More information, less precision: meta-analysis through random effects

Jan R. Magnus

Department of Econometrics and Data Science,

Vrije Universiteit Amsterdam and Tinbergen Institute, The Netherlands

Andrey L. Vasnev

University of Sydney, New South Wales, Australia

January 14, 2025

Abstract: Given several studies (inputs) of some phenomenon of interest, each input presents an estimate of a key parameter with an associated estimated precision. The random-effects model used in meta-analysis estimates this parameter based on a decomposition of the error term into within-input noise and across-input noise. Our interest is in the precision of this estimator, which leads to a confidence interval of the parameter. But we shall also be interested in the precision when we transform the inputs into one input, which leads to a (much wider) prediction interval. We review and extend the meta-analysis framework in a maximum likelihood context, paying special attention to conflict between the inputs, correlation between the inputs, and the difference between confidence and prediction intervals and the corresponding notions of precision. We illustrate our approach with three meta-analyses of clinical trials and observational studies.

JEL Classification: C13, C53, C83, I19.

Keywords: Conflicting evidence, confidence interval, prediction interval, information aggregation, meta-analysis, random-effects model.

Corresponding author: Andrey L. Vasnev, University of Sydney Business School, Abercrombie Building (H70), Sydney, NSW 2006, Australia

Email addresses: `jan@janmagnus.nl` (Magnus),
`andrey.vasnev@sydney.edu.au` (Vasnev).

1 Introduction

Suppose we are interested in the value of an unknown quantity β. We consult an expert who tells us that $\beta = 72$. The expert cannot be entirely certain, but she is confident that β lies between 70 and 74. After some time we consult a second, equally qualified, expert who tells us that $\beta = 58$. This expert is not certain either, but he is confident that β lies between 56 and 60. Based on this new information we decide to change our estimate from $\hat{\beta} = 72$ (the old information) to $\hat{\beta} = 65$ (the average of the old and the new information). But how much confidence should we have in this new estimate?

Let us think of the quantity β as the mean of a random variable y. The previous experiment gives us two observations: $y_1 = 72$ and $y_2 = 58$. Suppose we consult a third expert, what outcome could we expect? It seems reasonable, in the absence of other information, that we estimate the mean β of y_3 to be $\bar{y} = 65$. But what is a reasonable estimate of the variance of y_3?

The purpose of the current paper is to address both questions and discuss generalizations. The two questions are closely related, but they are not the same. The first question is answered by studying the distribution of $\hat{\beta}$ with mean β and variance $\operatorname{var}(\hat{\beta})$, leading to a confidence interval of the form

$$\hat{\beta} - 1.96\sqrt{\widehat{\operatorname{var}}(\hat{\beta})} < \beta < \hat{\beta} + 1.96\sqrt{\widehat{\operatorname{var}}(\hat{\beta})}, \tag{1}$$

while the second question is answered by studying the distribution of y with mean β and variance $\operatorname{var}(y)$, leading to a prediction interval of the form

$$\hat{\beta} - 1.96\sqrt{\widehat{\operatorname{var}}(y)} < y < \hat{\beta} + 1.96\sqrt{\widehat{\operatorname{var}}(y)}. \tag{2}$$

Now, $\operatorname{var}(\hat{\beta})$ tells us something about how precisely we can estimate the *location* of y, while $\operatorname{var}(y)$ tells us something about the *variability* of y. These two variances are thus conceptually (and numerically) quite different, and hence the answers to our two questions are also quite different.

A naive approach to the first question would be to argue as follows. We have two observations $y_1 = 72$ and $y_2 = 58$ with variances $v_1^2 = v_2^2 = 1$, and this leads to $\hat{\beta} = \bar{y} = 65$ with $\operatorname{var}(\hat{\beta}) = \operatorname{var}(\bar{y}) = 1/2$, and hence to a narrow confidence interval $63.6 < \beta < 66.4$. The combined estimate $\bar{y}$ has a greater precision than y_1 and y_2 individually, which makes sense because we have added information, and more information leads to more precision.

But does it? The two pieces of information are far apart, which happens frequently in practice. In Bayesian analysis, for example, the prior and the sample information may deliver conflicting messages. In the normal framework (normal prior, normal likelihood), the posterior mean (our estimate) lies somewhere in-between the mean of the prior and the mean of the sample, which is reasonable. The posterior variance will be *smaller* than the variance of the prior and the variance of the sample, which seems also reasonable, because we have added information, so the precision should increase. But it is also counter-intuitive, because the conflicting information makes us *less* confident about the resulting estimate: more information, less confidence.

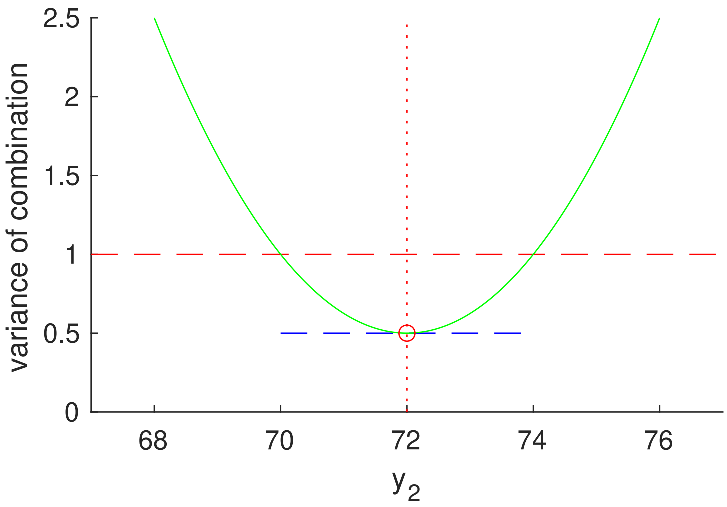

Figure 1: Common sense parabola, two observations, $y_1 = 72$, $v_1 = v_2 = 1$

Our questions are phrased in frequentist rather than in Bayesian terms, but the idea is the same, as illustrated in Figure 1. If only y_1 is available, then our estimate is $y_1 = 72$ with $\mathrm{var}(y_1) = 1$, the red dashed line. If both y_1 and y_2 are available, then standard statistical reasoning leads to an estimate $\bar{y} = 65$ with $\mathrm{var}(\bar{y}) = 1/2$, the blue dashed line, where we note that the variance does not depend on the value of y_2. Such a high precision (low variance) does not seem reasonable, and is not in line with common sense.[1] We should expect precision to increase when the two values support each other, and decrease when they contradict each other — something like the green 'common sense' parabola, given by the equation

$$V_{CS} = \frac{1}{2} + \frac{(y_2 - 72)^2}{8}. \tag{3}$$

When $70 < y_2 < 74$ (supporting evidence), the precision of the combination is higher than the precision of y_1, but otherwise the precision is lower

[1] In a recent survey among Japanese students, a similar question was asked, and more than one-half of the respondents indicated that the confidence interval should include the two observations, that is, the interval should be larger than $(58, 72)$; see Hanaki et al. (2023).

(conflicting evidence). The green parabola thus mimics common sense.

Is there a statistical theory that leads to such a figure? In fact, there is. The random-effects model used in meta-analysis essentially reconciles the apparent contradiction. We shall review and expand this theory in a maximum likelihood context, and attempt to answer both questions raised above.

When combining estimates or forecasts, there are two issues to be studied: the combined estimate or forecast, and its precision. Most (almost all) attention in the literature has gone to determining the 'best' point estimate based on a combination of point estimates, typically a weighed average of the underlying point estimates, where the weights are functions of their respective precisions; see for example Wang et al. (2023) for a recent review. But the precision of this estimate is equally important and is not yet fully understood. Measuring this uncertainty, especially in the presence of conflicting information, is our primary interest in the current paper.

We agree with Wooldridge (2023) when he emphasizes that there is a need for better and proper ways to compute standard errors — whether one takes a model-based, design-based, sampling-based, or even a Bayesian approach. Magnus and Vasnev (2023) study the critical role of correlation, Wang et al. (2024) provide a systematic review of interval forecasting, and Peng et al. (2024) propose a new aggregation operator to obtain more accurate interval prediction results. Menkveld et al. (2024) study the evidence-generating process across 164 teams and find that the variation across researchers (non-standard errors) is sizeable. However, to the best of our knowledge, the issue of conflict in the data is not discussed in these papers.

This is where meta-analysis pioneered by Glass (1976) can help, because in meta-analysis more attention is devoted to uncertainty. The standard tool to aggregate different studies in meta-analysis is the random-effects model (Borenstein et al., 2009), and this model is widely used to compare clinical studies, but also in many other areas such as economics (Havránek et al., 2020), organisation studies (Orlitzky at al., 2003), transportation (Button, 2019), supply chain management (Geng et al., 2017), and education and cognitive development (Peng et al., 2019). Different estimation methods, including maximum likelihood, have been suggested in the literature (Schmid et al., 2021, Section 4.5.2.2), and these are accessible through statistical packages, for example the metafor package in R (Viechtbauer, 2010).

Our approach is also through the random-effects model and maximum likelihood estimation, where we concentrate on three issues that seem to be outstanding or underdeveloped in meta-analysis. First, as in the forecast combination literature, the issue of conflict in the data has so far not been addressed. Second, no attention has been given to correlation between studies, which are typically assumed to be independent. Third, while meta-analyses usually focus on the mean treatment effect and its uncertainty, prediction intervals are also important but attract little attention in empirical studies; see however Schmid et al. (2021, Section 4.4.4.2).

Guddat et al. (2012) proposed a graphical representation of prediction intervals, but the adoption of this proposal has been slow. IntHout et al. (2016) showed that in almost three-quarters of statistically significant meta-analyses in the Cochrane Database of Systematic Reviews (Issues 2009–2013), the prediction interval contained the null effect, which means that, at least for some patient populations, the treatment might be more harmful than the comparator, in spite of the fact that the confidence interval of the point estimate suggests benefits. Borg et al. (2024) also advocate the use of prediction intervals, emphasizing their importance in accurately interpreting meta-analysis results in sports medicine and medical journals, and they note that only a few meta-analyses report such intervals, which can lead to misinterpretation of the results and potentially harmful treatments: in 2021, the proportion of studies reporting a prediction interval was 8% in sports and 15% in medicine. Seehra et al. (2021) perform a similar investigation for orthodontic meta-analyses, and find that only 19% of them reported prediction intervals. Borenstein (2020) stresses that prediction intervals provide important information about heterogeneity, while the commonly used I^2 index does not.

Yet, out of the above-mentioned studies, only IntHout et al. (2016) gained any momentum, as it was cited 1,527 times according to Google Scholar and 1,365 times according to scite.ai. Other studies had much less impact. Guddat et al. (2012) was cited 70 times (Google Scholar) and 58 times according (scite.ai), while Borenstein (2020) was cited 106 times (Google Scholar) and 96 times (scite.ai). The studies by Seehra et al. (2021) and Borg et al. (2024) are hardly cited at all.[2] Even when one or more of these studies is cited, the authors do not always follow the suggestions; see, for example, Botelho Miguel et al. (2025).

The remainder of the paper is organized as follows. Section 2 outlines the general setup of the random-effects model commonly used in meta-analysis. Section 3 presents the maximum likelihood estimators in the base case. Section 4 applies the theory to the situation where inputs are in conflict with each other, as discussed above. Section 5 extends the theory to the cases of relative precisions and correlated inputs. Section 6 demonstrates the effects of adopting our framework in three popular clinical trials. Section 7 concludes.

[2]Citations reported in Google Scholar and scite.ai as of 9 January 2025.

2 Meta-analysis and random effects

We consider the linear regression model

$$y = X\beta + u, \tag{4}$$

which is somewhat more general than the simple setup in the Introduction. Here, y denotes an $n \times 1$ vector of observations (typically called 'studies' or 'inputs'), X is an $n \times k$ matrix of nonrandom regressors, β is a $k \times 1$ vector of unknown coefficients, and u is an $n \times 1$ vector of random errors. In many cases of interest (as in the Introduction), the regressor matrix will be $X = \imath$ (the vector of ones) in which case the inputs y_i have a common mean β, but we shall not make this simplifying assumption just yet.

In standard regression one assumes that $\mathrm{E}(u) = 0$ and $\mathrm{var}(u) = \sigma^2 I_n$ which leads to the least-squares (LS) estimators for β and σ^2. But in meta-analysis one source of error does not suffice — we need two sources of error: errors of measurement *within* each of the inputs and errors of measurement *across* inputs. Thus, we assume that we can decompose the error vector u as

$$u = \zeta + \epsilon, \tag{5}$$

where u denotes 'system' noise, ζ denotes 'within-input' noise, and ϵ denotes 'across-input' noise.

We emphasize that, while borrowing language from the panel-data literature (within versus across), our data are y_i and not y_{ij}, that is, our data are one-dimensional, not two-dimensional. One can envisage a two-dimensional setting, where y_{ij} denotes the jth observation in the ith study. But this is not the usual setup in meta-analyses, where we simply look at the results of the ith study. This one-dimensionality implies that we have to be careful in the identification of the two components. For example, the introduction of autocorrelation is less trivial than in a panel-data setting, see Section 5.2. In Section 6 we shall see that an equicorrelation structure (which would be an obvious choice) is not possible, because the maximum likelihood estimator does not exist in that case.

If all inputs were measured perfectly (according to the authors), then $\zeta = 0$ and we are back at the (generalized) least-squares situation, where the only noise is generated by the errors across inputs. In the more realistic situation where within-input noise is present, we assume that $\mathrm{E}(\zeta) = \mathrm{E}(\epsilon) = 0$ and that ζ and ϵ are uncorrelated with $\mathrm{var}(\zeta) = V_\zeta$ and $\mathrm{var}(\epsilon) = V_\epsilon$. This implies that

$$\mathrm{E}(u) = 0, \qquad V = \mathrm{var}(u) = V_\zeta + V_\epsilon, \tag{6}$$

where V is assumed to be positive definite, while V_ζ and V_ϵ are positive semidefinite and may depend on unknown parameters.

To answer our first question from the Introduction, we need to estimate $\operatorname{var}(\hat{\beta}) = (X'V^{-1}X)^{-1}$. To answer the second question, we need to estimate the 'system variance' σ^2, which we define as the average variance of y, that is,

$$\sigma^2 = \operatorname{tr} V/n = \sigma_\zeta^2 + \sigma_\epsilon^2, \tag{7}$$

where

$$\sigma_\zeta^2 = \operatorname{tr} V_\zeta/n, \qquad \sigma_\epsilon^2 = \operatorname{tr} V_\epsilon/n. \tag{8}$$

To make further progress, we need to specify V_ζ and V_ϵ, decide how we wish to estimate the unknown parameters, and make sure these parameters are identified.

3 Maximum likelihood estimation in the base case

We shall estimate the unknown parameters by maximum likelihood (ML) and define the simplest possible 'base case,' where V depends on only one parameter. In particular, we specify

$$V_\epsilon = \sigma_\epsilon^2 I_n \qquad (\sigma_\epsilon^2 \geq 0), \tag{9}$$

and we assume the V_ζ is a known positive definite diagonal matrix, containing the variances of the underlying inputs. Notice that σ_ϵ^2 is identified, even in the case where V_ζ is proportional to the identity matrix. The variance $V = \sigma_\epsilon^2 I_n + V_\zeta$ thus depends on only one parameter. Assuming normality, the likelihood is given by

$$L = (2\pi)^{-n/2}|V|^{-1/2} \exp -\frac{1}{2}u'V^{-1}u. \tag{10}$$

Following Magnus (1978) we obtain the general formulas

$$2d \log L = -\operatorname{tr} V^{-1}(dV) + u'V^{-1}(dV)V^{-1}u + 2u'V^{-1}X(d\beta) \tag{11}$$

and

$$-\operatorname{E} d^2 \log L = \frac{1}{2} \operatorname{tr} V^{-1}(dV)V^{-1}(dV) + (d\beta)'X'V^{-1}X(d\beta). \tag{12}$$

In our base case we have $dV = (d\sigma_\epsilon^2)I_n$, so that (11) and (12) simplify to

$$2d \log L = \left(u'V^{-2}u - \operatorname{tr} V^{-1}\right)(d\sigma_\epsilon^2) + 2u'V^{-1}X(d\beta) \tag{13}$$

and

$$- \operatorname{E} d^2 \log L = \frac{1}{2}(\operatorname{tr} V^{-2})(d\sigma_\epsilon^2)^2 + (d\beta)'X'V^{-1}X(d\beta). \tag{14}$$

From (13) and (14) we obtain the first-order conditions:

$$\hat{\beta} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}y,$$
$$\operatorname{tr} \hat{V}^{-1} = (y - X\hat{\beta})'\hat{V}^{-2}(y - X\hat{\beta}). \tag{15}$$

We could try and solve these first-order conditions, but in many cases a more efficient method is to construct the concentrated (with respect to β) loglikelihood $\log L_c$ defined by

$$2 \log L_c = -n \log(2\pi) - \log |V_\zeta + \sigma_\epsilon^2 I_n| - \hat{u}'(V_\zeta + \sigma_\epsilon^2 I_n)^{-1}\hat{u}, \tag{16}$$

where

$$\hat{u} = y - X(X'(V_\zeta + \sigma_\epsilon^2 I_n)^{-1}X)^{-1}X'(V_\zeta + \sigma_\epsilon^2 I_n)^{-1}y. \tag{17}$$

Maximizing $\log L_c$ with respect to σ_ϵ^2 subject to the inequality constraint $\sigma_\epsilon^2 \geq 0$, yields the ML estimate $\hat{\sigma}_\epsilon^2$, from which we can compute $\hat{V} = \hat{\sigma}_\epsilon^2 I_n + V_\zeta$, and then $\hat{\beta} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}y$. There are many algorithms to maximize $\log L_c$. We used a simple grid search, which we found to be simple, fast, and accurate.

From (14) we obtain the information matrix and hence an approximation for the variances of our ML estimators:

$$\operatorname{var}(\hat{\beta}) \approx (X'V^{-1}X)^{-1}, \qquad \operatorname{var}(\hat{\sigma}_\epsilon^2) \approx \frac{2}{\operatorname{tr}(V^{-2})}. \tag{18}$$

In the special case $X = \imath$, we have

$$y_i = \beta + \zeta_i + \epsilon_i, \tag{19}$$

where $\operatorname{E}(\zeta_i) = \operatorname{E}(\epsilon_i) = 0$, all correlations are zero, and $\operatorname{var}(\zeta_i) = \sigma_{\zeta i}^2$ and $\operatorname{var}(\epsilon_i) = \sigma_\epsilon^2$. Letting

$$w_i = \frac{1}{\sigma_{\zeta i}^2 + \sigma_\epsilon^2}, \qquad \hat{w}_i = \frac{1}{\sigma_{\zeta i}^2 + \hat{\sigma}_\epsilon^2}, \tag{20}$$

we obtain the first-order conditions

$$\hat{\beta} = \frac{\sum_i \hat{w}_i y_i}{\sum_i \hat{w}_i}, \qquad \frac{\sum_i \hat{w}_i^2 (y_i - \hat{\beta})^2}{\sum_i \hat{w}_i} = 1, \tag{21}$$

and the estimated variance of $\hat{\beta}$,

$$\widehat{\operatorname{var}}(\hat{\beta}) = \frac{1}{\sum_i \hat{w}_i} = \frac{1}{n} \cdot \frac{1}{\sum_i \hat{w}_i / n}. \tag{22}$$

The solution $\hat{\beta}$ to (21) is known in the meta-analysis literature as the random-effects estimator. If $\sigma_\epsilon^2 = 0$, so that β is simply estimated by generalized least squares, we obtain the fixed-effects estimator.

In contrast, the system variance is estimated by

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\sigma_{\zeta i}^2 + \hat{\sigma}_\epsilon^2 = \frac{\sum_i (1/\hat{w}_i)}{n}. \tag{23}$$

The confrontation of (22) and (23) highlights the essential difference between the two questions raised in the Introduction. We have $\widehat{\text{var}}(\hat{\beta}) \leq \hat{\sigma}^2/n$ and, more generally, by Kantorovich' inequality,

$$1 \leq \frac{\hat{\sigma}^2/n}{\widehat{\text{var}}(\hat{\beta})} \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}, \tag{24}$$

where $\lambda_1 = \min_i \hat{w}_i$ and $\lambda_n = \max_i \hat{w}_i$; see Magnus and Neudecker (1988, 2019, p. 267).

If we specialize further and assume that $\sigma_{\zeta i}^2 = \sigma_\zeta^2$ (constant, as in our story in the Introduction), then w_i and $\hat{w}_i$ will also be constant, and we obtain $\widehat{\text{var}}(\hat{\beta}) = \hat{\sigma}^2/n$, but this is only true in this very special case.

4 Base case with two observations

Let us apply the base case proposed in the previous section to the problem discussed in the Introduction, and, in particular, present a figure resembling Figure 1 based on the random-effects approach.

Let $n = 2$, $y_1 = 72$, and assume that $v_1 = v_2 = 1$. Since $v_1 = v_2$, it follows that $w_1 = w_2 = w$, say, and hence the first-order conditions (21) simplify to $\hat{\beta} = \bar{y}$ and

$$\hat{\sigma}_\epsilon^2 + 1 = \frac{1}{\hat{w}} = \frac{1}{2}\sum_i (y_i - \bar{y})^2 = \frac{(y_2 - 72)^2}{4}, \tag{25}$$

so that, from (22),

$$\widehat{\text{var}}(\hat{\beta}) = \frac{1/\hat{w}}{2} = \begin{cases} (y_2 - 72)^2/8 & \text{if } |y_2 - 72| > 2, \\ 1/2 & \text{if } |y_2 - 72| \leq 2, \end{cases} \tag{26}$$

where the kink occurs because the variance in meta-analysis is the sum of two variances, and both must be nonnegative. This variance is represented by the blue parabola (labeled ML) in Figure 2. The variance of $\hat{\beta}$ is now only smaller than 1 when y_2 is 'close' to y_1, more precisely when $69.2 < y_2 < 74.8$.

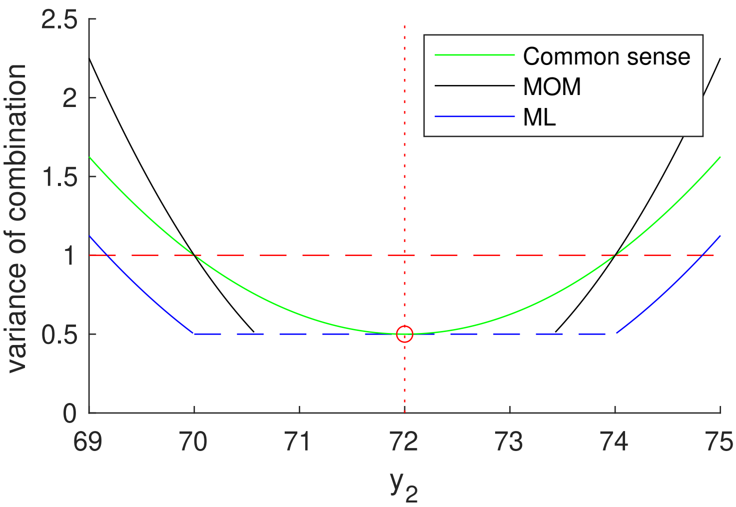

Figure 2: Random effects meta-analysis with two observations, $y_1 = 72$, $v_1 = v_2 = 1$

When y_2 is not close to y_1, then the variance can be large, and this is reflected in the confidence interval

$$\bar{y} - 1.96\sqrt{\widehat{\text{var}}(\hat{\beta})} < \beta < \bar{y} + 1.96\sqrt{\widehat{\text{var}}(\hat{\beta})}. \tag{27}$$

In particular, the confidence interval when $y_2 = 58$ is given by $55.3 < \beta < 74.7$, which is rather larger and more realistic than the naive confidence interval $63.6 < \beta < 66.4$ based on $\text{var}(\hat{\beta}) = 1/2$.

Our estimation approach is maximum likelihood, but one can estimate σ_ϵ^2 also by the method of moments (MOM). In that case, we obtain

$$\widehat{\text{var}}(\hat{\beta}) = \begin{cases} (y_2 - 72)^2/4 & \text{if } |y_2 - 72| > \sqrt{2}, \\ 1/2 & \text{if } |y_2 - 72| \leq \sqrt{2}, \end{cases} \tag{28}$$

using the formulas in Borenstein et al. (2009, pp. 72–74), leading to the black parabola (labeled MOM). The difference between (26) and (28) is caused by the fact that $\sum_i (y_i - \bar{y})^2$ is divided by $n = 2$ in the case of ML, and by $n - 1 = 1$ in the case of MOM. Either method of estimation leads to a figure which closely mimics the idealized green parabola proposed in Figure 1 and reproduced in Figure 2.

If our interest is in a prediction interval for y rather than a confidence interval of β, then we need

$$\hat{\sigma}^2 = \sigma_\zeta^2 + \hat{\sigma}_\epsilon^2 = \frac{(58 - 72)^2}{4} = 49, \tag{29}$$

using (25), leading to the prediction interval

$$51.3 = 65 - 13.7 = 65 - 1.96\hat{\sigma} < y < 65 + 1.96\hat{\sigma} = 65 + 13.7 = 78.7. \quad (30)$$

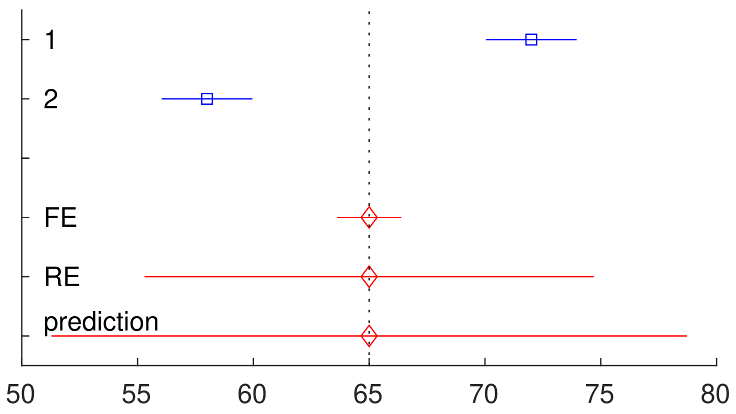

Figure 3: Base case with two observations,
confidence versus prediction intervals

Figure 3 illustrates this example. We have the two observations $y_1 = 58$ and $y_2 = 72$ with their standard deviations $v_1 = v_2 = 1$. The fixed-effects (FE, with $\sigma_\epsilon^2 = 0$) and the random-effects (RE, with σ_ϵ^2 estimated) estimators both provide confidence intervals of β. The FE estimator ignores the across-input noise ($\epsilon = 0$), and provides $\widehat{\text{var}}(\hat{\beta}) = 1/2$, which is much to small. The RE estimator takes both error sources into account, and provides the more reasonable estimate $\widehat{\text{var}}(\hat{\beta}) = 49/2$. The prediction interval for y is wider than the corresponding RE confidence interval, as it is based on $\widehat{\text{var}}(y) = 49$.

5 Extensions

The base case can be generalized in various directions. We shall discuss two generalizations where the variance matrix V depends on two (rather than on one) unknown parameters.

5.1 Relative precisions

In meta-analysis it is quite common to interpret the input variances in V_ζ as an indication of the relative rather than the absolute precision of the inputs, and our first extension analyzes the consequences of this situation. Let V_0

denote the diagonal matrix of (absolute) input variances and define

$$V_\zeta = \sigma_\zeta^2 V_0^*, \qquad V_0^* = \frac{V_0}{\operatorname{tr} V_0/n}. \tag{31}$$

The parameter σ_ζ^2 has the same interpretation as before, because $\sigma_\zeta^2 = \operatorname{tr} V_\zeta/n$, but now it is a parameter to be estimated, while in the base case it was set equal to the constant $\operatorname{tr} V_0/n$.

Since $V = \sigma_\zeta^2 V_0^* + \sigma_\epsilon^2 I_n$, we have $dV = (d\sigma_\zeta^2)V_0^* + (d\sigma_\epsilon^2)I_n$ and hence, from (11) and (12),

$$\begin{aligned} 2d\log L &= \left(u'V^{-1}V_0^*V^{-1}u - \operatorname{tr} V^{-1}V_0^*\right)(d\sigma_\zeta^2) \\ &\quad + \left(u'V^{-2}u - \operatorname{tr} V^{-1}\right)(d\sigma_\epsilon^2) + 2u'V^{-1}X(d\beta) \end{aligned} \tag{32}$$

and

$$\begin{aligned} -\operatorname{E} d^2\log L &= \frac{1}{2}\operatorname{tr} V^{-1}V_0^*V^{-1}V_0^*(d\sigma_\zeta^2)^2 + \operatorname{tr} V^{-1}V_0^*V^{-1}(d\sigma_\zeta^2)(d\sigma_\epsilon^2) \\ &\quad + \frac{1}{2}\operatorname{tr} V^{-2}(d\sigma_\epsilon^2)^2 + (d\beta)'X'V^{-1}X(d\beta). \end{aligned} \tag{33}$$

The first-order conditions are therefore

$$\begin{aligned} \hat\beta &= (X'\hat V^{-1}X)^{-1}X'\hat V^{-1}y, \\ \operatorname{tr} \hat V^{-1}V_0^* &= (y - X\hat\beta)'\hat V^{-1}V_0^*\hat V^{-1}(y - X\hat\beta), \\ \operatorname{tr} \hat V^{-1} &= (y - X\hat\beta)'\hat V^{-2}(y - X\hat\beta), \end{aligned} \tag{34}$$

generalizing (15).

To find the ML estimates, it is, as in Section 3, computationally more efficient to construct the concentrated loglikelihood $\log L_c$ defined by (16) and (17), where $V_\zeta = \sigma_\zeta^2 V_0^*$. Maximizing $\log L_c$ with respect to σ_ζ^2 and σ_ϵ^2 subject to the inequality constraints $\sigma_\zeta^2 \geq 0$ and $\sigma_\epsilon^2 \geq 0$ then yields the required ML estimates. The variance of $\hat\beta$ is again approximated by $(X'V^{-1}X)^{-1}$, and the variance matrix of the variance components by

$$\operatorname{var}\begin{pmatrix} \hat\sigma_\zeta^2 \\ \hat\sigma_\epsilon^2 \end{pmatrix} \approx 2\begin{pmatrix} \operatorname{tr}(V^{-1}V_0^*)^2 & \operatorname{tr} V^{-1}V_0^*V^{-1} \\ \operatorname{tr} V^{-1}V_0^*V^{-1} & \operatorname{tr} V^{-2} \end{pmatrix}^{-1} \tag{35}$$

generalizing (18).

Although the relative-precisions approach is quite common, we do not recommend it, as the interpretation of within-errors versus across-errors gets confused. We shall see examples of this in Section 6.

5.2 Correlated inputs

As a second extension of the base case we consider the situation where the inputs may be correlated. Let P denote the first-order autocorrelation matrix

$$P = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-2} & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-3} & \rho^{n-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \dots & 1 & \rho \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & \rho & 1 \end{pmatrix}. \tag{36}$$

The question then arises where to place the autocorrelation. In Section 2 we defined V_ζ as the variance of the within-input noise and V_ϵ as the variance of the across-input noise, and it seems plausible that autocorrelation due to the fact that researchers know each others' results shows up in the across-input variance rather than in the within-input variance. But these labels (borrowed from the panel-data literature) should be interpreted with some caution, because in the current study we are *not* dealing with panel data, and hence the placement of autocorrelation is less obvious.

To illustrate, we may specify the variance matrix either as

$$V_1 = V_\zeta + V_\epsilon, \qquad V_\zeta = V_0^{1/2} P V_0^{1/2}, \qquad V_\epsilon = \sigma_\epsilon^2 I_n, \tag{37}$$

or as

$$V_2 = V_\zeta + V_\epsilon, \qquad V_\zeta = V_0, \qquad V_\epsilon = \sigma_\epsilon^2 P, \tag{38}$$

where V_0 denotes again the diagonal matrix of input variances $\sigma_1^2, \dots, \sigma_n^2$. To compare V_1 with V_2, let $\tau_i^2 = \sigma_i^2/\sigma_\epsilon^2$ and note that the diagonal elements are the same in the two matrices:

$$(V_1)_{ii} = (V_2)_{ii} = \sigma_i^2 + \sigma_\epsilon^2 = \sigma_\epsilon^2(1 + \tau_i^2), \tag{39}$$

and the off-diagonal elements are given by

$$(V_1)_{ij} = \sigma_i \sigma_j \rho^{|i-j|} = \tau_i \tau_j \sigma_\epsilon^2 \rho^{|i-j|}, \qquad (V_2)_{ij} = \sigma_\epsilon^2 \rho^{|i-j|}, \tag{40}$$

respectively. The question is what is the more realistic representation of the correlation between studies. We can write the off-diagonal elements in the correlation matrices corresponding to V_1 and V_2 as

$$(P_1)_{ij} = \frac{\rho^{|i-j|}}{\sqrt{(1 + \tau_i^{-2})(1 + \tau_j^{-2})}}, \qquad (P_2)_{ij} = \frac{\rho^{|i-j|}}{\sqrt{(1 + \tau_i^2)(1 + \tau_j^2)}}, \tag{41}$$

If the τ_i are 'small,' that is, if σ_ϵ is 'large' relative to the σ_i, then $(P_1)_{ij} \approx 0$ and $(P_2)_{ij} \approx \rho^{|i-j|}$, while if the τ_i are 'large,' that is, if σ_ϵ is 'small' relative to the σ_i, then $(P_1)_{ij} \approx \rho^{|i-j|}$ and $(P_2)_{ij} \approx 0$.

Whether $V = V_1$ or $V = V_2$, the diagonal elements of V_ζ are the same as the diagonal elements of V_0, so that

$$\sigma_\zeta^2 = \operatorname{tr} V_\zeta / n = \operatorname{tr} V_0 / n \tag{42}$$

is not affected by the correlation structure. But $\hat{\beta}$ and $\hat{\sigma}_\epsilon$ (and hence $\hat{\sigma}$) will be affected by the correlation structure.

In the case $V = V_1$ we obtain, from (10), the concentrated (with respect to β) loglikelihood

$$2 \log L_c = -n \log(2\pi) - \log |V_1| - \hat{u}' V_1^{-1} \hat{u}, \tag{43}$$

where

$$\hat{u} = y - X(X'V_1^{-1}X)^{-1}X'V_1^{-1}y. \tag{44}$$

If $P = I_n$, we obtain (16) as a special case. Letting $\dot{P}$ denote the $n \times n$ matrix containing the derivatives of the elements of P with respect to ρ, that is, $\dot{P}_{ij} = dP_{ij}/d\rho$, we obtain

$$dV_1 = (d\rho) V_0^{1/2} \dot{P} V_0^{1/2} + (d\sigma_\epsilon^2) I_n, \tag{45}$$

so that the variance matrix of the variance components can be approximated by

$$\operatorname{var} \begin{pmatrix} \hat{\rho} \\ \hat{\sigma}_\epsilon^2 \end{pmatrix} \approx 2 \begin{pmatrix} \operatorname{tr}(V_1^{-1} V_0^{1/2} \dot{P} V_0^{1/2})^2 & \operatorname{tr} V_1^{-1} V_0^{1/2} \dot{P} V_0^{1/2} V_1^{-1} \\ \operatorname{tr} V_1^{-1} V_0^{1/2} \dot{P} V_0^{1/2} V_1^{-1} & \operatorname{tr} V_1^{-2} \end{pmatrix}^{-1}. \tag{46}$$

In the case $V = V_2$ we obtain

$$dV_2 = (d\sigma_\epsilon^2) P + (d\rho) \sigma_\epsilon^2 \dot{P}, \tag{47}$$

so that the variance matrix of the variance components can be approximated by

$$\operatorname{var} \begin{pmatrix} \hat{\rho} \\ \hat{\sigma}_\epsilon^2 \end{pmatrix} \approx 2 \begin{pmatrix} \sigma_\epsilon^4 \operatorname{tr}(V_2^{-1} \dot{P})^2 & \sigma_\epsilon^2 \operatorname{tr} V_2^{-1} \dot{P} V_2^{-1} P \\ \sigma_\epsilon^2 \operatorname{tr} V_2^{-1} P V_2^{-1} \dot{P} & \operatorname{tr}(V_2^{-1} P)^2 \end{pmatrix}^{-1}. \tag{48}$$

6 Some applications to the meta-analysis of heterogeneous clinical trials

Meta-analysis has been particularly popular in the assessment of clinical trials and observational studies (DerSimonian and Laird, 1986, 2015), and three of these analyses were reviewed by Doi et al. (2015a,b). The research questions underlying these three clusters of studies were:

1. Are diuretics (substances that promote an increased production of urine) useful in the prevention of pre-eclampsia (a disorder in late pregnancy)? (Collins et al., 1985);
2. Is fruit and vegetable consumption good for you? (Wang et al., 2014);
3. Does eating beans, chickpeas, lentils, and peas (so-called dietary pulses) help to improve dyslipidemia (a metabolic disorder and risk factor for developing heart diseases)? (Ha et al., 2014).

Doi et al. (2015a,b) criticise the random-effects confidence intervals of β used in the three meta-analyses as being too narrow, and they propose the 'IVhet' (inverse variance heterogeneity) method which produces wider intervals.[3] These wider confidence intervals seem to reconcile some of the contradictions in the data, but perhaps they are too wide, as they lead to negative answers to each of the three above research questions, and hence to rejections of each of the implied null hypotheses. Specifically, meta-analysis using the IVhet method finds no evidence that eating fresh fruit and vegetables is good for you.[4] Our ML-based formulae for the random-effects model also reconcile the contradictions, and the results do agree with common sense, as they typically produce slightly narrower confidence intervals of β than IVhet, so that a hypothesis about β which is not rejected under IVhet may be rejected under ML.

Table 1: Doi examples — the base case

Data	n	$\hat{\sigma}_\epsilon$	σ_ζ	$\hat{\sigma}$	$\sigma_\zeta^2/\hat{\sigma}^2$
Doi-1	9	0.485	0.441	0.656	0.453
Doi-1C	5	0.000	0.696	0.696	1.000
Doi-2	7	0.021	0.023	0.031	0.537
Doi-3	25	0.195	0.287	0.347	0.684

Referring to the three studies as Doi-1, Doi-2, and Doi-3, respectively, we have $n = 9$ inputs in Doi-1, 7 inputs in Doi-2, and 25 inputs in Doi-3. The

[3] The IVhet method uses the fixed-effects estimator, but inflates its variance using a quasi-likelihood approach with a scale parameter based on an inter-class correlation.

[4] There is, of course, the possibility that meta-analysis through the IVhet method is right and general consensus is wrong, perhaps through a confounding variable, wealth. Rich people eat more fresh fruit and vegetables (expensive) than poor people, and they have easier access to health care. Hence, they are healthier, not because they eat more fruit but because they get better medical support.

Doi-1 study, taken from Collins et al. (1985), was revisited by Churchill et al. (2007) in the Cochrane Library[5]. The latter study drops six inputs from Doi-1 (so that only three inputs remain, labeled 1, 4, and 7), and adds a more recent input by Sibai et al. (1984). The Landesman et al. (1965) input (labeled 6 in Doi-1) was awaiting assessment at the time, but we include it, so that we have a sufficient number of points to estimate our parameters in the extensions. We refer to this additional study as Doi-1C, and it has $n = 5$ inputs.

Each input has a reported variance $\sigma^2_{\zeta i}$ from which we compute $\sigma^2_\zeta = (1/n)\sum_i \sigma^2_{\zeta i}$. We next estimate σ^2_ϵ as described in Section 3, from which we obtain the system variance $\hat{\sigma}^2 = \sigma^2_\zeta + \hat{\sigma}^2_\epsilon$ and the ratio $\sigma^2_\zeta/\hat{\sigma}^2$. The results presented in Table 1 show that about one-half of the system variance can be attributed to within-input noise and one-half to across-input noise. In Doi-1C, all variance is attributed to within-input noise.

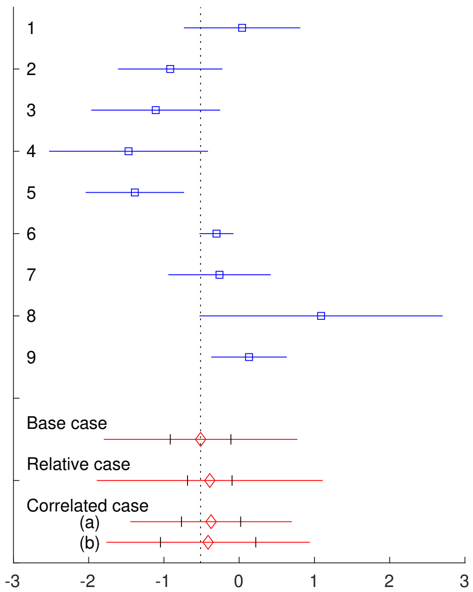

Figure 4: Doi-1, prediction and confidence intervals

A graphical illustration of the Doi-1 example is provided in Figure 4. The nine inputs are depicted in blue together with their confidence intervals based on $\sigma_{\zeta i}$. The prediction interval for y (the top red line) is $-1.80 < y < 0.77$ and the confidence interval of β (indicated by the ticks) is $-0.92 < \beta < -0.11$, so that the null hypothesis $\beta = 0$ is rejected, and diuretics are thus likely to be useful in the prevention of pre-eclampsia. This conclusion agrees with

[5]See https://www.cochranelibrary.com

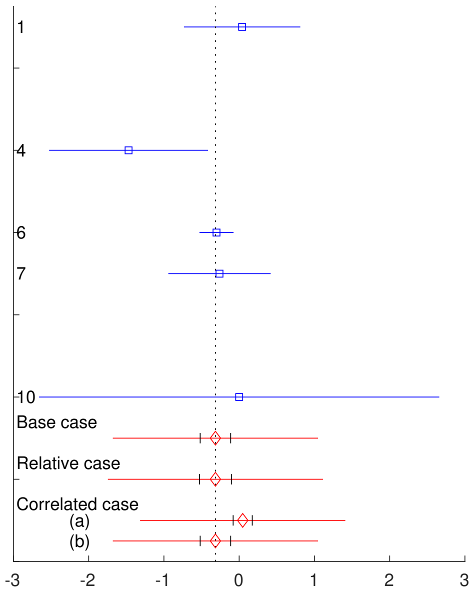

Figure 5: Doi-1C, prediction and confidence intervals

Collins et al. (1985), but not with Doi et al. (2015a, Section 5).

The figure also demonstrates the difference between the confidence interval of β and the prediction interval for y, which is more than $\sqrt{n} = 3$ times as wide in accordance with the inequality (24). Thus, while the null hypothesis $\beta = 0$ is rejected on the basis of nine inputs, we would expect a hypothetical tenth study to produce a value between -1.80 and 0.77, and it is this wider interval which should be employed when we wish to summarize the nine inputs into one input.

The Doi-1C example is illustrated in Figure 5 with inputs 1, 4, 6, and 7 together with the new input 10 (corresponding to Sibai et al., 1984). The prediction interval for y (the top red line) is $-1.68 < y < 1.05$ and the confidence interval of β (indicated by the ticks) is $-0.52 < \beta < -0.11$, so that the null hypothesis $\beta = 0$ is still rejected, and diuretics are thus likely to be useful in the prevention of pre-eclampsia.

Our conclusions are confirmed by the Doi-2 and Doi-3 examples. Using the data set Doi-2 we find $0.93 < \beta < 0.98$ (null hypotheses $\beta = 1$ is rejected), while Doi et al. (2015a) do not reject the null hypothesis. The prediction interval for y is $0.89 < y < 1.02$, more than $\sqrt{7} = 2.65$ the width of the confidence interval. Similarly, using the data set Doi-3 we find $-0.27 < \beta < -0.08$ (null hypothesis $\beta = 0$ is rejected), while Doi et al. (2015b) do not reject the null hypothesis. The prediction interval for y is $-0.86 < y < 0.50$, more than 5 times the width of the confidence interval.

Table 2: Doi examples — the relative case

Data	$\hat{\sigma}_\epsilon$	$\hat{\sigma}_\zeta$	$\hat{\sigma}$	$\hat{\sigma}_\zeta^2/\hat{\sigma}^2$
Doi-1	0.000	0.766	0.766	1.000
Doi-1C	0.000	0.729	0.729	1.000
Doi-2	0.025	0.000	0.025	0.000
Doi-3	0.077	0.527	0.533	0.979

Our ML approach allows us to consider two extensions as described in Section 5. First, we may estimate an additional parameter to allow the precisions of our inputs to be interpreted as relative rather than as absolute precisions. Table 2 summarizes these results and leads to the confidence intervals

$$-0.69 < \beta < -0.09, \qquad 0.93 < \beta < 0.97, \qquad -0.20 < \beta < -0.05, \tag{49}$$

and the corresponding prediction intervals

$$-1.89 < y < 1.11, \qquad 0.90 < y < 1.00, \qquad -1.17 < y < 0.92, \tag{50}$$

for Doi-1, Doi-2, and Doi-3, respectively, where the results for Doi-1 are presented as the second red line in Figure 4. The results for Doi-1C are presented in Figure 5, and the confidence intervals are

$$-0.53 < \beta < -0.10, \qquad -1.75 < y < 1.11. \tag{51}$$

While the conclusions do not change — the null hypotheses are still rejected in all four cases — we notice from Table 2 that the relative method tends to push one of the two variances towards zero. In Doi-1, Doi-1C, and Doi-3, the error ϵ essentially vanishes, so that all variation is captured by the within-input noise ζ; and in Doi-2, the error ζ vanishes, so that all variation is captured by the across-input noise ϵ. The two parameters σ_ζ and σ_ϵ are still identified, but to allow relative precisions for the inputs makes the separation of the two effects more difficult. This method is therefore not recommended.

Our second extension is different. Here we allow the inputs to be correlated, which makes sense because it is likely that authors are familiar with and influenced by earlier studies (although it is not immediately obvious whether we should expect the correlation to be positive or negative). One possible specification for the correlation structure would be to assume that all correlations are the same (equicorrelation). However, no ML solution exists in this case (De Luca et al., 2025). Another possibility is to assume a first-order autoregressive scheme, based upon the publication dates of the studies. Here we have two options as explained in Section 5.2, and we refer to the option corresponding to V_1 as case (a) and the option corresponding to V_2 as case (b).

Table 3: Doi examples — the correlated case

	Data	$\hat{\sigma}_\epsilon$	σ_ζ	$\hat{\sigma}$	$\sigma_\zeta^2/\hat{\sigma}^2$	$\hat{\rho}$
case (a)	Doi-1	0.324	0.441	0.547	0.650	0.518
	Doi-1C	0.000	0.696	0.696	1.000	−0.536
	Doi-2	0.022	0.023	0.032	0.519	1.000
	Doi-3	0.194	0.287	0.346	0.687	−0.041
case (b)	Doi-1	0.531	0.441	0.691	0.408	0.572
	Doi-1C	0.000	0.696	0.696	1.000	−0.280
	Doi-2	0.021	0.023	0.031	0.552	−0.279
	Doi-3	0.166	0.287	0.331	0.749	−0.443

When we implement first-order autoregression into our ML procedure, we obtain the results in Table 3. The implied intervals in case (a) are

$$-0.77 < \beta < 0.02, \qquad 0.94 < \beta < 1.00, \qquad -0.35 < \beta < -0.16, \tag{52}$$

and

$$-1.45 < y < 0.70, \qquad 0.90 < y < 1.03, \qquad -0.93 < y < 0.42, \tag{53}$$

and in case (b) they are

$$-1.05 < \beta < 0.22, \qquad 0.94 < \beta < 0.97, \qquad -0.23 < \beta < -0.11, \tag{54}$$

and

$$-1.77 < y < 0.94, \qquad 0.90 < y < 1.02, \qquad -0.82 < y < 0.48, \tag{55}$$

for Doi-1, Doi-2, and Doi-3, respectively. The intervals for Doi-1 are graphically presented in Figure 4: the third red line for case (a) and the forth red line for case (b). The corresponding results for Doi-1C are presented in Figure 5. The confidence intervals are $-0.08 < \beta < 0.17$ and $-1.32 < y < 1.41$ for case (a), and $-0.52 < \beta < -0.11$ and $-1.68 < y < 1.05$ for case (b).

In case (a), the correlation estimate $\hat{\rho}$ is mildly positive in Doi-1. The intervals become slightly shorter, but the effect on the estimates and the intervals is small. In Doi-1C, the correlation is negative, and the interval is shifted to the right, though the interval for β still includes zero as in Doi-1. The studies in Doi-2 appear to be very highly correlated, but the effect on the implied intervals is again small. The studies in Doi-3 are essentially uncorrelated and the effect on the estimates is therefore negligible.

In case (b), the Doi-1 correlation increases slightly and the confidence intervals widen, while the Doi-1C correlation becomes less negative; the interval shifts to the left and is now comparable to the base and relative cases. In Doi-2 and Doi-3, the correlation changes are more substantial (even changing the sign in Doi-2), but the intervals are not sensitive to this change. The possibility to take correlation between the inputs into account seems important, and our theory allows us to do this.

7 Concluding remarks

This paper has been concerned with the somewhat counter-intuitive situation that more information leads to less precision, and we have seen that a decomposition of the error into within-input and across-input components is sufficient to reconcile the conflict. Error decomposition is a standard tool in many areas, but the data are then typically two-dimensional. For example, if we wish to explain household consumption y_{it} of the ith household at time t, then it is quite common to write the error term as $u_{it} = \zeta_i + \eta_t + \epsilon_{it}$ (three error components) or as $u_{it} = \zeta_i + \epsilon_{it}$ or $u_{it} = \eta_t + \epsilon_{it}$ (two error components). In the current paper we have only one dimension and still we wish to write the error term as a sum of two orthogonal components, because we need to account for two different types of noise — a situation recently highlighted by Kahneman et al. (2021).

More information can lead to less precision, but can less precision also lead to more information? Rothenberg (2005) asked the following question:

> Suppose we wish to know the length and the width of a rectangular table based on n observations on the area of the table. Can we estimate the length and the width?

It turns out that we can, but only if the data are sufficiently noisy. If there is no noise, then all measurements of the area are the same and we cannot recover the length and width. If there is almost no noise, then we can recover the length and width, but only very imprecisely. If there is too much noise, then our estimates will also be bad. Hence, there exists some optimal level of noise which will give the best estimates: more (but not too much) noise produces more information.

Acknowledgements

We are grateful to the editor and the referee for their constructive comments.

References

Borenstein, M. (2020). Research Note: In a meta-analysis, the I^2 index does not tell us how much the effect size varies across studies, *Journal of Physiotherapy*, 66(2), 135–139.

Borenstein, M., L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein (2009). *Introduction to Meta-Analysis*, John Wiley: Chichester/New York.

Borg, D. N., F. M. Impellizzeri, S. J. Borg, K. P. Hutchins, I. B. Stewart, T. Jones, B. J. Baguley, L. B. R. Orssatto, A. J. E. Bach, J. O. Osborne, B. S. McMaster, R. L. Buhmann, J. J. Bon, and A. G. Barnett (2024). Meta-analysis prediction intervals are under reported in sport and exercise medicine, *Scandinavian Journal of Medicine & Science in Sports*, 34(3):e14603. doi: 10.1111/sms.14603. PMID: 38501202.

Button, K. (2019). The value and challenges of using meta-analysis in transportation economics, *Transport Reviews*, 39(3), 293–308.

Churchill, D., G. D. G Beevers, S. Meher, and C. Rhodes (2007). Diuretics for preventing pre-eclampsia, *Cochrane Database of Systematic Reviews*, Issue 1, Art. No.: CD004451.

Collins, R., S. Yusuf, and R. Peto (1985). Overview of randomised trials of diuretics in pregnancy, *British Medical Journal (Clinical Research Edition)*, 290 (6461), 17–23.

De Luca, G., J. R. Magnus, and A. L. Vasnev (2025). Maximum likelihood estimation of the linear model with equicorrelated errors, *Communications in Statistics—Theory and Methods*, to appear.

DerSimonian, R. and N. M. Laird (1986). Meta-analysis in clinical trials, *Controlled Clinical Trials*, 7, 177–188.

DerSimonian, R. and N. M. Laird (2015). Meta-analysis in clinical trials revisited, *Contemporary Clinical Trials*, 45, 139–145.

Doi, S. A. R., J. J. Barendrecht, S. Khan, L. Thalib, and G. M. Williams (2015a). Advances in the meta-analysis of heterogeneous clinical trials

I: The inverse variance heterogeneity model, *Contemporary Clinical Trials*, 45, 130–138.

Doi, S. A. R., J. J. Barendrecht, S. Khan, L. Thalib, and G. M. Williams (2015b). Advances in the meta-analysis of heterogeneous clinical trials II: The quality effects model, *Contemporary Clinical Trials*, 45, 123–129.

Geng, R., S. A. Mansouri, and E. Aktas (2017). The relationship between green supply chain management and performance: A meta-analysis of empirical evidences in Asian emerging economies, *International Journal of Production Economics*, 183, 245–258.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research, *Educational Researcher*, 5, 3–8.

Guddat, C., U. Grouven, R. Bender, and G. Skipka (2012). A note on the graphical presentation of prediction intervals in random-effects meta-analyses, *Systematic Reviews*, 1:34.
https://doi.org/10.1186/2046-4053-1-34.

Ha, V., J. L. Sievenpiper, R. J. de Souza, V. H. Jayalath, A. Mirrahimi, A. Agarwal, L. Chiavaroli, S. Blanco Mejia, F. M. Sacks, M. Di Buono, A. M. Bernstein, L. A. Leiter, P. M. Kris-Etherton, V. Vuksan, R. P. Bazinet, R. G. Josse, J. Beyene, C. W. C. Kendall, and D. J. A. Jenkins (2014). Effect of dietary pulse intake on established therapeutic lipid targets for cardiovascular risk reduction: A systematic review and meta-analysis of randomized controlled trials, *Canadian Medical Association Journal*, 186(8), E252–E262.

Hanaki, N., J. R. Magnus, and D. Yoo (2023). Statistics and common sense, *Journal of Statistics and Data Science Education*, 31, 295–304.

Havránek, T., T. D. Stanley, H. Doucouliagos, P. Bom, J. Geyer-Klingeberg, I. Iwasaki, W. R. Reed, K. Rost, and R. C. M. van Aert (2020). Reporting guidelines for meta-analysis in Economics, *Journal of Economic Surveys*, 34, 469–475.

IntHout, J., J. P. A. Ioannidis, M. M. Rovers, and J. J. Goeman (2016). Plea for routinely presenting prediction intervals in meta-analysis, *BMJ Open*, 6:e010247. doi:10.1136/bmjopen-2015-010247.

Kahneman, D., O. Sibony, and C. Sunstein (2021). *Noise: A Flaw in Human Judgment*, Little, Brown Spark: New York.

Landesman, R., O. Aguero, K. Wilson, R. La Russa, W. Campbell, and O. Penaloza (1965). The prophylactic use of chlorthalidone, a sulfonamide diuretic in pregnancy, *Journal of Obstetrics and Gynaecology of the British Commonwealth*, 72, 1004–1010.

Magnus, J. R. (1978). Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix, *Journal of Econometrics*, 7, 281–312.

Magnus, J. R. and H. Neudecker (1988, 2019). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, third edition, John Wiley: Chichester/New York.

Magnus, J. R. and A. L. Vasnev (2023). On the uncertainty of a combined forecast: The critical role of correlation, *International Journal of Forecasting*, 39(4), 1895–1908.

Menkveld, A. J., A. Drieber, F. Holzmeister, J. Huber, M. Johannesson, M. Kirchler, S. Neusüss, M. Razen, U. Weitzel, et al. (2024). Non-standard errors, *Journal of Finance*, 00, 000–000.

Botelho Miguel, C., R. de Souza Andrade, L. Mazurek, M. C. Martins-de-Abreu, J. Miguel-Neto, A. de Melo Barbosa, G. P. Silva, A. Góes-Neto, S. de Castro Soares, J. E. Lazo-Chica, and W. F. Rodrigues (2025). Emerging pharmacological interventions for chronic venous insufficiency: A comprehensive systematic review and meta-analysis of efficacy, safety, and therapeutic advances, *Pharmaceutics*, 17(1), 59, 1–27.

Orlitzky, M., F. L. Schmidt, and S. L. Rynes (2003). Corporate social and financial performance: A meta-analysis, *Organization Studies*, 24(3), 403–441.

Peng, K., C. Kang, X. Ru, and L. Zhou (2024). The optimal interval combination prediction model based on vectorial angle cosine and a new aggregation operator for social security level prediction, *Journal of Forecasting*, 43(2), 490–505.

Peng, P., T. Wang, C. Wang, and X. Lin (2019). A meta-analysis on the relation between fluid intelligence and reading/mathematics: Effects of tasks, age, and social economics status, *Psychological Bulletin*, 145(2), 189–236.

Rothenberg, T. J. (2005). Incredible structural inference, in: *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg* (Eds: D. W. K. Andrews and J. H. Stock), Cambridge University Press: New York, pp. 3–10.

Seehra, J., D. Stonehouse-Smith, and N. Pandis (2021). Prediction intervals reporting in orthodontic meta-analyses, *European Journal of Orthodontics*, 43(5), 596–600.

Sibai, B. M., R. A. Grossman, and H. G. Grossman (1984). Effects of diuretics on plasma volume in pregnancies with long-term hypertension, *American Journal of Obstetrics and Gynecology*, 150, 831–835.

Schmid, C. H., T. Stijnen, and I. R. White (2021). *Handbook of Meta-Analysis*, CRC Press: New York.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package, *Journal of Statistical Software*, 36(3), 1–48. https://doi.org/10.18637/jss.v036.i03.

Wang, P., S. H. Gurmani, Z. Tao, J. Liu, and H. Chen (2024). Interval time series forecasting: A systematic literature review, *Journal of Forecasting*, 43(2), 249–285.

Wang, X., R. J. Hyndman, F. Li, and Y. Kang (2023). Forecast combinations: An over 50-year review, *International Journal of Forecasting*, 39(4), 1518–1547.

Wang, X., Y. Ouyang, J. Liu, M. Zhu, G. Zhao, W. Bao, and F. B. Hu (2014). Fruit and vegetable consumption and mortality from all causes, cardiovascular disease, and cancer: Systematic review and dose-response meta-analysis of prospective cohort studies, *British Medical Journal*, 349, g4490.

Wooldridge, J. M. (2023). What is a standard error? (And how should we compute it?), *Journal of Econometrics*, 237, 105517.