to the formalities, take $X_1 = \boldsymbol{\omega}$, $X_2 = \boldsymbol{\theta}_{A1}$ and $X_3 = \boldsymbol{\theta}_{A2}$ in (8), which yields

$$\begin{aligned} \operatorname{var}(\boldsymbol{\omega}) &= \mathrm{E}_{\boldsymbol{\theta}_A}\left[\operatorname{var}(\boldsymbol{\omega} \mid \boldsymbol{\theta}_A)\right] && (29)\\ &\quad + \mathrm{E}_{\boldsymbol{\theta}_{A1}}\left\{\operatorname{var}_{\boldsymbol{\theta}_{A2}}\left[\mathrm{E}(\boldsymbol{\omega} \mid \boldsymbol{\theta}_A)\right]\right\} && (30)\\ &\quad + \operatorname{var}_{\boldsymbol{\theta}_{A1}}\left[\mathrm{E}(\boldsymbol{\omega} \mid \boldsymbol{\theta}_{A1})\right]. && (31) \end{aligned}$$

(That all moments are also conditional on $\mathbf{Y}_T^o$ and A is suppressed in this notation to avoid clutter.) The first term (29) is the intrinsic variance of $\boldsymbol{\omega}$. The terms (30)-(31) decompose the extrinsic variance of $\boldsymbol{\omega}$. The last term (31) measures systematic location movement in $\boldsymbol{\omega}$ in response to changes in $\boldsymbol{\theta}_{A1}$, where the relevant changes in $\boldsymbol{\theta}_{A1}$ are those in its posterior distribution. Since by definition all extrinsic variance is accounted for by $\boldsymbol{\theta}_A$, (30) assigns the remainder to $\boldsymbol{\theta}_{A2}$ while also accounting for the fact that the magnitude of the remainder may depend on $\boldsymbol{\theta}_{A1}$ itself. If the joint distribution of $\boldsymbol{\theta}_A$ and $\boldsymbol{\omega}$ were Gaussian and the last two terms were divided by extrinsic variance $\operatorname{var}_{\boldsymbol{\theta}_A}\left[\mathrm{E}(\boldsymbol{\omega} \mid \boldsymbol{\theta}_A)\right]$ then (31) would provide the R-squared, the fraction of extrinsic variance due to $\boldsymbol{\theta}_{A1}$.

As a specific instance, suppose a central bank using a dynamic stochastic general equilibrium (DSGE) model is interested in the extent to which extrinsic variation the one-step-ahead predictive distribution $p(\mathbf{y}_{T+1} \mid \mathbf{Y}_T^o, A)$ is driven by the structural parameters of the model (e.g., the parameters of utility, production and policy response functions) as opposed to those describing the dynamics of shocks to equations (e.g., the variances and autoregressive coefficients). The foregoing analysis provides two answers to these questions, depending on which parameters are cast as $\boldsymbol{\theta}_{A1}$ and which as $\boldsymbol{\theta}_{A2}$.

It is important to keep in mind that by definition the parameters as a whole account for all of the extrinsic variance. Taking $\boldsymbol{\theta}_{A1}$ as the vector of structural parameters provides the measure (31) of the extent to which posterior variation in structural parameters accounts for extrinsic variance, and taking $\boldsymbol{\theta}_{A1}$ as the shock dynamics parameters does the same for those parameters. If the two groups of parameters are independent in the posterior distribution, then the sum of the two measures is the extrinsic variance. But, in general, this will not be the case and it will turn out that the two measures sum to more than extrinsic variance. Nevertheless the exercise can identify contrasting contributions of groups of parameters to extrinsic variance. This ambiguity in attributing explained variance uniquely to alternative sources is precisely the same one that arises in linear regression, and is just as fundamental here as in that more elementary situation.

5 An illustration

We illustrate the methods set forth using the dynamic stochastic general equilibrium (DSGE) model with price and wage stickiness and monopolistic competition due to Smets and Wouters (2007). There are seven structural shocks in the model: innovations in total factor productivity and the risk premium, an investment specific technology shock, innovations to wage and price mark up, and policy shocks to fiscal and monetary policy. See Smets and Wouters (2007) for further detail. The model predicts seven macroeconomic time series:

1. Consumption (quarterly percentage growth rate in per capita real consumption);

2. Investment (quarterly percentage growth rate in per capita real investment);

3. Output (quarterly percentage growth rate in per capita real GDP);

4. Hours (log per capita weekly hours);

5. Inflation (quarterly percentage growth rate growth rate in GDP deflator);

6. Real wage (quarterly percentage growth rate growth rate in real wage);

7. Interest rate (Federal Funds Rate on a quarterly basis).

The data used in this illustration begin with the first quarter of 1951.

With respect to the notation introduced in Section 4 this seven-dimensional time series constitutes $\{\mathbf{y}_t\}$. DSGE modes like the one in Smets and Wouters (2007) are widely used in central banks. In the most common implementation the one-step-ahead predictive density is taken to be

$$p\left(\mathbf{y}_{T+1} \mid \mathbf{Y}_T^o, \widehat{\boldsymbol{\theta}}_A^{(T)}, A\right), \text{ where } \widehat{\boldsymbol{\theta}}_A^{(T)} = \arg\max_{\boldsymbol{\theta}_A} p\left(\boldsymbol{\theta}_A \mid \mathbf{Y}_T^o, A\right). \tag{32}$$

This does not conform with the formal Bayesian rule

$$p\left(\mathbf{y}_{T+1} \mid \mathbf{Y}_T^o, A\right) = \int_{\Theta_A} p\left(\mathbf{y}_{T+1} \mid \mathbf{Y}_T^o, \boldsymbol{\theta}_A, A\right) p\left(\boldsymbol{\theta}_A \mid \mathbf{Y}_T^o, A\right) d\boldsymbol{\theta}_A. \tag{33}$$

If there were no extrinsic variance in $\mathbf{y}_{T+1}$ in the predictive density (33) then (32) would be equivalent to (33). Of course this is not the case, but if extrinsic variance is a small enough component of predictive variance then the preference for (33) over (32) is an academic rather than a practical point. That is one of the questions investigated here.

We accessed the posterior density $p\left(\boldsymbol{\theta}_A \mid \mathbf{Y}_T^o, A\right)$ by means of a conventional Metropolis random walk posterior simulation algorithm to generate the sequence $\left\{\boldsymbol{\theta}_A^{(m_2)}\right\}$ described in Section 4 by thinning a chain of 10,000 MCMC simulations to $M_2 = 100$ equally-spaced draws. Our vector of interest is $\boldsymbol{\omega}' = (\boldsymbol{\omega}_1', \boldsymbol{\omega}_2')$, with $\boldsymbol{\omega}_1 = \mathbf{y}_{T+1}$ and

$\boldsymbol{\omega}_2 = (1/4)\sum_{s=1}^{4}\mathbf{y}_{T+s}$. Following the procedures described in the previous section we simulated $M_3 = 100$ draws $\boldsymbol{\omega}_1^{(m_2,m_3)} = \mathbf{y}_{T+1}^{(m_2,m_3)}$ from the one-quarter-ahead density $p\left(\mathbf{y}_{T+1} \mid \mathbf{Y}_T^o, \boldsymbol{\theta}_A^{(m_2)}\right)$ conditional on each of these 100 parameter drawings. Starting with each of the 10,000 pairs $\left(\boldsymbol{\theta}_A^{(m_2)}, \boldsymbol{\omega}_1^{(m_2,m_3)}\right)$ we simulated in succession $M_1 = 100$ drawings

$$\mathbf{y}_{T+s}^{(m_1,m_2,m_3)} \sim p\left(\mathbf{y}_{T+s} \mid \mathbf{Y}_T^o, \mathbf{y}_{T+1}^{(m_2,m_3)}, \ldots, \mathbf{y}_{T+s-1}^{(m_1,m_2,m_3)}, \boldsymbol{\theta}_A^{(m_2)}\right) \quad (s = 2,3,4)$$

and then formed

$$\boldsymbol{\omega}_2^{(m_1,m_2,m_3)} = (1/4)\left(\mathbf{y}_{T+1}^{(m_2,m_3)} + \mathbf{y}_{T+2}^{(m_1,m_2,m_3)} + \mathbf{y}_{T+3}^{(m_1,m_2,m_3)} + \mathbf{y}_{T+4}^{(m_1,m_2,m_3)}\right). \tag{34}$$

The simulations of $\boldsymbol{\omega}_1^{(m_2,m_3)}$ and $\boldsymbol{\omega}_2^{(m_1,m_2,m_3)}$ required just under one minute using Matlab code on a laptop computer. The analysis of variance computations (8), which produce 7×7 variance matrices, required just under one-quarter second.

		Predictive variance		Intrinsic variance	
	Total	Extrinsic	Intrinsic	Due to $\boldsymbol{\omega}_1$	Remainder
Consumption	0.2340	0.0077 (0.033)	0.2263 (0.967)	0.0706 (0.322)	0.1557 (0.688)
Investment	3.3394	0.0763 (0.023)	3.2631 (0.977)	1.2786 (0.392)	1.9845 (0.608)
Output	0.4209	0.0066 (0.016)	0.4143 (0.984)	0.1376 (0.332)	0.2767 (0.668)
Hours	1.0106	0.0152 (0.015)	0.9954 (0.985)	0.5537 (0.556)	0.4417 (0.444)
Inflation	0.0766	0.0022 (0.029)	0.0744 (0.971)	0.0696 (0.936)	0.0048 (0.064)
Wages	0.1461	0.0024 (0.016)	0.1438 (0.984)	0.0658 (0.458)	0.0779 (0.542)
Fed funds	0.0894	0.0013 (0.014)	0.0881 (0.986)	0.0489 (0.556)	0.0391 (0.444)

Table 1: Decomposition of predictive variance for one-year growth rates 2008:1- 2008:4, based on the posterior at 2007:4

The analysis of variance decomposes the predictive variance of growth rates over the next four quarters into extrinsic and intrinsic variance, and then decomposes intrinsic variance into a component that is resolved after one quarter and the remainder that is not resolved until all four quarters have been observed. We undertook this exercise at two recent points in time. The first is the end of 2007:4, using data through that quarter and analyzing the predictive distribution for growth rates over calendar year 2008. The second is the end of 2009:2, using data through that quarter and analyzing the predictive distribution for growth rates over the one-year period 2009:3 through 2010:2. The first exercise is positioned just before the onset of the global financial crisis: compared with historical values there was nothing particularly remarkable about recent quarterly growth rates. In the second exercise recent quarterly growth rates exhibit values beyond the range seen in the 60-year time series on which the posterior distribution conditions.

Table 1 shows the decomposition of variance for each of the seven series in the first exercise. The main entries are the variance terms estimated from the simulations.

The entries in parentheses in columns 3 and 4 indicate the fractional decomposition of variance between extrinsic and intrinsic variance, while the entries in parentheses in the last two columns indicate the fractional decomposition of intrinsic variance between the component resolved after one quarter and the remaining variance. Table 2 does the same thing for the second exercise.

The predictive variance (column 2 in both tables) is substantially different for the seven time series, but these differences simply reflect the units of measurement detailed at the start of this section. Except for inflation, predictive variance is modestly higher in the second exercise than in the first, and the same thing is true of intrinsic variance.

There is a marked contrast in the allocation of predictive variance to extrinsic variance (column 3 in both tables) in the two exercises. Extrinsic variance is never more than 10% of predictive variance. However, in the second exercise the fraction of predictive variance that is extrinsic is always much higher, ranging from almost three times as high to over four times as high. The explanation for this contrast lies in the values of recent time series in the two exercises. With values atypical of the sample (the second exercise) uncertainty about parameter values is magnified in the predictive distribution: the impact of unusually high or low values, which is imperfectly known, is more important. Replacing (33) with (32) does not result in serious understatement of the dispersion of the predictive distribution at either time, but the understatement is substantially greater in the second case than in the first. This finding should be generally applicable: when the data that drive the predictive distribution have been unusual, relative to the sample, the allocation of predictive variance to extrinsic variance will be greater and using point estimates in place of full predictive distributions will more seriously understate variance in near-term predictive distributions.

The allocation of intrinsic variance of the one-year forecasts to near-term and longer term reflects the volatility inherent in the values of the model's parameters that are plausible in the posterior distribution. Over future quarters, the spread in the predictive distribution will reflect actual behavior in recent quarter more strongly in the near term and the volatility implicit in the model's parameters more strongly in the far term. Thus, for example, were the model to infer unrealistically low volatility then the near term would dominate the decomposition of intrinsic variance into near and longer term (and conversely).

Intrinsic variance itself (column 4 in both tables) is virtually identical in the two exercises. For growth rates in GDP and components (the first three series) 30% to 40% of the variance is resolved in the near term. This is intuitively plausible: the near term both reveals one of the four components in (34) and in addition brings the following three quarters one quarter closer, further reducing uncertainty about these growth rates. For hours, wages and Federal funds even more of the variance, roughly half, is resolved in the near term. In the case of inflation well over 90% of the intrinsic variance is so resolved, a remarkable finding. This is consistent with the model assigning volatility to inflation that is much too low in the case of inflation, and perhaps somewhat too low in the case of hours, wages and the Federal funds rate. Of course, other explanations are possible as well, but these findings motivate a closer examination of inflation dynamics in this

		Predictive variance		Intrinsic variance	
	Total	Extrinsic	Intrinsic	Due to $\boldsymbol{\omega}_1$	Remainder
Consumption	0.2782	0.0264 (0.095)	0.2518 (0.905)	0.0817 (0.324)	0.1702 (0.676)
Investment	4.2767	0.3218 (0.075)	3.9549 (0.925)	1.6159 (0.409)	2.3390 (0.591)
Output	0.4714	0.0308 (0.065)	0.4406 (0.935)	0.1513 (0.343)	0.2893 (0.657)
Hours	1.2488	0.0771 (0.062)	1.1717 (0.938)	0.6689 (0.571)	0.5029 (0.429)
Inflation	0.0745	0.0052 (0.070)	0.0693 (0.930)	0.0652 (0.941)	0.0041 (0.059)
Wages	0.1558	0.0072 (0.046)	0.1487 (0.954)	0.0660 (0.444)	0.0826 (0.556)
Fed funds	0.0936	0.0055 (0.059)	0.0881 (0.941)	0.0489 (0.555)	0.0392 (0.445)

Table 2: Decomposition of predictive variance for one-year growth rates 2009:3 - 2010:2, based on the posterior at 2009:2

model.

6 Conclusion

From a formal but reasonable perspective the goal of Bayesian analysis can generally be cast as providing a predictive distribution relevant for a decision at hand. In doing so it integrates information from several sources, including increments to information sets as predictive distributions are updated in real time. This paper has provided a corresponding analysis of variance. Because integration of information does not typically lead to linear (Gaussian) models this analysis is necessarily more complex than the familiar treatment that has been central to statistics for a century. But the complexity poses no essential complication for simulation methods that are fast, practical, and natural in the context of modern Bayesian inference. We believe that systematic application of this analysis of variance will provide greater insight into the structure of models and information aggregation, and hope that ultimately it will be useful in improving models, predictions and decisions.

References

Box GEP (1980). Sampling and Bayes inference in scientific modeling and robustness. Journal of the Royal Statistical Society Series A 143: 383-430.

Geweke J (2005). Contemporary Bayesian Econometrics and Statistics. Englewood Cliffs NJ: Wiley.

Geweke J (2010). Complete and Incomplete Econometric Models. Princeton: Princeton University Press.

Good IJ (1956). The surprise index for the multivariate normal distribution. Annals of Mathematical Statistics 27: 1130-1135.

Lancaster T (2004). An Introduction to Modern Bayesian Econometrics. Malden MA: Blackwell Publishing.

Roberts HV (1965). Probabilistic prediction. Journal of the American Statistical Association 60: 50-62.

Smets F, Wouters R (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. American Economic Review 97: 586–606.

Weiss NA (2005). A Course in Probability. Boston: Addison Wesley.

Zellner A (1971). An Introduction to Bayesian Inference in Econometrics and Statistics. New York: Wiley.

Appendix

The following Matlab code implements the simulation-approximation of variance decomposition described in Section 3, and was used in for the illustration in Section 5.

```
function [term1 term2 term3 total]=aov(x)
% This function computes simulation-based method of moments estimates
% of the population decomposition in the paper.  The explicit
% computations are in the displays (16)-(18).
% Input:
% x This is a three-dimensional structure.  The first dimension
% corresponds to "X_2" and has the m_2 index; the second
% dimension corresponds to "X_3" and has the m_3 index; the
% third dimension corresponds to "X_1" and has the m_1 index.
% Outputs:
% term1 Estimate (16)of the first term of (8)
% term2 Estimate (17) of the second term of (8)
% term3 Estimate (18) of the third term of (8)
%
[M2 M3 M1]=size(x);
term1=var(mean(reshape(x,M2,M3*M1),2),1,1);
term2=mean(var(mean(x,3),1,2),1);
term3=mean(var(reshape(x,M3*M2,M1),1,2),1);
total=term1+term2+term3;
end
```