On the uncertainty of a combined forecast: The critical role of correlation

December 21, 2021

Jan R. Magnus
Vrije Universiteit, Amsterdam, The Netherlands
and Tinbergen Institute, Amsterdam, The Netherlands

Andrey L. Vasnev
University of Sydney, New South Wales, Australia

Abstract: The purpose of this paper is to show that the effect of the zero-correlation assumption in combining forecasts can be huge, and that ignoring (positive) correlation can lead to confidence bands around the forecast combination that are much too narrow. In the typical case where three or more forecasts are combined, the estimated variance increases without bound when correlation increases. Intuitively, this is because similar forecasts provide little information if we know that they are highly correlated. Although we concentrate on forecast combinations and confidence bands, our theory applies to any statistic where the observations are linearly combined. We apply our theoretical results to explain why forecasts by Central Banks (in our case, the Bank of Japan) are so frequently misleadingly precise. In most cases, a correlation above 0.7 is required to produce reasonable confidence bands.

Key words: Combining information, Correlation, Growth forecasting, Inflation forecasting, Central Banks.

JEL Codes: C52, C53, E31, E37, O40.

Corresponding author:
Andrey L. Vasnev
The University of Sydney Business School Abercrombie Building (H70)
Sydney, NSW 2006
Australia
E-mail: andrey.vasnev@sydney.edu.au

1 Introduction

Models are defined by the restrictions they impose, and a good model is not sensitive to small violations of these restrictions. For example, if we think of approximating a function $y = \mu(X)$, then the linear function $\mu(X) = X\beta$ provides a first-order approximation, so that even if μ is in fact nonlinear, the linear model may still work well in practice. Similarly, the normal distribution serves as an approximation, in this case a second-order approximation, because for an arbitrary density function $f(u)$ with a maximum at $u = 0$ (the mode), the derivative of f vanishes at the mode, so that

$$\log f(u) \approx \log f(0) + (1/2)u'H(0)u, \tag{1}$$

where $H(u)$ denotes the Hessian matrix. Hence,

$$f(u) \approx f(0) \exp[(-1/2)u'(-H(0))u], \tag{2}$$

which we recognize as the normal distribution with mean zero and variance $[-H(0)]^{-1}$. This tells us that the normal distribution serves as a good approximation to an arbitrary density function, certainly at the center of the distribution.

Things are different with the assumption of zero-correlation between the observations. Letting r denote the correlation coefficient, the problem is not so much that the function of interest may be ill-behaved around $r = 0$, but rather that, if there is correlation, then this correlation may not be close to zero, and that the behavior of the function at $r = 0.7$ or $r = 0.9$ may be very different from its behavior at $r = 0$.

The purpose of this paper is to show that the effect of the zero-correlation assumption in combining forecasts can be huge, and that ignoring (positive) correlation can lead to confidence bands around the forecast combination that are much too narrow. Although we concentrate on forecast combinations and confidence bands, our theory applies to any statistic where the observations are linearly combined and even to a Bayesian puzzle (Ikefuji et al., 2021).

Bates and Granger (1969) were the first to highlight the benefits of combining forecasts, and these benefits are now well known and well documented; see Timmermann (2006) for an exhaustive literature review and Elliott and Timmermann (2016) for a recent and detailed treatment. The M4 competition by Makridakis et al. (2018) and the more recent M5 competition by Makridakis et al. (2021) show that the overwhelming majority of the most accurate methods are combinations. Combinations are expected to become the default method in forecasting. Yet, the *uncertainty* of the combined forecast is not yet fully understood, especially the critical role of correlation.

The M4 competition was the first M competition that looked at prediction intervals in addition to point estimates. Their performance is investigated in detail in Grushka-Cockayne and Jose (2019), and they concluded that "the submissions fail to estimate the uncertainty properly." In the same paper, Grushka-Cockayne and Jose investigated six simple heuristics of combining intervals (and reference many more) which improve the situation, but they offered no theoretical explanation. Meira et al. (2021) explored the area in-depth and concluded that "no work has considered looking at the information delivered by prediction intervals when conducting model selection and/or combination."

Since then, Wang et al. (2021) investigated the weighted average of upper and lower bounds of forecasting intervals and found good performance on the M4 data. Hounyo and Lahiri (2021) showed that bootstrap inference needs to be adjusted when estimating the variance of a combined forecast. Knüppel and Krüger (2021) focused on uncertainty in density forecasts, using a random mean and variance, but they did not consider correlation between experts.[1] They found that the variance of the linear pool is larger than the average variance, which foreshadows our findings.

Our framework is as follows. We consider a sequence of correlated random variables $x = (x_1, x_2, \ldots, x_n)$ with common mean μ, so that

$$\mathrm{E}(x) = \mu\, \imath, \qquad \mathrm{var}(x) = \sigma^2 V, \tag{3}$$

where $\imath$ denotes the vector of ones and V is positive definite. Without loss of generality we normalize V by imposing the restriction $\mathrm{tr}(V) = n$, which implies that σ^2 is the *average* variance of the x_j.

The generalized least-squares estimator for μ is

$$\hat{\mu} = \frac{\imath' V^{-1} x}{\imath' V^{-1} \imath} = \omega' x, \qquad \omega = \frac{V^{-1}\imath}{\imath' V^{-1} \imath}, \tag{4}$$

with variance

$$\tau^2 = \mathrm{var}(\hat{\mu}) = \frac{\sigma^2}{\imath' V^{-1} \imath}, \tag{5}$$

where we note that $\imath'\omega = 1$, so that $\hat{\mu}$ is a "weighted average" of the x_j in the sense that the ω_j add to one. However, these "weights" are not proper weights as it is well known that, in the presence of correlation, the ω_j do not necessarily lie between zero and one; see e.g. Bates and Granger (1969) and Radchenko et al. (2021).

[1] Correlation is present only in the example in their Section 3, but it is assumed to be zero in their Monte Carlo simulations.

The corresponding estimator for σ^2 is

$$\hat{\sigma}^2 = \frac{(x-\hat{\mu}\imath)'V^{-1}(x-\hat{\mu}\imath)}{n} = \frac{(\imath'V^{-1}\imath)(x'V^{-1}x) - (\imath'V^{-1}x)^2}{n\,\imath'V^{-1}\imath}. \tag{6}$$

The estimator $\hat{\sigma}^2$ is bounded by $0 \leq \hat{\sigma}^2 \leq x'V^{-1}x/n$, so that one might think that $\hat{\sigma}^2$ is always finite. This is indeed true when the x_j are uncorrelated, but it is no longer true when the x_j are correlated. We shall prove and illustrate this simple fact below. The main message of the current paper is therefore that ignoring possible correlation in the observations can lead to estimates of σ^2 which are (much) too small.

Heuristically it is easy to see what happens by writing $V^{-1} = |V|^{-1}C_V$, where C_V denotes the cofactor matrix of V. In (4) the determinant $|V|$ cancels in numerator and denominator, but in (6) is doesn't and we are left with one factor $|V|$ in the denominator. This factor goes to zero, so that the ratio goes to ∞. This is the essence of our story. Notice the difference between σ^2 and τ^2. The parameter σ^2 is the variance of the process, while τ^2 measures the precision of the estimator $\hat{\mu}$, which goes to zero, because the determinant $|V|$ doesn't cancel in (5) and we are left with one factor $|V|$ in the numerator.

The remainder of this paper is organized as follows. In Section 2 we consider a general setup and show that $\hat{\sigma}^2$ in the most common case increases without bound when correlation gets closer and closer to one. In other words, we provide sufficient but not necessary conditions. In Section 3 we consider a more constrained setup, but within this setup we present a full characterization of the behavior of $\hat{\sigma}^2$ (and the other estimators of interest). In Section 4 we apply these results to understand why forecasts by Central Banks (in our case, the Bank of Japan) are so frequently misleadingly precise. Section 5 concludes. The appendix contains proofs of the three propositions.

2 General case: a sufficient condition

Let us denote the diagonal elements of V by $v_1^2, \ldots, v_n^2$, and introduce the notation $w_j = 1/v_j$ and $y_j = x_j/v_j$ for $j = 1, \ldots, n$. If we let V_0 denote the diagonal matrix containing the diagonal elements of V, then

$$v = V_0^{1/2}\imath, \qquad w = V_0^{-1/2}\imath, \qquad y = V_0^{-1/2}x. \tag{7}$$

The correlation matrix associated with the variance matrix V is defined as

$$P = V_0^{-1/2}VV_0^{-1/2}, \tag{8}$$

so that we can write our statistics of primary interest (4) and (6) as

$$\hat{\mu} = \frac{w'P^{-1}y}{w'P^{-1}w} \tag{9}$$

and

$$\hat{\sigma}^2 = \frac{(w'P^{-1}w)(y'P^{-1}y) - (w'P^{-1}y)^2}{n\, w'P^{-1}w}. \tag{10}$$

The variance of $\hat{\mu}$ is given by

$$\tau^2 = \frac{\sigma^2}{\imath' V^{-1}\imath} = \frac{\sigma^2}{w'P^{-1}w}, \tag{11}$$

and its estimator by

$$\hat{\tau}^2 = \frac{\hat{\sigma}^2}{\imath' V^{-1}\imath} = \frac{(w'P^{-1}w)(y'P^{-1}y) - (w'P^{-1}y)^2}{n\,(w'P^{-1}w)^2}. \tag{12}$$

Suppose now that the correlation matrix $P = P(r)$ is a function of some parameter r which we interpret as a correlation parameter. We wish to investigate the behavior of the (estimated) moments (9)–(12) as $r \to 1$ and we assume that P becomes singular in the limit, that is,

$$\lim_{r\to 1} |P(r)| = 0. \tag{13}$$

Let

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0 \tag{14}$$

be the eigenvalues of P, which will, in general, depend on r. Since P becomes singular in the limit, at least one of the eigenvalues (namely λ_n) will vanish as $r \to 1$. In practice, several eigenvalues will vanish in the limit, but what matters is not how many eigenvalues vanish, but how many vanish *at the same speed* as λ_n.

To appreciate this aspect we define $\xi_j = \lambda_n/\lambda_j$, so that

$$0 < \xi_1 \leq \xi_2 \leq \cdots \leq \xi_{n-1} \leq \xi_n = 1, \tag{15}$$

and let $1 \leq m \leq n-1$ be such that

$$\xi_m \to 0, \qquad \xi_{m+1} \to \xi^*_{m+1} > 0. \tag{16}$$

This tells us that the $n-m$ eigenvalues $\lambda_{m+1}, \ldots, \lambda_n$ go to zero at the same speed, while the remaining m eigenvalues don't go to zero or go to zero at a lower speed than λ_n. Notice that m must be strictly larger than 0 and strictly smaller than n. If $m = 0$ then all eigenvalues go to zero and this is impossible

because P is a correlation matrix and hence the sum of its eigenvalues is n. If $m = n$ then no eigenvalue goes to zero which contradicts our assumption that the determinant $|P|$ vanishes in the limit.

We decompose $P = S\Lambda S'$, where S is orthogonal and Λ is diagonal, and we partition $S = (S_1, S_2)$ and $\Lambda = \text{diag}(\Lambda_1, \Lambda_2)$ where Λ_1 contains the largest m eigenvalues and Λ_2 contains the remaining $n - m$ eigenvalues. Then,

$$\lambda_n P^{-1} = \lambda_n S\Lambda^{-1}S' = \lambda_n S_1\Lambda_1^{-1}S_1' + \lambda_n S_2\Lambda_2^{-1}S_2' = S_1\Xi_1 S_1' + S_2\Xi_2 S_2', \quad (17)$$

where $\Xi_1 = \text{diag}(\xi_1, \ldots, \xi_m)$ and $\Xi_2 = \text{diag}(\xi_{m+1}, \ldots, \xi_n)$.

Using this decomposition we write, for any two $n \times 1$ vectors a and b,

$$(\lambda_n/n)\, a'P^{-1}b = B_{ab} + C_{ab}, \quad (18)$$

where

$$B_{ab} = a'S_1\Xi_1 S_1' b/n, \qquad C_{ab} = a'S_2\Xi_2 S_2' b/n, \quad (19)$$

and this allows us to express our estimators of interest as

$$\hat{\mu} = \frac{B_{wy} + C_{wy}}{B_{ww} + C_{ww}} \quad (20)$$

and

$$\hat{\sigma}^2 = \frac{(B_{ww} + C_{ww})(B_{yy} + C_{yy}) - (B_{wy} + C_{wy})^2}{\lambda_n(B_{ww} + C_{ww})}, \quad (21)$$

the variance of $\hat{\mu}$ as

$$\tau^2 = \frac{\lambda_n}{B_{ww} + C_{ww}} \cdot \frac{\sigma^2}{n}, \quad (22)$$

and its estimator as

$$\hat{\tau}^2 = \frac{(B_{ww} + C_{ww})(B_{yy} + C_{yy}) - (B_{wy} + C_{wy})^2}{n\,(B_{ww} + C_{ww})^2}. \quad (23)$$

Now consider the limit as $r \to 1$. We already know that

$$\Xi_2 \to \Xi_2^* = \text{diag}(\xi_{m+1}^*, \ldots, \xi_n^*) > 0 \quad (24)$$

as $r \to 1$. Letting $C_{ab}^* = \lim_{r\to 1} C_{ab}$, we obtain the following result.

Proposition 1. *If* $C_{wy}^{*2} < C_{ww}^* C_{yy}^*$, *then*

$$\hat{\mu} \to \frac{C_{wy}^*}{C_{ww}^*}, \qquad \hat{\sigma}^2 \to \infty, \qquad \tau^2 \to 0, \qquad \hat{\tau}^2 \to \frac{C_{ww}^* C_{yy}^* - C_{wy}^{*2}}{n\, C_{ww}^{*2}} > 0,$$

as $r \to 1$.

The proposition tells us that — when a certain condition is satisfied — the estimator of σ^2 increases without bound when the correlation r converges to one. Let's consider this condition. We know from the Cauchy–Schwarz inequality that $0 \leq C_{wy}^{*2} \leq C_{ww}^{*} C_{yy}^{*}$. Our condition thus amounts to requiring that the two $m \times 1$ vectors $\Xi_2^{1/2} S_2' w$ and $\Xi_2^{1/2} S_2' y$ are linearly independent in the limit when $r \to 1$, and, since Ξ_2^{*} is positive definite, this occurs if and only if $S_2' w$ and $S_2' y$ are linearly independent in the limit. This situation is the most common and thus of most interest in practice.

3 Equicorrelation case: full characterization

In the previous section we obtained a (relatively weak) sufficient condition for $\hat{\sigma}^2 \to \infty$ when $r \to 1$ in a general setup. Now we consider a more specialized structure for the correlation matrix P and obtain a full characterization, that is, conditions that are not only sufficient but also necessary.

For this special structure we choose the case where the correlations are all equal so that the correlation matrix takes the form

$$P = \begin{pmatrix} 1 & r & r & \dots & r \\ r & 1 & r & \dots & r \\ \vdots & \vdots & \vdots & & \vdots \\ r & r & r & \dots & 1 \end{pmatrix}. \tag{25}$$

The matrix P in (25) is called the *equicorrelation* matrix, and it can be written as

$$P = \alpha A + \beta(I_n - A), \qquad A = \imath\imath'/n, \tag{26}$$

where A is an idempotent matrix of rank 1 and

$$\alpha = (n-1)r + 1, \qquad \beta = 1 - r. \tag{27}$$

Since A and $I - A$ are idempotent and $A(I - A) = 0$, the eigenvalues of P are α (once) and β ($n - 1$ times), and we see that P is positive definite if and only if

$$\frac{-1}{n-1} < r < 1. \tag{28}$$

The inverse and the determinant of P are given by

$$P^{-1} = \frac{1}{\alpha} A + \frac{1}{\beta}(I - A), \qquad |P| = \alpha \beta^{n-1}. \tag{29}$$

We are interested in the limiting case where $r \to 1$ and hence $|V|$ approaches zero. When r goes to one, $n - 1$ eigenvalues of P go to zero (all at the same

speed) and the remaining eigenvalue goes to n. Hence the number m, defined in (16), is here equal to $n-1$ with $\lambda_1 = \alpha$ and $\lambda_2 = \cdots = \lambda_n = \beta$.

The matrix C_{ab} introduced in (19) played a key part in Proposition 1, and it plays a key part again in our next result. It now takes the form

$$C_{ab} = \frac{a'(I-A)b}{n} = \frac{\sum_{j=1}^{n}(a_j - \bar{a})(b_j - \bar{b})}{n}, \tag{30}$$

and we see that it does not depend on r because the eigenvectors of P don't depend on r.

We obtain the following characterization.

Proposition 2. *Let $v = V_0^{1/2}\imath$, $w = V_0^{-1/2}\imath$, and $y = V_0^{-1/2}x$. Then, for fixed n, the limiting behavior as $r \to 1$ of $\hat{\mu}$, $\hat{\sigma}^2$, τ^2, and $\hat{\tau}^2$ depends on the specification of v and x, as follows:*

A:	$v = \imath$	$x = \bar{x}\imath$	$LC = Yes$	$\hat{\mu} \to \bar{x}$	$\hat{\sigma}^2 \to 0$	$\tau^2 \to \sigma^2$	$\hat{\tau}^2 \to 0$
B:	$v = \imath$	$x \neq \bar{x}\imath$	$LC = No$	$\hat{\mu} \to \bar{x}$	$\hat{\sigma}^2 \to \infty$	$\tau^2 \to \sigma^2$	$\hat{\tau}^2 \to \infty$
C:	$v \neq \imath$	$x = \bar{x}\imath$	$LC = Yes$	$\hat{\mu} \to \bar{x}$	$\hat{\sigma}^2 \to 0$	$\tau^2 \to 0$	$\hat{\tau}^2 \to 0$
D:	$v \neq \imath$	$x \neq \bar{x}\imath$	$LC = Yes$	$\hat{\mu} \to R_1$	$\hat{\sigma}^2 \to R_2$	$\tau^2 \to 0$	$\hat{\tau}^2 \to 0$
E:	$v \neq \imath$	$x \neq \bar{x}\imath$	$LC = No$	$\hat{\mu} \to R_1$	$\hat{\sigma}^2 \to \infty$	$\tau^2 \to 0$	$\hat{\tau}^2 \to R_3$

where

$$R_1 = \frac{C_{wy}}{C_{ww}}, \qquad R_2 = \frac{C_{xx}}{nC_{vv}}, \qquad R_3 = \frac{C_{ww}C_{yy} - C_{wy}^2}{nC_{ww}^2},$$

and LC stands for "x is a linear combination of $\imath$ and v."

There are thus five distinct cases depending on whether $v = \imath$ or not and $x = \bar{x}\imath$ or not. In *A*, *B*, and *C*, it follows from the restrictions on v and x whether or not x is a linear combination of $\imath$ and v. But when $v \neq \imath$ and $x \neq \bar{x}\imath$, then x may or may not be a linear combination of $\imath$ and v, and we need to distinguish between these two cases. Of the five cases, only *E* corresponds to the setup in Proposition 1.

The main message in the equicorrelation case confirms our analysis in the previous section, namely that $\hat{\sigma}^2 \to \infty$ when x is not a linear combination of $\imath$ and v. But it tells us, in addition, what happens when x *is* a linear combination of $\imath$ and v. Since Proposition 2 considers a special case while Proposition 1 considers the general case, it must be true that if $S_2'w$ and $S_2'y$ are linearly independent (the condition in Proposition 1), then x is not a linear combination of $\imath$ and v. To see this, assume that x is a linear combination of $\imath$ and v. Then, y is a linear combination of w and $\imath$, and hence $(I-A)y$ is a multiple of $(I-A)w$ (since $(I-A)\imath = 0$). This shows

that $S_2S_2'w$ and $S_2S_2'y$ are linearly dependent, and hence that $S_2'w$ and $S_2'y$ are linearly dependent.

Of interest is also the different behavior of τ^2 and $\hat{\tau}^2$. While $\hat{\tau}^2$ converges to a positive constant in the standard case E, $\tau^2 \to 0$ in that case, in accordance with Proposition 1.

We illustrate the behavior of our four statistics by considering examples for the cases B–E, leaving out the somewhat trivial case A.

Case	v	x
B	$(1,1,1,1,1)'$	$(10,30,11,24,36)'$
C	$(1,2,2,1.5,2.5)'/\sqrt{3.5}$	$10\imath$
D	$(1,2,2,1.5,2.5)'/\sqrt{3.5}$	$10\imath + 20v$
E	$(1,2,2,1.5,2.5)'/\sqrt{3.5}$	$(10,30,11,24,36)'$

Table 1: Examples in the equicorrelation case for $n = 5$

The four examples are listed in Table 1, where we set $n = 5$. Notice that $v'v = 5$ in all cases as required by the restriction $\text{tr}(V) = n$.

FIGURE 1 HERE

The behavior of the four statistics in these four cases is graphed in Figure 1. From Proposition 2 we know what happens in the limit when $r \to 1$. In particular, we see that $\hat{\sigma}^2$ goes to ∞ in Examples B and E, and that $\hat{\tau}^2$ goes to ∞ in B and to zero in C and D. The proposition does not, however, tell us the path towards the limit. In the typical case E, we see that $\hat{\mu} = \bar{x}$ at $r = 0.5$ and is monotonic in r, that $\hat{\sigma}^2$ is also monotonic and increases sharply after $r = 0.7$ or $r = 0.8$. More specifically we have $\hat{\sigma} = 11.2$ when $r = 0.0$, $\hat{\sigma} = 16.8$ when $r = 0.7$, $\hat{\sigma} = 19.5$ when $r = 0.8$, and $\hat{\sigma} = 25.5$ when $r = 0.9$. So, even for moderately positive correlation (say $r = 0.8$) the standard deviation is about twice what is was without correlation. We also see that $\hat{\tau}^2$ is concave with a maximum at $r = 0.88$. The concavity is induced by $\imath'V^{-1}\imath$. The $\hat{\tau}^2$ curves all originate from the same point $r = -0.25$, while the $\hat{\sigma}^2$ curves don't. This, again, is due to the term $\imath'V^{-1}\imath$ term in the denominator.

The special case $n = 2$ differs from the case $n \geq 3$ because of the fact that the matrix $(\imath, v)$ is now nonsingular if and only if $v_1 \neq v_2$. Hence, x will be a linear combination of $\imath$ and v if and only if either $v_1 \neq v_2$ or $x_1 = x_2$ (or both), and this implies that E cannot occur. We obtain the following proposition as a special case of Proposition 2.

Proposition 3. *For $n = 2$ and $r \to 1$, we have the following limiting behavior of $\hat{\mu}$, $\hat{\sigma}^2$, τ^2, and $\hat{\tau}^2$:*

A:	$v_1 = v_2$	$x_1 = x_2$	$\hat{\mu} \to \bar{x}$	$\hat{\sigma}^2 \to 0$	$\tau^2 \to \sigma^2$	$\hat{\tau}^2 \to 0$
B:	$v_1 = v_2$	$x_1 \neq x_2$	$\hat{\mu} \to \bar{x}$	$\hat{\sigma}^2 \to \infty$	$\tau^2 \to \sigma^2$	$\hat{\tau}^2 \to \infty$
C:	$v_1 \neq v_2$	$x_1 = x_2$	$\hat{\mu} \to \bar{x}$	$\hat{\sigma}^2 \to 0$	$\tau^2 \to 0$	$\hat{\tau}^2 \to 0$
D:	$v_1 \neq v_2$	$x_1 \neq x_2$	$\hat{\mu} \to R_1$	$\hat{\sigma}^2 \to R_2$	$\tau^2 \to 0$	$\hat{\tau}^2 \to 0$

where

$$R_1 = \frac{v_2 x_1 - v_1 x_2}{v_2 - v_1}, \qquad R_2 = \frac{(x_2 - x_1)^2}{2(v_2 - v_1)^2}.$$

When $n = 2$ there are only four distinct cases, because E cannot occur as we have seen. While E was the typical case when $n \geq 3$, the typical case when $n = 2$ is D, and this affects the limiting behavior of both $\hat{\sigma}^2$ and $\hat{\tau}^2$. The typical situation when $n \geq 3$ is that $\hat{\sigma}^2 \to \infty$ and $\hat{\tau}^2$ converges to a finite positive limit, while the typical situation when $n = 2$ is that $\hat{\sigma}^2$ converges to a finite limit and $\hat{\tau}^2 \to 0$. Notice also that R_1 never lies in-between x_1 and x_2, so that, in the typical case D, $\hat{\mu}$ does not lie in-between x_1 and x_2 for large enough r.

Case	v	x
B	$(1, 1)'$	$(10, 30)'$
D	$(1, 2)'/\sqrt{2.5}$	$(10, 30)'$

Table 2: Examples in the equicorrelation case for $n = 2$

We provide two examples for the case $n = 2$. In Example B we have $v_1 = v_2$, and in D we have $v_1 \neq v_2$.

FIGURE 2 HERE

The typical case is now D, where $\hat{\sigma}^2$ remains finite when $r \to 1$, as illustrated in Figure 2. At $r = 0$ we have $\hat{\sigma} = 15.0$, at $r = 0.8$ we have $\hat{\sigma} = 25.0$, and at $r = 1.0$ we have $\hat{\sigma} = 33.5$. So, the value of $\hat{\sigma}$ more than doubles when r increases from 0 to 1. The shape of all curves is similar to the case $n = 5$. However, $\hat{\mu}$ is negative for $r \geq 0.8$.

4 The Bank of Japan's GDP and CPI forecasts

We illustrate the theory developed in the previous sections by analyzing GDP and CPI forecasts of the Bank of Japan. The Bank of Japan (BOJ)

has a Policy Board consisting of nine members: the governor, two deputy governors, and six members. Each of the nine members provide their view, and these are aggregated into the official BOJ's forecast. The forecasts are published in January, April, July, and October of each year in the Outlook Report (formally, the Outlook for Economic Activity and Prices) soon after the monetary policy meetings and they provide the real GDP and CPI (all items less fresh food) forecasts for the next three years. The Outlook Report presents the Bank's outlook for developments in economic activity and prices, assesses upside and downside risks, and outlines its views on the future course of monetary policy. It is published online in a short version ("the Bank's view") and a long version ("full text").

The published forecasts contain the aggregate forecast, but also the forecasts by the individual members of the Policy Board. The individual members' forecasts are presented graphically, and numerical values are provided only for the median forecast and the range. If we order the forecasts $f_1 \geq f_2 \geq \cdots \geq f_9$, then the range is defined as $|f_8 - f_2|$, hence the trimmed spread after removing the highest and the lowest forecast. The range is interpreted by the BOJ as a simple measure of disagreement among the Policy Board members about future economic development. Unfortunately, the range has a coverage rate (inclusion rate) of only 10–20% for GDP and 10–35% for inflation; see also Tsuchiya (2021, Table 1). The BOJ is aware of this and provides a warning that "the range does not indicate the forecast errors" as a footnote under every table of forecasts that they publish.

Still, the coverage rate is alarmingly low and our theory provides an instrument to obtain more realistic coverage rates. To apply our theory, we shall assume that the forecasts of the members of the Policy Board are positively correlated. This seems realistic, because the members know each other, they talk together, they use the same or similar datasets, and they received similar training in economics and statistics. What we don't know is the degree of correlation.

The individual members' forecasts are not provided numerically, only graphically. We have to extract the numerical values from the graphs, and this is possible from October 2015 onwards (with the exception of April 2020, where the graph does not depict the individual forecasts).[2] The resolution in the published graphs is sufficiently high that we are confident that our extracted forecasts are accurate up to one decimal place. Thus, we obtain nine ordered individual forecasts for each of the Bank's GDP or CPI fore-

[2]The extracted values, which constitute our dataset, are contained in our Data Document which is available upon request. This document contains not only the extracted values, but also an account of how we dealt with several consistency issues.

casts. We cannot link a specific forecast to a specific member of the Policy Board, and therefore we cannot follow the members through time or compute correlations between the members' forecasts or their variances.

The data collected from the graphs in each report were checked against the numerical data in the tables. Some care is required before we can use the extracted data for analysis. This applies in particular to inflation, where the members are asked to forecast the CPI (all items less fresh food). In several reports (for example, April 2018), the data on the graphs exclude the effect of the consumption tax hike. However, all members agree that this effect is 0.5 percentage points in 2019 and 2020 (see April 2018, footnote 4, page 8), and we therefore added 0.5 to the forecast made by the Policy Board members to produce the final CPI (all items less fresh food) forecasts for 2019 and 2020. Several other adjustments are required so that forecasts from different reports and the actual realizations of the CPI (all items less fresh food) can be compared. All adjustments are explained and justified in our Data Document.

FIGURE 3 HERE

We assume equicorrelation as analyzed in Section 3, and use four correlations for our Policy Board members: $r = 0.0$ (as a benchmark), $r = 0.70$ and 0.95 (as realistic values), and 0.99 (as an extreme value).

We first consider the evolution of forecasts made for the same year, where the forecast is made at twelve different points in time, each one quarter closer to the realization. For the year 2018 we consider forecasts made from 2016/Q2 up to 2019/Q1 (twelve quarters), given that the "true" realized value is determined in 2019/Q2. For 2019 and 2020 we also consider twelve quarters starting four and eight quarters later. In Figure 3 the left panels contain the forecasts for GDP, the right panels for CPI. The members' forecasts are represented by small crosses. (We often see fewer than nine crosses in the figures because the forecasts are often numerically close or the same, so that crosses are on top of each other.) There are two solid lines: a red horizontal line (with small circles) representing the realized values of the forecast, and a blue line (not horizontal) representing the median of the forecasts. The two dotted lines in blue denote the range, and we see immediately that the range typically does not include the realized value. The other dotted lines in black denote confidence intervals based on different values of r. The confidence interval when $r = 0$ is typically wider than the range, which already indicates that the range is too tight. When r takes on more realistic values, then the confidence interval becomes wider, and it is clear that only with correlations as high as $r = 0.95$ or $r = 0.99$ do we obtain confidence intervals that include the actual realization. So, it is in fact possible to obtain realistic confidence

intervals, but only if we assume that the correlation between the individual forecasts is of the order 0.95 or higher.

				Equicorrelation r			
	Year	n	BOJ	0.00	0.70	0.95	0.99
GDP	2018	12	0.00	0.00	0.33	0.75	1.00
	2019	12	0.00	0.00	0.00	0.08	0.83
	2020	11	0.09	0.09	0.18	0.27	0.27
CPI	2018	12	0.25	0.67	0.92	1.00	1.00
	2019	12	0.17	0.17	0.50	1.00	1.00
	2020	11	0.09	0.18	0.27	0.55	1.00

Table 3: Coverage rates associated with Figure 3

The coverage rate of the intervals is summarized in Table 3. There are data on $n = 12$ quarters in 2018 and 2019, but only on $n = 11$ quarters in 2020, because April 2020 is missing. For the forecasts of GDP in 2018, we need to assume that the correlation is between 0.95 and 0.99 to achieve the nominal coverage of 95%. For 2019, even a correlation of 0.99 delivers a coverage of only 83%. The numbers for 2020 need to be treated with caution because of the COVID pandemic. Once the impact of COVID is included in the forecasts from July 2020, then the intervals for $r = 0.95$ cover the actual values. Similarly, for the forecasts of CPI in 2018 and 2019, a correlation between 0.70 and 0.95 is required to achieve the nominal coverage. CPI in 2020 has experienced the effect of COVID as well, but a correlation between 0.95 and 0.99 still delivers an interval with nominal coverage.

FIGURE 4 HERE

Next, we consider forecasts for different years, but using the same forecast horizon. We consider forecasts made in April two years before the actual release (forecast horizon $h = 11$ quarters), forecasts from July of the previous year ($h = 6$ quarters), and forecasts made in October of the same year ($h = 1$ quarter). Figure 4 presents this situation for both GDP (left panels) and CPI (right panels) for the three forecast horizons, respectively. As in Figure 3, the member forecasts are presented in crosses and the actual realization and the median forecast as solid lines. The dotted lines closest to the median represent the BOJ range, while all other dotted lines represent the confidence intervals under different correlation assumptions as indicated in the figure. Again, a high correlation is required to cover the actual realizations using the long term ($h = 11$) and medium-term ($h = 6$) horizons. This is even true

for the short term ($h = 1$) when forecasting GDP, but not when forecasting CPI.

	Horizon			Equicorrelation r			
	(quarters)	n	BOJ	0.00	0.70	0.95	0.99
GDP	$h = 11$	3	0.00	0.00	0.33	0.33	0.67
	$h = 6$	4	0.00	0.25	0.25	0.50	0.75
	$h = 1$	4	0.25	0.25	0.50	0.50	1.00
CPI	$h = 11$	3	0.00	0.33	0.67	0.67	1.00
	$h = 6$	4	0.25	0.50	0.50	1.00	1.00
	$h = 1$	4	0.50	1.00	1.00	1.00	1.00

Table 4: Coverage rates associated with Figure 4

The corresponding coverage rates are given in Table 4. For the long- and medium-term forecasts of GDP, the COVID impact is severe, but a correlation up to 0.99 is sufficient to achieve nominal coverage if we exclude 2020, as can be seen on the graphs. The same conclusion holds for the short-term GDP forecasts even when 2020 is included, since the impact was well understood in October 2020 when the forecasts were made. For CPI, a correlation between 0.95 and 0.99 is required to achieve nominal coverage for the long-term forecasts and between 0.70 and 0.95 for the medium-term forecasts. The short-term intervals with $r = 0$ cover all realizations, although the forecasts themselves are very close.

5 Concluding remarks

We have considered a sequence $x = (x_1, x_2, \ldots, x_n)$ of correlated random variables with common mean μ and variance $\sigma^2 V$, where V is positive definite and normalized by the restriction $\mathrm{tr}(V) = n$. In this framework, the generalized least-squares estimator for μ is

$$\hat{\mu} = \frac{\imath' V^{-1} x}{\imath' V^{-1} \imath},$$

where $\imath$ denotes the vector of ones, and the corresponding estimator for σ^2 is

$$\hat{\sigma}^2 = \frac{(\imath' V^{-1} \imath)(x' V^{-1} x) - (\imath' V^{-1} x)^2}{n\, \imath' V^{-1} \imath}.$$

If there is no correlation, then $\text{var}(x) = \sigma^2 I_n$ and the parameters μ and σ^2 are estimated by

$$\hat{\mu} = \bar{x}, \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2.$$

The main point of the paper is to demonstrate that the no-correlation assumption if often unrealistic *and that it matters*. In particular, when correlation is positive and converges to 1, the estimator $\hat{\sigma}^2$ diverges to infinity.

The lesson for practitioners is that ignoring positive correlation in the data can be dangerous, because it leads to confidence intervals that are (much) too narrow. This phenomenon is quite general, and it is clearly visible in the precision estimates of the GDP and especially the CPI forecasts of Central Banks. We have illustrated our theory using forecasts from the Bank of Japan, because they provide information about the data underlying the Bank's forecasts. Macro variables such as GDP and CPI are difficult to forecast, but for policy applications it is important to provide confidence bands with credible coverage probabilities. To achieve this we must assume a high level of positive correlation between the individual forecasters.

The analysis in this paper also helps to explain a Bayesian puzzle (Ikefuji at al., 2021), where we have data (n observations) and a prior, both normally distributed, which together lead to a posterior, also normal. According to Bayesian theory, the variance of the posterior is smaller than the variance in the data and also smaller than the variance in the prior. But, in practice, when we look at the posterior (a published report) and the data (some weighted average of the observations), we often see that the variance in the data is smaller than the variance in the posterior, which contradicts the theory. In other words, the data variance is too small and this can be understood and remedied by allowing for correlation among the n observations.

Acknowledgements

We are grateful to Masako Ikefuji and Nao Sudo for helpful comments.

Appendix: Proofs

Proof of Proposition 1

The assumptions imply that C_{ww}, C_{yy}, and $C_{ww}C_{yy} - C_{wy}^2$ remain strictly positive in the limit, while B_{ww}, B_{yy}, and B_{wy} go to zero. The result then follows from the fact that $\lambda_n \to 0$.

Proof of Proposition 2

Let P be the correlation matrix defined in (25). Then, for arbitrary vectors a and b,

$$a'P^{-1}b = \frac{a'Q_0b - \beta a'Q_1b}{\alpha\beta},$$

where

$$Q_0 = n(I - A), \qquad Q_1 = Q_0 - I.$$

This gives

$$\hat{\mu} = \frac{w'P^{-1}y}{w'P^{-1}w} = \frac{w'Q_0y - \beta w'Q_1y}{w'Q_0w - \beta w'Q_1w}, \tag{31}$$

$$\begin{aligned} n\hat{\sigma}^2 &= (y - \hat{\mu}w)'P^{-1}(y - \hat{\mu}w) = y'P^{-1}(y - \hat{\mu}w) \\ &= \frac{y'P^{-1}yw'Q_0w - \beta y'P^{-1}yw'Q_1w - y'P^{-1}ww'Q_0y + \beta y'P^{-1}ww'Q_1y}{w'Q_0w - \beta w'Q_1w} \\ &= \frac{S_0 + \beta S_1 + \beta^2 S_2}{\beta(T_0 + \beta T_1 + \beta^2 T_2)}, \end{aligned} \tag{32}$$

and, using the fact that $\alpha = n - (n-1)\beta$,

$$\hat{\tau}^2 = \frac{S_0 + \beta S_1 + \beta^2 S_2}{T_0 + \beta T_1 + \beta^2 T_2} \cdot \frac{1 - \beta(n-1)/n}{w'Q_0w - \beta w'Q_1w}, \tag{33}$$

where

$$\begin{aligned} S_0 &= (w'Q_0w)(y'Q_0y) - (w'Q_0y)^2, \\ S_1 &= -(w'Q_0w)(y'Q_1y) - (w'Q_1w)(y'Q_0y) + 2(w'Q_0y)(w'Q_1y), \\ S_2 &= (w'Q_1w)(y'Q_1y) - (w'Q_1y)^2, \end{aligned}$$

and

$$\begin{aligned} T_0 &= nw'Q_0w, \\ T_1 &= -(n-1)w'Q_0w - nw'Q_1w, \\ T_2 &= (n-1)w'Q_1w. \end{aligned}$$

Before discussing separate cases, we make two preliminary observations. First, since Q_0 is positive semidefinite, we have $w'Q_0w \geq 0$ with

$$w'Q_0w = 0 \iff (I - A)w = 0 \iff w = \bar{w}\imath \iff v = \imath, \tag{34}$$

because the components of $w = V_0^{-1/2}\imath$ are all equal if and only if all diagonal elements of V_0 are all equal, which occurs if and only if all these diagonal

elements are one, since $\mathrm{tr}(V_0) = \mathrm{tr}(V) = n$ by assumption. Note that $v = \imath$ implies $\bar{w} = 1$.

Second, we know from the Cauchy–Schwarz inequality that $S_0 \geq 0$ with equality if and only if the matrix

$$(I - A)(w, y) = (I - A)V_0^{-1/2}(\imath, x)$$

has rank < 2. This occurs if and only if either $(I - A)w = 0$ or $(I - A)w \neq 0$ and $(I - A)y = \theta_1(I - A)w$ for some θ_1 (possibly zero). The latter condition occurs if and only if $y - \theta_1 w = \theta_2 \imath$ for some θ_2 (possibly zero), that is, if and only if $x = \theta_1 \imath + \theta_2 v$. Hence,

$$S_0 = 0 \iff v = \imath \text{ or } x \text{ is a linear combination of } \imath \text{ and } v. \tag{35}$$

The case of $\hat{\mu}$

The limiting behavior of $\hat{\mu}$ depends on whether $w'Q_0w$ is zero or not. We know from (34) that $w'Q_0w = 0$ if and only if $v = \imath$. If $v = \imath$, then

$$\hat{\mu} = \frac{w'Q_1y}{w'Q_1w} = \frac{w'y}{w'w} = \frac{\imath'x}{\imath'\imath} = \bar{x}.$$

If $v \neq \imath$, then

$$\hat{\mu} = \frac{w'Q_0y - \beta w'Q_1y}{w'Q_0w - \beta w'Q_1w} \to \frac{w'Q_0y}{w'Q_0w} = \frac{C_{wy}}{C_{ww}}.$$

The case of $\hat{\sigma}^2$

The limiting behavior of $\hat{\sigma}^2$ depends on whether S_0 is zero or not, and we know from (35) that $S_0 = 0$ if and only if either $v = \imath$ or x is a linear combination of $\imath$ and v.

Case 1: $v = \imath$. If $v = \imath$ then $(I - A)w = 0$, and hence $w = \imath$, $Q_0w = 0$, and $Q_1w = -\imath$. This gives $S_0 = T_0 = 0$ and

$$\begin{aligned} S_1 &= -(w'Q_1w)(y'Q_0y) = n\, x'Q_0x, \\ T_1 &= -nw'Q_1w = n^2 > 0. \end{aligned}$$

As a result,

$$n\hat{\sigma}^2 = \frac{S_1 + \beta S_2}{\beta T_1 + \beta^2 T_2},$$

so that we need to distinguish between $S_1 = 0$ and $S_1 \neq 0$, that is between $x = \bar{x}\imath$ and $x \neq \bar{x}\imath$. If $v = \imath$ and $x = \bar{x}\imath$, then $S_1 = S_2 = 0$ and $\hat{\sigma}^2 = 0$. If $v = \imath$ and $x \neq \bar{x}\imath$, then $S_1 > 0$ and $\hat{\sigma}^2 \to \infty$.

Case 2: $v \neq \imath$. Here we distinguish between two subcases. If $v \neq \imath$ and x is not a linear combination of $\imath$ and v, then by (35), $S_0 > 0$ and $T_0 > 0$, so that $\hat{\sigma}^2 \to \infty$.

If $v \neq \imath$ and x is a linear combination of $\imath$ and v, then $S_0 = 0$ and $T_0 > 0$, and hence

$$n\hat{\sigma}^2 = \frac{S_1 + \beta S_2}{T_0 + \beta T_1 + \beta^2 T_2} \to \frac{S_1}{T_0}.$$

Write $x = \theta_1 \imath + \theta_2 v$. Then $(I - A)x = \theta_2(I - A)v$ and hence $\theta_2^2 = C_{xx}/C_{vv}$. Premultiplying by $V_0^{-1/2}$ gives $y = \theta_1 w + \theta_2 \imath$, and hence

$$Q_0 y = \theta_1 Q_0 w, \qquad Q_1 y = \theta_1 Q_1 w - \theta_2 \imath.$$

This gives

$$w'Q_0 y = \theta_1 w'Q_0 w, \qquad y'Q_0 y = \theta_1^2 w'Q_0 w,$$

and

$$w'Q_1 y = \theta_1 w'Q_1 w - n\theta_2 \bar{w}, \qquad y'Q_1 y = \theta_1^2 w'Q_1 w - 2n\theta_1\theta_2\bar{w} - n\theta_2^2,$$

so that

$$S_1 = -(w'Q_0 w)\left(\theta_1^2 w'Q_1 w - 2\theta_1 w'Q_1 y + y'Q_1 y\right) = n\theta_2^2 w'Q_0 w$$

and $S_1/T_0 = \theta_2^2 = C_{xx}/C_{vv}$.

The case of τ^2

We have

$$\tau^2 = \frac{\sigma^2}{w'P^{-1}w} = \frac{\alpha\beta\sigma^2}{w'Q_0 w - \beta w'Q_1 w}.$$

If $v = \imath$ then $w'Q_0 w = 0$ and $w'Q_1 w = -n$, and hence $\tau^2 \to \sigma^2$. If $v \neq \imath$ then $w'Q_0 w > 0$ and hence $\tau^2 \to 0$.

The case of $\hat{\tau}^2$

We have

$$n\hat{\tau}^2 = \frac{n\hat{\sigma}^2}{w'P^{-1}w} = \frac{S_0 + \beta S_1 + \beta^2 S_2}{T_0 + \beta T_1 + \beta^2 T_2} \cdot \frac{n - \beta(n-1)}{w'Q_0 w - \beta w'Q_1 w}.$$

Case 1: $v = \imath$. As in the case of $\hat{\sigma}^2$ we have $S_0 = T_0 = 0$. As a result,

$$n\hat{\tau}^2 = \frac{S_1 + \beta S_2}{\beta T_1 + \beta^2 T_2} \cdot \frac{n - \beta(n-1)}{n},$$

so that we again need to distinguish between $S_1 = 0$ and $S_1 \neq 0$, that is between $x = \bar{x}\imath$ and $x \neq \bar{x}\imath$. If $v = \imath$ and $x = \bar{x}\imath$, then $S_1 = S_2 = 0$ and $\hat{\tau}^2 = 0$. If $v = \imath$ and $x \neq \bar{x}\imath$, then, $S_1 = n\, x'Q_0x > 0$ and hence $\hat{\tau}^2 \to \infty$.

Case 2: $v \neq \imath$. Then $T_0 = nw'Q_0w > 0$, so that

$$n\hat{\tau}^2 \to \frac{S_0}{T_0} \cdot \frac{n}{w'Q_0w} = \frac{C_{ww}C_{yy} - C_{wy}^2}{C_{ww}^2},$$

where we note that $C_{ww}C_{yy} - C_{wy}^2 = 0$ when $x = \bar{x}\imath$. This concludes the proof.

Proof of Proposition 3

This is the special case of Proposition 2 where $n = 2$. In this special case, x is not a linear combination of $\imath$ and v if and only if $v_1 = v_2$ and $x_1 \neq x_2$. Also,

$$C_{ab} = \frac{\sum_{j=1}^{2}(a_j - \bar{a})(b_j - \bar{b})}{2} = \frac{(a_1 - a_2)(b_1 - b_2)}{4}.$$

Hence, for $v_1 \neq v_2$,

$$\frac{C_{wy}}{C_{ww}} = \frac{y_1 - y_2}{w_1 - w_2} = \frac{x_1/v_1 - x_2/v_2}{1/v_1 - 1/v_2} = \frac{v_2x_1 - v_1x_2}{v_2 - v_1},$$

and

$$\frac{C_{xx}}{C_{vv}} = \frac{(x_1 - x_2)^2}{(v_1 - v_2)^2}.$$

Finally, since $C_{ww}C_{yy} - C_{wy}^2 = 0$, the results follow.

References

Bank of Japan (quarterly). *Outlook for Economic Activity and Prices.* https://www.boj.or.jp/en/mopo/outlook/index.htm.

Bates, J. M. and C. W. J. Granger (1969). The combination of forecasts. *Operational Research Quarterly*, 20, 451–468.

Elliott, G. and A. Timmermann (2016). Forecast combinations. In G. Elliott and A. Timmermann (Eds), *Economic Forecasting*, Chapter 14, pp. 310–344. Princeton University Press.

Grushka-Cockayne, Y. and V. R. R. Jose (2019). Combining prediction intervals in the M4 competition. *International Journal of Forecasting*, 36, 178–185.

Hounyo, U. and K. Lahiri (2021). Estimating the variance of a combined forecast: Bootstrap-based approach. CREATES Research Paper Nr. 2021-14.

Ikefuji, M., J. R. Magnus, and T. Yamagata (2021). Revealing priors from posteriors. Submitted for publication.

Knüppel, M. and F. Krüger (2021). Forecast uncertainty, disagreement, and the linear pool. *Journal of Applied Econometrics*, forthcoming. doi:10.1002/jae.2834.

Makridakis, S., E. Spiliotis, and V. Assimakopoulos (2018). The M4 competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, 34, 802–808.

Makridakis, S., E. Spiliotis, and V. Assimakopoulos (2021). The M5 accuracy competition: Results, findings and conclusions. *International Journal of Forecasting*, forthcoming.

Meira, E., F. L. C. Oliveira, and J. Jeon (2021). Treating and pruning: New approaches to forecasting model selection and combination using prediction intervals. *International Journal of Forecasting*, 37, 547–568.

Radchenko, P., A. L. Vasnev, and W. Wang (2021). Too similar to combine? On negative weights in forecast combination. *International Journal of Forecasting*, forthcoming. doi:10.1016/j.ijforecast.2021.08.002.

Timmermann, A. (2006). Forecast combinations. In G. Elliott, C. Granger, and A. Timmermann (Eds), *Handbook of Economic Forecasting*, Vol. 1, pp. 135–196. North-Holland.

Tsuchiya, Y. (2021). The value added of the Bank of Japan's range forecasts. *Journal of Forecasting*, 40, 817–833.

Wang, X., Y. Kang, F. Petropoulos, and F. Li (2021). The uncertainty estimation of feature-based forecast combinations. *Journal of the Operational Research Society*, forthcoming. doi:10.1080/01605682.2021.1880297.

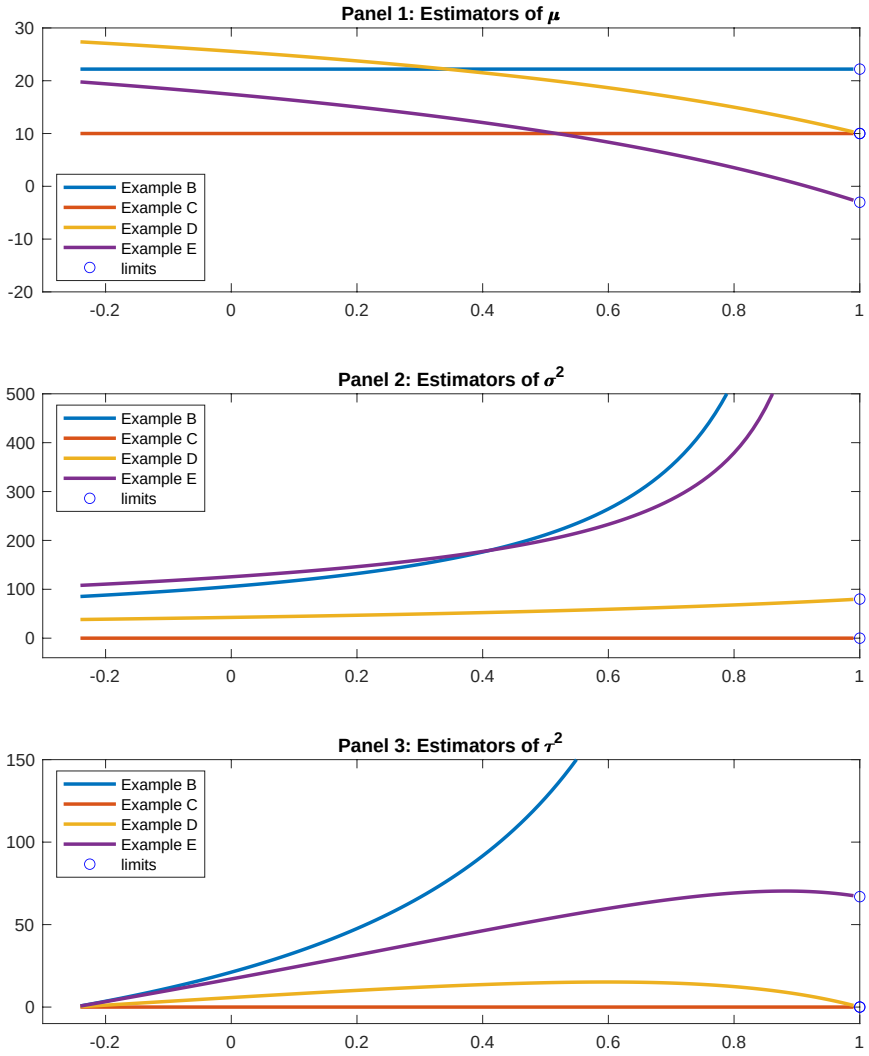

Figure 1: Behavior of $\hat{\mu}$ (Panel 1), $\hat{\sigma}^2$ (Panel 2), and $\hat{\tau}^2$ (Panel 3) as a function of r when $n = 5$

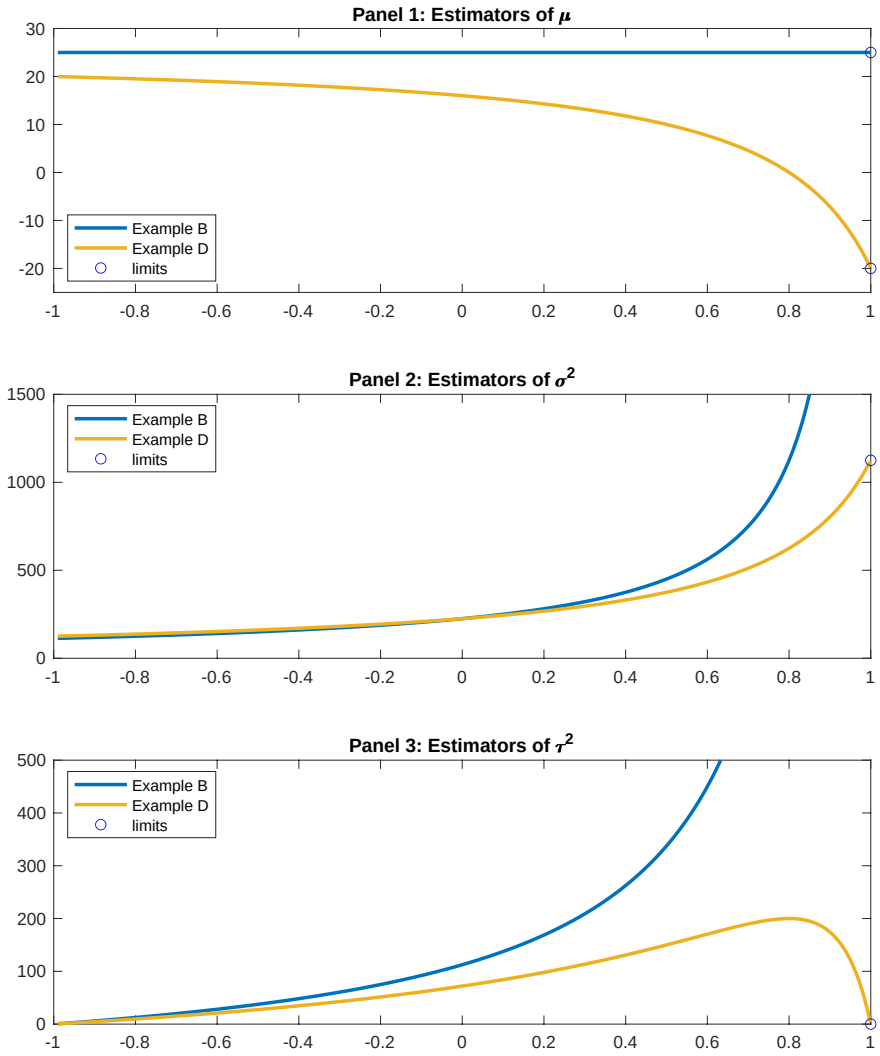

Figure 2: Behavior of $\hat{\mu}$ (Panel 1), $\hat{\sigma}^2$ (Panel 2), and $\hat{\tau}^2$ (Panel 3) as a function of r when $n = 2$

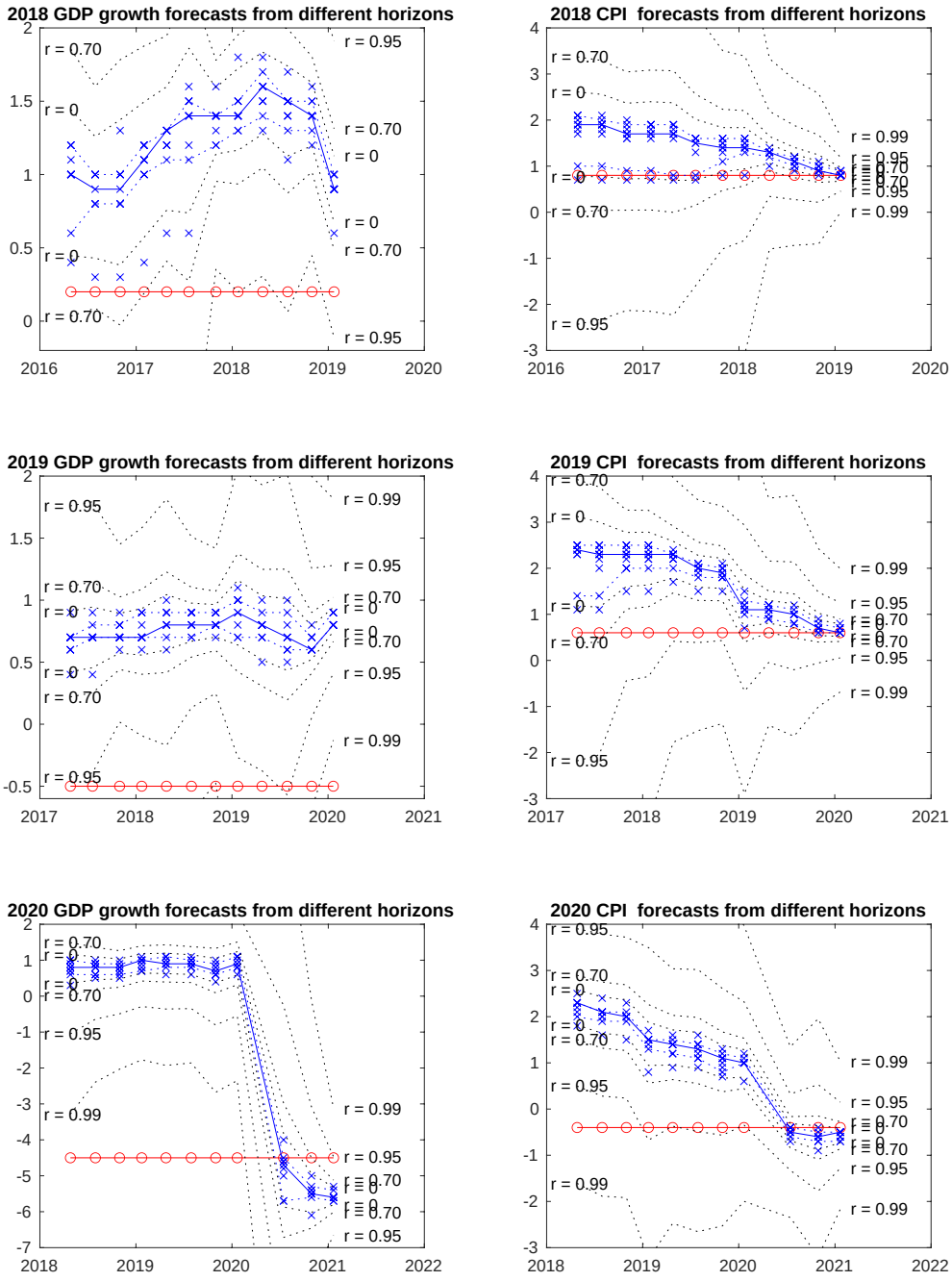

Figure 3: GDP and CPI forecasts with decreasing forecast horizons

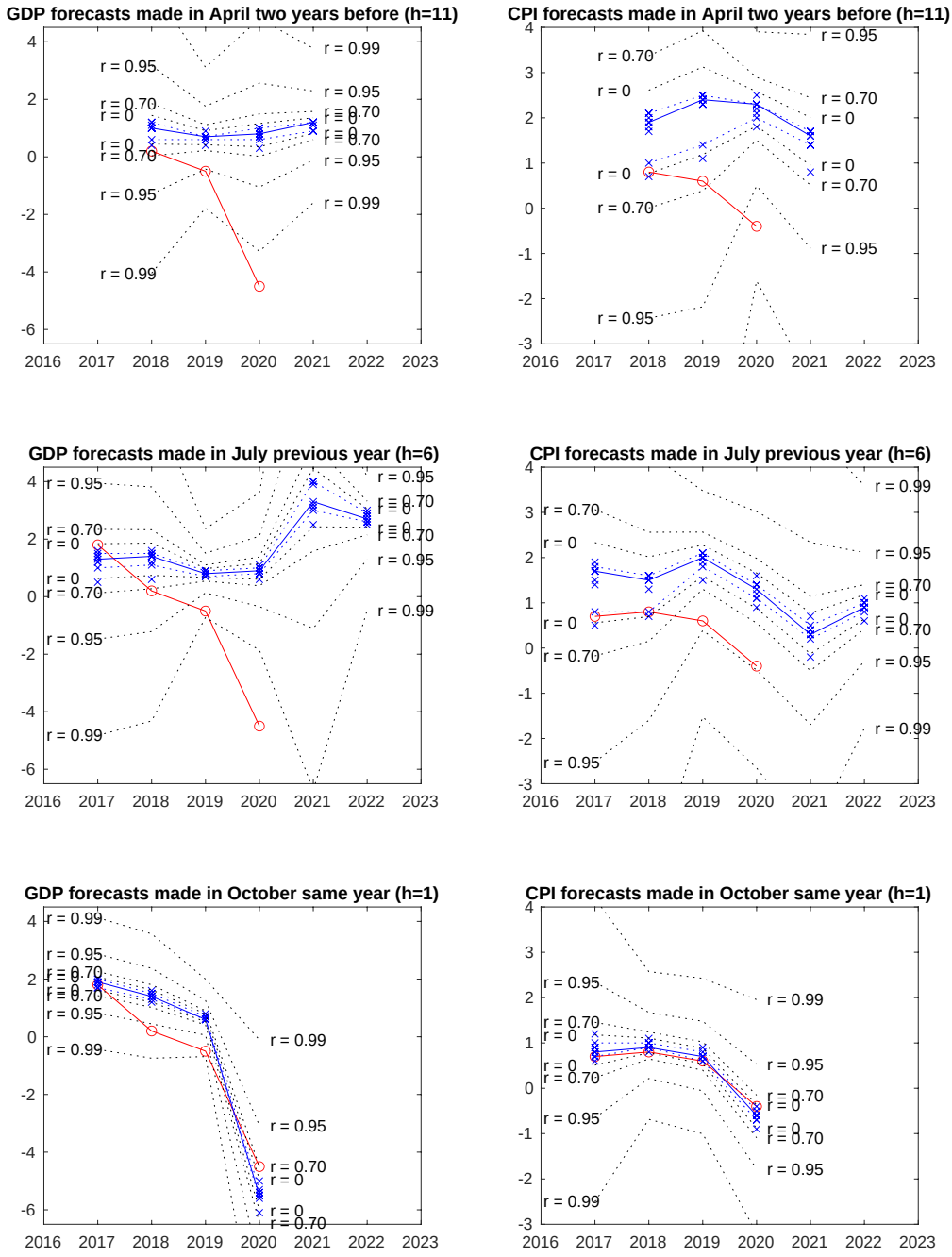

Figure 4: GDP and CPI forecasts with the same forecast horizon

On the uncertainty of a combined forecast: The critical role of correlation

Data Document (for online publication only)

December 21, 2021

Jan R. Magnus
Vrije Universiteit, Amsterdam, The Netherlands
and Tinbergen Institute, Amsterdam, The Netherlands

Andrey L. Vasnev
University of Sydney, New South Wales, Australia

The Bank of Japan (BOJ) has a Policy Board consisting of nine members: the governor, two deputy governors, and six members. Each of the nine members provide their view, and these are aggregated into the official BOJ's forecast. The forecasts are published in January, April, July, and October of each year in the *Outlook Report* (formally, the Outlook for Economic Activity and Prices) soon after the monetary policy meetings. The Outlook Reports are available from

https://www.boj.or.jp/en/mopo/outlook/index.htm.

and they provide the real GDP and CPI (all items less fresh food) forecasts for the next three years. The Outlook Report presents the Bank's outlook for developments in economic activity and prices, assesses upside and downside risks, and outlines its views on the future course of monetary policy. It is published online in a short version ("the Bank's view") and a long version ("full text").

The published forecasts contain the aggregate forecast, but also the forecasts by the individual members of the Policy Board. The individual members' forecasts are presented graphically, and numerical values are provided only for the median forecast and the range. If we order the forecasts $f_1 \geq f_2 \geq \cdots \geq f_9$, then the range is defined as $|f_8 - f_2|$, hence the trimmed spread after removing the highest and the lowest forecast. The range is interpreted by the BOJ as a simple measure of disagreement among the Policy Board members about future economic development. Unfortunately, the range has

a coverage rate (inclusion rate) of only 10–20% for GDP and 10–35% for inflation. The BOJ is aware of this and provides a warning that "the range does not indicate the forecast errors" as a footnote under every table of forecasts that they publish.

The individual members' forecasts are not provided numerically, only graphically. We have to extract the numerical values from the graphs, and this is possible from October 2015 onwards (with the exception of April 2020, where the graph does not depict the individual forecasts). The extraction is done by drawing a horizontal line to project the points to the axis and then reading off the corresponding measurement.

The resolution in the published graphs is sufficiently high that we are confident that our extracted forecasts are accurate with an absolute precision of one decimal place. Thus, we obtain nine ordered individual forecasts for each of the Bank's GDP or CPI forecasts. We cannot link a specific forecast to a specific member of the Policy Board, and therefore we cannot follow the members through time or compute correlations between the members' forecasts.

The data collected from the graphs in each report were checked against the numerical data in the tables. Some care is required before we can use the extracted data for analysis. This applies in particular to inflation, where the members are asked to forecast the CPI (all items less fresh food). In several reports (for example, April 2018), the data on the graphs exclude the effect of the consumption tax hike. However, all members agree that this effect is 0.5 percentage points in 2019 and 2020 (see April 2018, footnote 4, page 8), and we therefore added 0.5 to the forecast made by the Policy Board members to produce the final CPI (all items less fresh food) forecasts for 2019 and 2020. Several other adjustments are required so that forecasts from different reports and the actual realizations of the CPI (all items less fresh food) can be compared. All adjustments are explained and justified below.

The extracted data are freely available as a Google spreadsheet at

https://bit.ly/3IOQTwG

and contain the following information extracted from BOJ reports:

- column A: report publication date.
- column B: report file name as downloaded from the BOJ webpage. It has the format "gorYYMMa.pdf" where YY is the year and MM is the month. For example, gor1804a.pdf is the April 2018 report.
- column C: variable name, rGDP for real GDP growth and CPI for inflation.

- column D: data code, $f1$ to $f9$ for ordered members' forecasts, where $f1$ is the largest value and $f9$ is the smallest value.
- columns E–M: members' forecasts extracted from the graphs at the end of the reports. The years from 2015 (in column E) to 2023 (column M) are forecasted in the reports starting from October 2015. Three years are forecasted at each time.
- columns O–AI: members' names for reference only. It is impossible to match the names against the forecasts using public data. The names in these columns indicate whether the person was a member when the report was published. Column O gives the total number of members in the position for each report, which is nine.

The data in the spreadsheet are the data extracted from the graphs in the BOJ reports. However, there are several exceptions and these need to be taken into account in the analysis. All exceptions in the data are highlighted in yellow in columns E–M with an added note for each group of highlighted cells.

- The April reports in each year provide forecasts for real GDP growth in the previous year, as the official numbers are not released. However, a lot of information about the previous year is already known, so many forecasts are identical, and the remaining are very close. Those numbers are not considered in the analysis.
- The April 2020 report does not provide the members' forecasts on the graph.
- In the October, July, and April 2018 reports, the CPI graph shows the 2019 and 2020 forecasts excluding the tax hike of 0.5. This number is added for comparison with other forecasts and the actual realizations. A similar adjustment is required for the 2019 forecasts in the reports produced in January 2018, and October, July, and April 2017.
- The forecasts for 2017 made in April, January 2016, and October 2017 also exclude the tax hike. However, these numbers are not used in our analysis.

The actual (realized) real GDP and CPI numbers

Year	GDP	CPI
2017	1.8	0.7
2018	0.2	0.8
2019	−0.5	0.6
2020	−4.5	−0.4

are obtained from the sk4.pdf report (GDP), available at

https://www.boj.or.jp/statistics/pub/sk/data/sk4.pdf,

and the sk5.pdf report (CPI), available at

https://www.boj.or.jp/statistics/pub/sk/data/sk5.pdf.