Subjective Prior Probability Distributions and Audit Risk

Paul J. Beck
Ira Solomon
Lawrence A. Tomassini
Subjective Prior Probability Distributions and Audit Risk

Paul J. Beck, Associate Professor
Department of Accountancy

Ira Solomon, Associate Professor
Department of Accountancy
Lawrence A. Tomassini
University of Texas at Austin

Acknowledgment: We wish to acknowledge the suggestions of the referee on earlier drafts and the comments of accounting workshop participants at the University of Illinois, Indiana University, Ohio State University, and the University of Minnesota. Other useful comments were provided by Ed Blocher, Dave Burgstahler, Dan Dhaliwal, Bill Felix, Jim Jiabalvo, D. Paul Newman, Jack Robertson, Mike Shields, and Bill Waller.
Subjective Prior Probability Distributions and Audit Risk

Abstract

This paper presents an analysis of the audit risk consequences of PPD extremeness deficiencies and miscalibration. While there is empirical evidence that auditors, like many other decision makers, assess miscalibrated PPDs, the attendant inferential risk consequences of such deficiencies have not been addressed in the extant literature. The comparative statics analysis performed in this study indicates that the risk effects of miscalibration and extremeness deficiencies on the auditor's (substantive testing) evaluation decision are complex and cannot be predicted from an examination of the planning (sampling size) decision.
1.0 Introduction

Bayesian models of the auditor's decision process (e.g., Bailey and Jensen [1977], Kinney [1975], and Scott [1973, 1975]) incorporate subjective beliefs formalized as prior probability distributions (PPDs). Recent studies by Solomon [1982] and Tomassini et al. [1982] investigated two conformance properties of auditors' account balance PPDs: extremeness and calibration. PPD extremeness is measured over a sequence of elicitation trials by computing the average subjective probability assigned to intervals containing the auditees' actual account values and, as such, can be viewed as a measure of predictive ability (see Seaver et al. [1978]). Calibration, in turn, is a measure of the ability to express an appropriate degree of confidence in such subjective estimates.\(^1\) Both the Solomon [1982] and Tomassini et al. [1982] studies reported that auditors' PPDs were miscalibrated and that there was limited evidence of underconfidence (see fn. 1). Other probability elicitation studies in psychology also have found miscalibration, but with the exception of experienced weather forecasters, these other studies have reported almost universal evidence of overconfidence (see Lichtenstein et al. [1982]).

Although several empirical studies have investigated PPD calibration and extremeness, their specific decision-making consequences have not been analyzed formally in either the psychology or accounting literatures. Libby [1981] discussed the effect of miscalibration on the audit sample size decision in compliance
testing and concluded that overconfidence impairs audit effectiveness, and underconfidence impairs audit efficiency. Tomassini et al. [1982] investigated the implications of underconfidence for the sampling decision in a substantive testing context, but did not consider the subsequent effect on the auditor's evaluation of account book values.

Simulation studies (e.g., Blocher [1981] and Cushing [1980]) have been used to investigate the sensitivity of audit planning (sample size) decisions to misspecifications of PPD parameters in the compliance and substantive testing contexts, respectively. While these studies indicated that planning decisions were sensitive to misspecifications of both PPD means and/or variances, the findings permit only very limited conclusions about the adverse consequences of extremeness deficiencies and miscalibration.

In our study, we identified specific audit risk consequences of PPD extremeness deficiencies and miscalibration for Bayesian hypothesis tests of account balances (which could be applicable to quality control and cost-variance decisions). Following Kinney [1975], we assumed that the auditor assesses account balance PPDs which are combined with mean-per-unit (MPU) sample estimates to form posterior probability distributions. The effects of extremeness deficiencies were then analyzed for a sequence of such posterior distributions by varying the PPD mean displacement relative to the corresponding outcome sequence. Similarly, the effects
of miscalibration were determined by varying the correspondence between the assessor's perceptions of PPD mean displacement and the actual displacement. We subsequently identified the attendant audit risk consequences by interpreting our comparative statics results in the context of the Kinney [1975] audit decision model.

Our study extends previous research first by formally analyzing the audit risk effects of miscalibration and extremeness deficiencies on both planning and evaluation of substantive tests. Second, by allowing subjects' PPD variances to adjust in response to the perceived PPD mean displacement, we were able to avoid experimentally confounding the effects of calibration and extremeness deficiencies.

Our analysis showed that the risk consequences of miscalibration (overconfidence and underconfidence) were complex and depended upon the actual population mean and the nature and severity of any extremeness deficiencies. Furthermore, unlike Libby [1981], we found that overconfidence actually can predispose auditors to commit either effectiveness or efficiency errors. The remaining sections of the paper are organized as follows: Section 2 presents our modeling assumptions which serve as a foundation for the comparative statics analysis in Section 3. This is followed in Section 4 by a discussion of the audit risk implications of our comparative statics findings. Concluding remarks, with implications for the design of audit training programs are provided in Section 5.
2.0 Audit Decision Context Assumptions

Audit verification of account balances is based upon both objective and subjective information. Objective information for asset accounts (such as inventory and receivables) is generally obtained by sampling from populations of subsidiary book values. Previous Bayesian studies in auditing have modelled such populations using a probability density with an uncertain mean, \( \bar{u} \). In these models, a PPD is assessed for each \( \bar{u} \) and then combined with sample information to form a posterior probability distribution. In practice, such PPDs would incorporate all of the available forms of evidence typically collected prior to substantive testing, including compliance tests.

2.1 Hypothesis Testing Assumptions

We model the audit evaluation decision as a Bayesian hypothesis test (i.e., we assume a two-point auditor loss function). This simplifying assumption implies that the losses borne by the auditor do not depend upon the magnitude of misstatement (see Moriarity [1975]) and ensures that the optimal decision rule is consistent both with the expected linear utility framework, adopted herein, and with traditional definitions of audit risk in the professional auditing literature. Specifically, for asset accounts, we assume that the null and alternative hypotheses are:
\( H_0: \bar{u} > X - M \) (i.e., account book value is not materially overstated)

\( H_a: \bar{u} \leq X - M \) (i.e., account book value is materially overstated),

where \( X \) is the auditee's reported account book value (expressed as an average), and \( M \) is the materiality limit defined with respect to the mean account book value. Note that the null hypothesis does not consider the case of material understatements, but otherwise is consistent with the hypothesis formulation initially presented in Elliott and Rodgers [1972] and subsequently adopted in Section 4 of the recently issued Audit and Accounting Guide: Audit Sampling (AICPA [1983]). A discussion of the modified risk implications for cases in which the auditor employs a two-sided hypothesis test is presented below (see fns. 8 and 10).

Figure 1 presents the four possible outcomes associated with the audit evaluation (acceptance-rejection) decision and the two hypothesized states of nature, \( H_0 \) and \( H_a \). When \( \bar{u} \leq X - M \), the null hypothesis (\( H_0 \)) should be rejected. Thus, if \( H_0 \) were accepted, audit effectiveness would be compromised and the loss is represented by \( L_{II} \). Alternatively, when \( \bar{u} > X - M \), audit efficiency would be compromised if \( H_0 \) were to be rejected, and \( L_I \) represents the associated loss.

Assuming minimization of expected losses, the optimal audit decision rule (see Berger [1980]) is to reject \( H_0 \) if and only if:
\[
\frac{\Pr(H_0)}{1 - \Pr(H_0)} < \frac{L_{II}}{L_I},
\]

(1)

where \(\Pr(H_0)\) represents the posterior probability that \(\bar{\mu} > X - M\).

The above formulation simplifies our analysis because, for a given loss ratio, \(L_{II}/L_I\), the optimal decision depends solely upon the posterior likelihood ratio, \(\Pr(H_0)/(1 - \Pr(H_0))\). The relationship between \(L_I\) and \(L_{II}\) affects the audit risk implications of miscalibration and extremeness deficiencies and will be discussed further in Section 4.

2.2 Distributional Assumptions

Consistent with previous research (e.g., Kinney [1975] and Scott [1973, 1975]), we assume that the distribution of subsidiary book values and auditor's PPD are normally distributed and that the ordinary MPU estimator is employed. We also assume that the variance of each distribution is known. Given these assumptions, the posterior distribution for each audited account value is normally distributed and parameterized (see Winkler [1972, p. 169]) by \((m'', v'')\), where:

\[
m'' = \frac{(\frac{1}{v'})m' + (\frac{n}{v})m}{(\frac{1}{v'}) + (\frac{n}{v})}
\]

(2)

\[
\frac{1}{v''} = \frac{1}{v'} + \frac{n}{v}
\]

(3)

and:
\[ m = \text{sample estimate of the population mean, } \bar{m}; \]
\[ m' = \text{PPD mean}; \]
\[ m'' = \text{posterior estimate of the mean}; \]
\[ n = \text{quantity of sample evidence}; \]
\[ v/n = \text{variance of } m; \]
\[ v' = \text{PPD variance for the assessed } m'; \]
\[ v'' = \text{posterior variance}; \]

The above assumptions are not as restrictive as they might appear since, as discussed below, our analysis of the simple MPU estimator can be generalized directly to the stratified MPU estimator. Also, the assumption of population normality is defensible when stratified sampling is employed since the strata, which are more internally homogeneous, are less skewed than the population taken as a whole. More importantly, sample MPU estimates of the mean should be normally distributed, given the central limit theorem. Previous empirical results (e.g., Cushing and Romney [1981], and Solomon, et al. [1982]) suggest also that auditors' PPDs assessed for highly skewed populations, such as those described in Neter and Loebbecke [1975], are typically symmetric and can be approximated by normal PPDs. Finally, relaxing the assumption of known sample variance by allowing an assessment of a second PPD for the variance parameter (actually a joint PPD for the mean and variance parameters) as in Felix and Grimlund [1977] would not change our analysis in any significant manner. Calibration and extremeness,
however, would have to be redefined with respect to each of the unknown parameters.

Our assumptions establish a simplified environment in which there is only one uncertain parameter and there are no sample estimation problems due to skewness and/or an unknown sample variance. If miscalibration can be shown to be a significant problem in the simplified environment modeled herein, we can reasonably expect it to be as important or more so when the probability elicitation task is more complex and there are statistical estimation problems.

2.3 Extremeness and Calibration Assumptions

Since calibration and extremeness are defined with respect to a sequence of PPDs and corresponding outcomes (see Lichtenstein et al. [1982]), we assumed that the probability assessment-hypothesis testing procedures were to be repeated for a sequence of trials with outcomes, \( \{\mu_t; t=1, T\} \). This assumption is consistent with traditional definitions of inferential risks (e.g., alpha and beta risks) which are defined with respect to a sequence of estimates.

PPD extremeness, as noted above, is measured \textit{ex post} by imputing the probability assigned to a suitably chosen interval containing each \( \mu_t \) and then averaging these probabilities over the sequence (see Solomon [1982]). An achievement of high levels of extremeness would then require that the auditor-assessor specify distributions which are centered appropriately, relative to the \( \mu_t \) sequence, and are tight (see Peters [1978]). Otherwise, the
highest probabilities would be assigned to population values which do not occur most frequently. We modeled extremeness deficiencies in our study by PPD mean displacement (i.e., \( m' = \mu + \Delta \), where \( \Delta \) denotes the PPD mean displacement). By varying \( \Delta \) in our comparative statics analysis, we manipulated the severity of the assessor-auditor's extremeness deficiency.

The auditor's calibration, (see above and also fn. 1), reflects the ability to express an appropriate degree of confidence in subjective estimates of population means so it is defined with respect to a given level of extremeness (i.e., magnitude of \( \Delta \)). Both overconfidence and underconfidence calibration deficiencies were modeled. An overconfident assessor's PPDs would be too tight and, therefore, over a sequence of elicitation trials, too many realizations would be captured in the outer fractile ranges, and too few captured in the inner fractile ranges (see Lichtenstein et al. [1982]). In contrast, underconfident assessors would specify PPDs which are too diffuse, in that the inner fractile ranges capture too many outcomes, while the outer fractile ranges capture too few outcomes. The variance of a normal PPD represents the assessor's uncertainty about the population mean and, thus, is related directly to the auditor's calibration (see Moskowitz and Bullers [1979]).

Calibration was modeled in our comparative statics analysis by assuming that the auditor would implicitly estimate the average
mean displacement that would exist if the elicitation process were repeated for the entire \( \mu^e \) sequence, and then would adjust the assessed PPD variance on each trial to reflect the estimated (average) displacement. Since each trial was assumed to be exchangeable, the assessor's perceptions about displacement would be the same for each trial in the sequence (see fn. 3). Furthermore, provided that an assessor were not completely miscalibrated, there should be some correspondence between the assessor's perceptions about the PPD mean displacement (\( \hat{\Delta} \)) and the actual \( \Delta \). Accordingly, we formalized this correspondence between \( \hat{\Delta} \) and \( \Delta \) by expressing the PPD variance on each trial as a function of \( \hat{\Delta} \) (i.e., \( v' = v'(\hat{\Delta}) \), where \( \Delta = \hat{\Delta}(\Delta) \)). Note that, within the context of our model, perfect calibration would exist if \( \hat{\Delta}(\Delta) = \Delta \) for each possible level of extremeness since the auditor would be able to spread the PPD variance appropriately (see fn. 2). Alternatively, \( \hat{\Delta}(\Delta) > \Delta \) would imply underconfidence, while \( \hat{\Delta}(\Delta) < \Delta \) would indicate overconfidence. The correspondence between \( \hat{\Delta} \) and \( \Delta \) will be varied in our comparative statics analysis to identify the consequences of the two types of miscalibration.

3.0 Comparative Statics Analysis

In this section we first investigate the effects of extremeness deficiencies and miscalibration on a sequence of posterior
distributions, after which (in Section 4) the audit risk implications of these effects will be discussed.

3.1 Extremeness

The effect of PPD mean displacement on the posterior mean for trial \( t \) can be determined by substituting \( \mu_t + \Delta \) for \( m' \) into (2). Assuming that \( m_t = \mu_t \), the following simplified expression for the posterior mean is obtained:

\[
m''_t = \mu_t + \frac{\Delta}{(1 + n_t v'/v)},
\]

(4)

where \( n_t = n(m'_t, v') \), \( m'_t = \mu_t + \Delta \), and \( v' = v'(\Delta(\Delta)) \).

The difference between \( m''_t \) and \( \mu_t \) represents the magnitude of posterior mean displacement on trial \( t \). Since \( (1 + n_t v'/v) > 0 \), equation (4) predictably indicates that the posterior mean on each trial will be displaced in the same direction as the PPD mean due to the presence of \( \Delta \). However, the magnitude of the posterior mean displacement can vary from trial to trial. Accordingly, we take an expectation (average) with respect to \( \{\mu_t\} \).

\[
E[m''_t] = E[\mu_t] + E[\Delta/(1 + n_t v'/v)].
\]

(5)

The consequences of altering PPD extremeness deficiencies for the sequence are determined by differentiating (5) totally with respect to \( \Delta \) using the quotient and chain rules,
\[
\frac{dE[m^*]}{d\Delta} = E\left\{ \frac{(1 + n'v'/v)}{(1 + n'v'/v)^2} \left[ \Delta \left( \frac{d\Delta}{d\Delta} + \left( \frac{v'v}{v} \right) \frac{dn}{d\Delta} \right) \right] \right\} \quad (6)
\]

Simplifying (6) algebraically,

\[
\frac{dE[m^*]}{d\Delta} = E\left\{ \frac{1}{(1 + n'v'/v)} \left[ \Delta \frac{d\Delta}{d\Delta} - \frac{n't}{v(1 + n'v'/v)^2} \frac{v'v}{d\Delta} - \frac{v'on}{v(1 + n'v'/v)^2} \frac{dn}{d\Delta} \right] \right\}. \quad (7)
\]

Equation (7) shows that the total effect of PPD mean displacement on the expected value of the posterior mean sequence can be decomposed into three terms. The first term on the right side of (7) represents the expected (upward) posterior mean displacement which would occur if \( \frac{dv'}{d\Delta} = \frac{dn}{d\Delta} = 0 \), whereas the second and third terms reflect the sensitivity of the PPD variance and sample size, respectively, to changes in the magnitude of PPD mean displacement. The second term can be further decomposed as follows:

\[
\frac{dv'}{d\Delta} = \frac{dv'}{d\Delta} \frac{d\hat{\Delta}}{d\Delta}. \quad \text{This decomposition indicates that the expected magnitude of posterior mean displacement in (7) is influenced by the auditor's calibration (sensitivity) to PPD mean displacement, as reflected by the} \ \frac{d\hat{\Delta}}{d\Delta} \ \text{component of the} \ \frac{dv'}{d\Delta} \ \text{term. In particular, if} \ \frac{d\hat{\Delta}}{d\Delta} < 0 \ (\text{which implies very severe overconfidence caused by the assessor's believing that the PPD mean displacement has been reduced when it actually has increased}), \ \text{the expected posterior mean displacement in (7) would be increased due to the minus sign preceding the second term.}
\]
Another possibility is that \( 0 < \frac{\hat{d}A}{dA} < 1 \). In this situation, the auditor would still be overconfident, because the PPD variance would not change commensurately with the magnitude of the PPD mean displacement. Hence, the auditor would fail to recognize the actual level of PPD mean displacement. Note that such overconfidence was effectively incorporated in Cushing's [1980] study in which the PPD mean and variance were perturbed independently (i.e., \( \frac{dv'}{d\hat{A}} \frac{d\hat{A}}{dA} = 0 \)).

Alternatively, if \( \frac{d\hat{A}}{dA} = 1.0 \), the auditor would exhibit an appropriate calibration sensitivity to the PPD mean displacement by increasing the PPD dispersion, thereby, effectively eliminating posterior mean displacement. Finally, if \( \frac{d\hat{A}}{dA} > 1.0 \) (i.e., auditor is underconfident), the assessed PPD variances would be too large given the actual level of PPD mean displacement. While (8) indicates that severe underconfidence can effectively drive the expected posterior mean displacement to zero, unnecessary sampling would still be performed.

Further insight regarding the sample size effect is obtained by decomposing the \( \frac{dn}{d\hat{A}} \) term in (7) which will show the dependence of \( n \) on \( \Delta \) through the \( \hat{A} \) and \( v' \) terms:

\[
\frac{dn_t}{d\hat{A}} = \frac{\partial n_t}{\partial m'} \frac{dm'}{d\hat{A}} + \frac{\partial n_t}{\partial v'} \frac{dv'}{d\hat{A}} \frac{d\hat{A}}{dA}.
\]

The \( \frac{\partial n_t}{\partial m'} \frac{dm'}{d\hat{A}} \) term in (8) can be either positive or negative, depending on the auditor's loss function.\(^5\) Hence, general conclu-
sions about the sample size effect and the magnitude of the posterior mean displacement are not possible. Nevertheless, the decomposition in (8) indicates that there is a "calibration component" embedded in the $\frac{\partial n_t}{\partial \Delta}$ term in (7). Since the sign of the terms multiplying $\frac{\partial n_t}{\partial \Delta}$ in (7) is negative, the expected magnitude of the posterior mean displacement will be reduced by the calibration component when $\frac{\partial v'}{\partial \Delta} \frac{\partial \hat{\Delta}}{\partial \Delta} > 0$, but increased when $\frac{\partial v'}{\partial \Delta} \frac{\partial \hat{\Delta}}{\partial \Delta} < 0$.

3.2 Calibration

This subsection extends the previous analysis by investigating the effects of PPD miscalibration on the posterior variance sequence. This is done by holding $\Delta$ constant, and varying the inaccuracy of the assessor's perception of the actual displacement. More specifically, we model the effects of miscalibration by assuming that $\hat{\Delta}(\Delta) = (1+\phi)\Delta$, where $\phi$ represents the inaccuracy of the assessor's perception of $\Delta$. In this framework, $\phi = 0$ would imply that the assessor's perception of the actual level of PPD mean displacement across the sequence of trials is accurate, so that the PPD variance can be adjusted appropriately. Alternatively, $\phi < 0$ and $\phi > 0$ would indicate overconfidence and underconfidence, respectively.

As a first step here, we solve (3) for $v''$. Upon simplification, the posterior variance for the $t^{th}$ trial can be expressed as:

$$v''_t = \frac{v}{n_t + v/v'},$$

(9)
where: \( n_t = n(m'_t, v'), \) \( m'_t = \mu_t + \Delta, \) \( v' = v'(\hat\Delta), \) and \( \hat\Delta = (1+\phi)\Delta. \)

Taking an expectation with respect to the \( \mu_t \) sequence, we obtain the average variance of the posterior sequence:

\[
E[v''_t] = vE[1/(n_t + v/v')]. \tag{10}
\]

Differentiating (10) with respect to \( \phi, \)

\[
\frac{dE[v'']}{d\phi} = -vE\left[\frac{1}{n_t + v/v'}\right] - 2\left[\frac{dn_t}{dv'} \frac{dv'}{d\hat\Delta} \frac{d\hat\Delta}{d\phi} - v\frac{dv'}{d\hat\Delta} \frac{d\hat\Delta}{d\phi}/v''\right]. \tag{11}
\]

Given that \( \frac{d\hat\Delta}{d\phi} = \Delta \) and \( \frac{dv'}{d\hat\Delta} > 0, \) (11) indicates that

\[
\frac{dE[v'']}{d\phi} > (\langle \rangle) 0 \quad \text{when} \quad \frac{dn_t}{dv'} < (\langle \rangle) v/v''^2. \tag{12}
\]

Since the direction of the inequality in (12) depends upon the specific parameter values, general conclusions about the effects either of underconfidence or overconfidence are not possible. An inspection of (12) indicates, however, that misspecification of the PPD variance will have the greatest impact on the posterior variance when the PPD variance is small in relation to the sample variance—i.e., the auditor has substantial subjective knowledge—or sampling costs are high so that \( \frac{dn_t}{dv'} \) is small. Hence, in situations in which the formal incorporation of subjective audit knowledge potentially would be most beneficial, underconfidence will cause the posterior variance to be biased upward, while overconfidence will cause the posterior variance to be biased downward.
3.3 **Stratification**

So far our analysis of extremeness deficiencies and miscalibration has assumed unstratified MPU sampling. In practice, however, population skewness often necessitates stratification. Application of Bayesian stratified methods requires the assessment of a vector of means and k(k+1)/2 elements of the variance-covariance matrix (see Ericson [1965]). But, in the special case in which the stratum means are independent (i.e., the variance-covariance matrix is diagonal), extremeness and calibration can be computed with respect to a PPD sequence for each stratum. This case represents the direct Bayesian analogue of the classical stratified sampling model in which the sample mean estimates are assumed to be distributed independently.

In such Bayesian stratified sampling models, posterior estimates of both the mean and variance are computed for each stratum and then aggregated linearly to obtain a stratified posterior estimate for the population as a whole. As a result, if our assumed hypothesis test (decision rule) were employed to evaluate the stratified posterior estimates, we would obtain comparative statics results similar to those reported above. For example, if the stratum PPD mean sequences are displaced in the same direction, the posterior means and stratified mean estimate for the population also will be displaced and in the same direction. If, in addition, the auditor is overconfident (underconfident) in
assessing the PPDs for one or more strata, the expected magnitude of posterior mean displacement will be increased (decreased) and the expected posterior variance also will be increased (decreased). Unfortunately we are unable to provide general results when the strata means are correlated or if calibration deficiencies differ across strata.

4.0 Audit Risk Implications

We now identify the risk implications of extremeness deficiencies and miscalibration for the audit evaluation decision. An ex post perspective is implicitly adopted in which we focus on a representative trial (t) and separately analyze Case I in which \( \mu_t \leq X - M \) and Case II in which \( \mu_t > X - M \). Separate risk analyses are required, because audit efficiency and effectiveness errors depend jointly upon the audit decision and whether the account balance actually contains a material (overstatement) error.

4.1 Case I: Book Value Materially Overstated

Since the account book value is assumed to be materially overstated (i.e., \( \mu_t \leq X - M \)), we focus on the auditor's risk of committing an effectiveness error in the four subcases below.

4.1.1 No Mean Displacement

We initially simplify by considering the risk effects of miscalibration in the absence of PPD mean displacement. Figure 2
presents two normal posterior probability distributions which, for expositional purposes, are assumed to be based upon PPDs assessed by auditors A and B, respectively. Since the PPDs and posterior distributions correspond to the same \( \mu_t \) outcome, we henceforth replace the \( t \) subscript on all PPD parameters with the letters A and B, so that each distribution can be associated with a particular auditor-assessor. While both auditors are assumed to assess PPDs which are properly centered vis-a-vis the actual account balance (i.e., \( m'_A = m'_B = \mu_t \)), Auditor A is more underconfident than Auditor B (i.e., \( \phi_A > \phi_B > 0 \)). Therefore, consistent with our assumption that \( \frac{dE[v^*]}{d\phi} > 0 \), distribution A is more diffuse than distribution B and, thus, can be interpreted as reflecting greater underconfidence.

---

Insert Figure 2 Here

---

The area under each distribution to the right of \( X-M \) represents the probability that the actual account balance mean (\( \mu_t \)) exceeds \( X-M \) and is denoted by \( Pr(H_0) \), while that to the left of \( X-M \) represent \( Pr(H_a) = 1 - Pr(H_0) \). Given the symmetry property of normal distributions and the further assumption that \( m_t = \mu_t \), the posterior means, \( m''_A, m''_B \preceq X - M \), so the computed \( Pr(H_0) \leq .5 \) for both distributions A and B.

The optimal decision, for either auditor, is to reject \( H_0 \) when the computed likelihood ratio, \( Pr(H_0)/(1-Pr(H_0)) < L_{II}/L_I \). Since
the costs of effectiveness and efficiency errors depend upon factors specific to the decision-making context, general conclusions cannot be drawn about the loss ratio \( \frac{L_{II}}{L_I} \). Typically, however, the assumption is that losses from audit effectiveness errors are larger than those from audit efficiency errors, implying that \( \frac{L_{II}}{L_I} > 1.0 \) (see Elliott and Rodgers [1972] and AICPA [1981, 1983]). This means that, in the present subcase in which \( \Pr(H_0)/(1-\Pr(H_0)) \leq 1.0 \), \( H_0 \) would be (correctly) rejected by either Auditor A or B. In fact, varying the magnitude of the posterior variance would not alter the rejection decision, because the computed \( \Pr(H_0) < .5 \) when \( m''_A = m''_B = \mu_t \leq X - M \). Therefore, in the absence of mean displacement, underconfidence does not affect audit risk exposure (see Table 1).

But the audit planning decision would be adversely affected by underconfidence. That is, ceteris paribus, an underconfident auditor would choose a larger sample size than a properly calibrated auditor when \( \frac{d_n}{d_{\nu'}} > 0 \). This would result in an opportunity cost due to insufficient reliance upon subjective knowledge. The next two subcases extend the preceding analysis by incorporating PPD mean displacement.

4.1.2 Moderate Upward Mean Displacement

We now assume that both Auditors A and B assess PPDs whose means are equally displaced upward (i.e., \( m'_A = m'_B > \mu_t \)), but
Auditor A is more underconfident than B. From our previous analysis, it is apparent that both posterior means will be displaced upward (i.e., $m''_A, m''_B > \mu_t$), so that the computed $Pr(H_0)$ and likelihood ratio will be increased. Here, we consider the risk effects when $\mu_t < m'_A = m'_B < X - M$, but in the next subsection we consider more serious upward mean displacements wherein

$$\mu_t < X - M < m'_A = m'_B.$$ 

When $\mu_t < m'_A = m'_B < X - M$, the posterior means, $m''_A, m''_B < X - M$, so the computed $Pr(H_0) < .5$ for both distributions. Therefore, given that $L_{II} > L_I$, both auditors will correctly reject $H_0$, so changes in the PPD variance due to miscalibration would not alter the rejection decision. Therefore, upward mean displacement and/or miscalibration do not affect risks provided that

$$m'_A = m'_B < X - M$$
(see Table 1).

### 4.1.3 Serious Upward Mean Displacement

In practice, the magnitude of the upward PPD mean displacement may be so great that $m'_A = m'_B > X - M$. If so, both likelihood ratios could exceed the loss ratio, which would result in the commission of effectiveness errors (see Table 1). The predisposition to commit an effectiveness error, however, is influenced by the severity and nature of the auditor's miscalibration. Figure 3 presents two posterior probability distributions for Auditors A and B. As in 4.1.1, $m'_A = m'_B$, but the posterior mean displacement is less for distribution A, because the (displaced) PPD mean
is weighted less heavily vis-à-vis the sample mean due to more serious underconfidence.

Reducing the upward posterior mean displacement, ceteris paribus, increases the computed $Pr(H_a)$ and reduces $Pr(H_0)$. Since the area under the lower tail of the distribution also is increased by underconfidence, the computed $Pr(H_a)$ is further increased, thereby reducing the computed likelihood ratio. Thus, underconfidence reduces the auditor's predisposition to commit an effectiveness error due to upward mean displacement (see Table 1), but at the cost of suboptimal (larger) sample sizes.

The risk effects of overconfidence can also be determined from Figure 3 by reinterpreting distribution B as exhibiting greater overconfidence than A. Predictably, the effects of overconfidence on the posterior distribution are contrary to those described above for underconfidence, because the weight accorded the PPD mean is increased and the audit sample size is simultaneously reduced. As a result, the posterior mean displacement is increased and the posterior variance is decreased. Both effects reduce the area to the left of the point $X-M$ thereby increasing the computed $Pr(H_0)$. Hence, when $m'_A = m'_B > X - M$, overconfidence increases the auditor's predisposition to commit an effectiveness error resulting from upward PPD mean displacement (see Table 1).
4.1.4 Downward Mean Displacement

We now assume that the PPD means are displaced downward (i.e., \( m_A' = m_B' < \mu_c \)). Since by assumption, \( m = \mu_c > X - M \), our previous analysis indicates that the posterior means, \( m_A'' \), \( m_B'' < X - M \). Therefore, the computed \( \Pr(H_0) < .5 \) for both distributions which ensures that both auditors will correctly reject \( H_0 \). Furthermore, as discussed above, changes in the PPD variance due to miscalibration would not increase the computed \( \Pr(H_0) \) above .5. Hence, no risk consequences are associated with downward PPD mean displacement and/or miscalibration (see Table 1).

4.2 Case II: Account Book Value NOT Materially Overstated

We now assume that the account book value is not materially overstated (i.e., \( \mu_c > X - M \)). This allows us to focus on the risk of committing an efficiency error due to PPD extremeness deficiencies and miscalibration.

4.2.1 No Mean Displacement

We first consider a subcase in which there is no PPD or sample mean displacement (i.e., \( m_A' = m_B' = m_c' = \mu_c \)), but the auditor is underconfident. Figure 4 presents two posterior probability distributions which are based upon the PPDs assessed by Auditors A and B and are analogous to those of Figure 2, except we now assume that \( \mu_c > X - M \). As in 4.1.1, the computed \( \Pr(H_a) \) is higher for the more diffuse distribution (A) which reduces the likelihood
ratio. However, in contrast with the earlier effects, the more underconfident auditor has a greater predisposition to commit an efficiency error by incorrectly rejecting $H_0$.

4.2.2 Upward Mean Displacement

Upward PPD mean displacement (i.e., $m'_A = m'_B > \mu_t$), *ceteris paribus* results in upward posterior mean displacement which, as in cases 4.1.2 and 4.1.3, increases the computed $\Pr(H_0)$. Here, however, increasing the likelihood ratio reduces the risk of efficiency errors (see Table 1).10 Figure 5 presents two posterior distributions depicting the joint effects of upward PPD mean displacement and miscalibration. Posterior distribution A is centered to the right of distribution B and is less diffuse to reflect greater overconfidence. As in subsection 4.1.3, both the displacement and dispersion effects upon the posterior distribution increase the computed $\Pr(H_0)$. However, the effect of increasing the computed $\Pr(H_0)$ now is to reduce the probability of efficiency errors.

Unfortunately, when there is underconfidence, the risk effects are more difficult to discern. Figure 5 can be reinterpreted to facilitate an analysis of underconfidence in conjunction with
upward PPD mean displacement. The more diffuse distribution B now represents greater underconfidence than distribution A. Note that greater PPD underconfidence reduces the upward posterior mean displacement while simultaneously increasing the posterior variance. This effect tends to mitigate the effect of upward PPD mean displacement. As a result, the ultimate risk consequences are indeterminate unless specific assumptions are made both about the magnitude of the PPD mean displacement and the severity of the auditor's underconfidence (see Table 1).

4.2.3 Downward Mean Displacement

We now analyze the case of downward PPD mean displacement in which \( m'_A = m'_B < \mu_t \). Since the posterior mean is displaced in the same direction as the PPD mean, the computed \( \Pr(H_0) \) is decreased which, ceteris paribus, predisposes the auditor to commit efficiency errors (see Table 1). Such predisposition, however, is also influenced by the auditor's miscalibration. Figure 6 presents two posterior distributions which depict the joint effects of downward PPD mean displacement and miscalibration. Auditor A's posterior distribution is displaced further than B's and also is tighter than distribution B due to greater overconfidence. Note that the areas to the left of \( X - M \) are approximately the same for both distributions, given the opposing mean and variance effects on the computed \( \Pr(H_a) \). Hence, the efficiency risk consequences cannot be determined. Finally, note that, while we have assumed
implicitly that \( X - M < m_A^* = m_B^* < \mu_t \), the same indeterminate risk effects also would arise when \( m_A^* = m_B^* < X - M < \mu_t \).

Insert Figure 6 Here

The risk consequences of underconfidence also can be determined from Figure 6 by reinterpreting the relationship between the two posterior distributions. Auditor B is now assumed to be more underconfident than A, so distribution B exhibits less downward mean displacement and also is more diffuse than distribution A. However, since the joint effects of underconfidence and downward mean displacement are in opposition to one another, the audit efficiency risk exposure cannot be determined again without specific assumptions both about the magnitude of PPD mean displacement and severity of the underconfidence (see Table 1).

5.0 Concluding Remarks

In this paper, we have investigated the audit risk consequences of PPD extremeness deficiencies and miscalibration in a Bayesian hypothesis testing context. These risk consequences, as summarized in Table 1, are complex and their identification requires an analysis of both the auditor's planning and evaluation decisions. In contrast with the existing literature, our analysis showed that overconfidence could predispose the auditor toward either efficiency or effectiveness errors. Furthermore, the risk consequences of each type of miscalibration were found to depend
jointly upon the actual population mean and the nature and severity of the extremeness deficiencies.

Public accounting firms devote significant resources to training programs designed to impart audit knowledge which can improve the subjective judgment accuracy component portrayed here by PPD extremeness. Our analysis suggests the need to give additional consideration to training directed specifically at improving auditor calibration.
FOOTNOTES

1. A sequence of continuous PPDs would exhibit proper calibration (and the assessor's confidence would be appropriate) if the fractiles of the PPDs correspond with the relative outcome frequencies. For example, only 10% of the outcomes should fall to the right of the .90 fractiles of the PPD sequence. If more (less) than 10% of the outcomes should occur in the right tails of the PPDs, overconfidence (underconfidence) is indicated.

2. The auditor has two choices: 1) conclude that the account book value is materially misstated and possibly extend substantive tests or 2) conclude that the account book value is fairly stated and discontinue testing. If the former choice is made, the expected loss is:

$$p_r(H_0) \cdot 0 + (1 - p_r(H_0))L_{II}.$$ 
Alternatively, if $H_0$ is rejected, the expected loss is:

$$p_r(H_0)L_{I} + (1 - p_r(H_0))0.$$ 
The rejection decision is optimal when:

$$p_r(H_0)L_{I} < (1 - p_r(H_0))L_{II}$$
which is equivalent to the condition that:

$$p_r(H_0)/(1 - p_r(H_0)) \leq L_{II}/L_{I}.$$ 

3. Calibration must be measured with respect to a sequence of PPDs assessed for exchangeable events (i.e., the assessor perceives the events as equally uncertain). This implies that, for normally distributed PPDs, the same $v'$ would be specified for each trial in the sequence. For purposes of
identifying audit risk consequences, however, we must further assume that, for all uncertain account balances in the sequence, the auditor's loss ratio is the same. Such an assumption might be satisfied if the PPDs are assessed for the same account and the auditees are in similar circumstances. We also assume that the PPD sequence is assessed by a single auditor. Multiple auditor-assessors could be introduced (as in most calibration studies, see Lichtenstein et al. [1982]), provided that they are homogeneous with respect to their substantive knowledge, calibration deficiencies, and perceived uncertainty.

4. One could assume that the \( \{\mu_t\} \) are drawn from a superpopulation (see Godfrey and Andrews [1982]) modelled by a probability density function, \( g(\mu_t) \). Given such an assumption, the expectation in (5) would be taken with respect to \( g(\cdot) \). Alternatively, one could view the expectation as a simple average over \( \{\mu_t\} \).

5. Raiffa and Schlaiffer [1961] have shown that \( \frac{3n}{\partial m'} \frac{d\nu'}{d\Delta} \frac{d\hat{\Delta}}{d\Delta} \) is positive for most commonly employed loss functions. The \( \frac{3n}{\partial m'} \frac{d\hat{m}}{d\Delta} \) term is related to the expected value of sample information and depends, therefore, on the location and dispersion of the auditor's PPD. In our model, the term would be positive if the likelihood ratio, \( P_r(H_0)/(1 - P_r(H_0)) \), initially were close to the critical ratio \( (L_{II}/L_I) \) and the PPD mean
displacement changed the likelihood ratio, so that the auditor's optimal evaluation decision changed.

6. Our benchmark case deals with underconfidence which would occur if the auditor failed to recognize that $\Delta = 0$. When $\Delta = 0$, we would not expect to observe overconfidence on a representative trial from the sequence.

7. The audit risk consequences of PPD mean displacement and miscalibration (discussion below) are the same when the auditor adopts the Statement on Auditing Procedure (SAP) No. 54 [AICPA, 1972] decision rule (i.e., interchanges the null and alternate hypotheses), provided that the PPD sequence is assessed for account balances. However, if the PPD sequence were assessed for the monetary error in the account rather than the account value itself, the risk consequences associated with upward PPD mean displacement would be the same as downward displacement in our model and vice versa.

8. Given a one-sided hypothesis test, upward displacement of the posterior mean increases $P_r(H_0)$. If a two-sided hypothesis test were employed, $P_r(H_0)$ would not increase monotonically with the posterior mean. Hence, sufficiently large posterior mean displacement could result in rejection of $H_0$. Note that, while rejection of $H_0$ would be correct, the auditor's conclusion about the direction of the material error would be incorrect. This is considered in more detail below.
9. In the unlikely event that \( dE[v^+] / d\phi < 0 \), distribution A would be more diffuse than distribution B, so that the effects of overconfidence would be analogous to those described above for underconfidence. Similarly, the effects of underconfidence when \( dE[v^-] / d\phi < 0 \), would be similar to those for overconfidence when \( dE[v^-] / d\phi > 0 \).

10. As discussed above (see fn. 8), if a two sided hypothesis test were employed, upward posterior mean displacement could decrease \( P_r(H_0) \) and create a predisposition toward efficiency errors.
### States:

<table>
<thead>
<tr>
<th>Decision</th>
<th>$\tilde{\mu} &gt; X - M$</th>
<th>$\tilde{\mu} \leq X - M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accept $H_0$</td>
<td>0</td>
<td>$L_{II}$</td>
</tr>
<tr>
<td>Reject $H_0$</td>
<td>$L_I$</td>
<td>0</td>
</tr>
</tbody>
</table>

- $\tilde{\mu}$ ≡ actual population mean
- $X$ ≡ average value of subsidiary account reported by the client
- $L_I$ ≡ loss from audit inefficiency
- $L_{II}$ ≡ loss from audit ineffectiveness

**Auditor's Loss Matrix**

**Figure 1**
Fig. 2 - Underconfidence - No Mean Displacement

\[ m_A'' , m_B'' = \mu_e = x - M \]

Monetary Value
Fig. 3 - Miscalibration, Upward Mean Displacement
Fig. 4 – Underconfidence – No Mean Displacement
Monetary Value

Fig. 5 - Miscalibration, Upward Mean Displacement
Fig. 6 - Miscalibration - Downward Mean Displacement
### Ex Post Audit Risk Implications

#### Case I: Account Book Value Materially Overstated

<table>
<thead>
<tr>
<th>Basic Risk Effect of PPD Mean Displacement</th>
<th>Joint Risk Effect</th>
<th>Underconfidence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No PPD Mean Displacement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( m_A' = m_B' = \mu_c \leq X - H )</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td><strong>Upward PPD Mean Displacement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \mu_c &lt; m_A' = m_B' &lt; X - H )</td>
<td>No Risk Effect (Always Correctly Reject ( H_0 ))</td>
<td></td>
</tr>
<tr>
<td>( \mu_c \leq X - H &lt; m_A' = m_B' )</td>
<td>Predisposition Toward Effectiveness Errors</td>
<td></td>
</tr>
<tr>
<td><strong>Downward PPD Mean Displacement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( m_A' = m_B' &lt; \mu_c \leq X - H )</td>
<td>No Risk Effect (Always Correctly Reject ( H_0 ))</td>
<td></td>
</tr>
</tbody>
</table>

#### Case II: Account Book Value NOT Materially Overstated

<table>
<thead>
<tr>
<th>Basic Risk Effect of PPD Mean Displacement</th>
<th>Joint Risk Effect</th>
<th>Underconfidence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No PPD Mean Displacement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( m_A' = m_B' = \mu_c &gt; X - H )</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td><strong>Upward Mean Displacement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( m_A' = m_B' &gt; \mu_c &gt; X - H )</td>
<td>Reduction in Risk of Efficiency Error</td>
<td></td>
</tr>
<tr>
<td><strong>Downward PPD Mean Displacement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( X - H &lt; m_A' = m_B' &lt; \mu_c )</td>
<td>Predisposition Toward Efficiency Error</td>
<td></td>
</tr>
</tbody>
</table>

---

Note: The formulas and risk effects are interpreted in the context of statistical hypothesis testing, where \( m_A' \) and \( m_B' \) represent the observed means, \( \mu_c \) is the claimed mean, and \( X - H \) is a threshold or target value.
REFERENCES


