On Averaging-out Errors by Employing Substandard Indicators

Michal Dvorak
Department of Monetary Policy and Theory, University of Economics, Prague

April 7, 2014

Abstract

The article investigates three ways to form a final estimate using two different estimators, one of which exhibits higher variance or a bias: (1) taking the estimate produced by the better estimator, (2) taking their simple average, (3) taking their weighted average. It is shown that if there is no serious positive correlation, using both estimators is always preferable. Simple average is justified if both estimators exhibit similar variances or when variances are unknown. Weighted average is optimal in all other cases. The optimal weight rule is based on variances, not on standard deviations or coefficients of variation. The findings are useful for decision making in many situations, including the formation of the final valuation when valuing firms or securities with a multitude of approaches.

Keywords: Estimation, Final valuation.

JEL Classification: C13, G32

1 Introduction

In many practical situations we are to estimate an unobservable characteristic (say, price of a stock), using more indicators (say, valuations coming from two different multiples, for example P/E and MV/BV), each of them providing different estimate of the characteristic. We hypothesize both indicators are relevant. How to compose a final point estimate of the characteristic using these multiple indicators? Three options arise.

1. **Single indicator.** Take the estimate given by the indicator we believe it is the more precise one.
2. **Weighted average.** Assign weights to each of these estimates and calculate their weighted average.
3. **Simple average.** Take a simple average of these estimates.

Usually, the second approach is commonly adopted (Mařík et al., 2011; more applications should be given). It is based on a straightforward logic that better information should be weighted more heavily and no relevant information should be ommited. Nevertheless, it is difficult to come up with proper weights. Lack of guidance about weight assignment leads to arbitrary decisions (called *expert* judgment; in some cases this name is proper, in others it is a mere euphemism) and allows the analyst to justify *virtually any* value between the two indicators' estimates.

Which of these options is the best? Although the optimal solution differs situation by situation, certain general rules can be established. This article brings an easy guidance based on one of the most common dispersion measures – sample variances.

The article is structured as follows. In Chapter 2, the optimal weight variance rule is established for the baseline case of unbiasedness and independence of the indicators. Chapter 3 investigates if similar rules hold for standard deviations or coefficients of variation. Chapter 4 presents a numerical example. Chapters 5 a 6 study the variance rule if indicators are correlated or if one of them is biased. Chapter 7 concludes.

2 The Variance Rule

We consider the two-indicator case only, because, as we will see, the idea can be easily generalized to address multi-indicator cases. We label the indicators X_1 and X_2 and, in accord with the statistical terminology, we will call them estimators. Without the loss of generality, we consider X_1 to be the more precise (or, at least equally precise) one of these two. What is the precision measure? We will use Mean Squared Error[1] in this role. Mean Squared Error of an indicator X is defined as

$$\mathbf{MSE}(X) = \mathbb{E}\left\{(X-\mu)^2\right\} = (\mathbb{E}\{X\}-\mu)^2 + \mathbf{var}\{X\}, \tag{1}$$

where μ is the unobservable true characteristic (say, the correct price of a stock), $\mathbb{E}$ is the expectation operator and $\mathbf{var}\{X\} = \mathbb{E}\left\{(X-\mathbb{E}\{X\})^2\right\}$ represents sampling variance of the estimator. MSE is always nonnegative and smaller values define more precise estimators.

We will assume both estimators are unbiased (i.e. $\mathbb{E}\{X_1\} = \mathbb{E}\{X_2\} = \mu$). Even though biased estimators can improve aggregate precision, as we shall see in Chapter 6, in practice, nobody constructs biased estimators deliberately. For this reason, the unbiased case is taken as baseline in this article. By the unbiasedness property, only variance determines the MSE. Thus, estimators' standalone precision is

$$\mathbf{MSE}(X_1) = \mathbf{var}\{X_1\}, \tag{2}$$

$$\mathbf{MSE}(X_2) = \mathbf{var}\{X_2\}. \tag{3}$$

As we said before, X_1 is more (or equally) precise than X_2. This is is meant in the form of $\mathbf{MSE}(X_1) \leq \mathbf{MSE}(X_2)$, which, by the unbiasedness assumption, translates into $\mathbf{var}\{X_1\} \leq \mathbf{var}\{X_2\}$.

Now we consider a composite estimator C using X_1 and X_2. We denote it $C(\alpha)$ to emphasise it is a function of α:

$$C(\alpha) = \alpha \cdot X_1 + (1-\alpha) \cdot X_2, \tag{4}$$

[1] Often, Root Mean Squared Error, the square root of Mean Square Root, is used. This, however, does not alter the results, because square root is a strictly increasing function and a lower MSE always translates into a lower RMSE.

where $0 \leq \alpha \leq 1$ is the composition parameter.[2] The variance of the composite indicator $C(\alpha)$, as a function of α is

$$\begin{aligned} \mathbf{var}\{C(a)\} &= \alpha^2 \cdot \mathbf{var}\{X_1\} + (1-\alpha)^2 \cdot \mathbf{var}\{X_2\} + \qquad (5) \\ &+ 2 \cdot \alpha \cdot (1-\alpha) \cdot \mathbf{cov}\{X_1, X_2\}. \qquad (6) \end{aligned}$$

Under supposition the two original indicators are uncorrelated, which is likely to be the case when they are derived in a genuinely different way (this assumption is relaxed in Chapter 5), the covariance term is zero ($\mathbf{cov}\{X_1, X_2\} = 0$) and the variance shrinks to

$$\mathbf{var}\{C(a)\} = \alpha^2 \cdot \mathbf{var}\{X_1\} + (1-\alpha)^2 \cdot \mathbf{var}\{X_2\}. \qquad (7)$$

The α is our parameter of choice. It should be set to minimize the above-mentioned variance $\mathbf{var}\{C\}$. From calculus, the minimal variance can occur at

- **Boundaries** ($\alpha = 0$, $\alpha = 1$). The case of $\alpha = 0$ means the final estimate is based exclusively on X_2. The $\alpha = 1$ case means the final estimate is based exclusively on X_1. Out of these two, which is the preferred option? Because the variance of C is then equal to the variance of the estimator used, the more precise estimator, here X_1, should be utilized and the less precise estimator should be neglected. This corresponds to the single-indicator option mentioned of the introduction.

- **Local minima** ($0 < \alpha < 1$). How to find such points? Taking derivative of $\mathbf{var}\{C(a)\}$ with respect to α and setting it equal to zero produces

$$\frac{\partial \mathbf{var}\{C(\alpha)\}}{\partial \alpha} = 2 \cdot \alpha^* \cdot \mathbf{var}\{X_1\} - 2 \cdot (1-\alpha^*) \cdot \mathbf{var}\{X_2\} = 0 \qquad (8)$$

$$\alpha^* \cdot \mathbf{var}\{X_1\} + \alpha^* \cdot \mathbf{var}\{X_2\} = \mathbf{var}\{X_2\} \qquad (9)$$

$$\alpha^* = \frac{\mathbf{var}\{X_2\}}{\mathbf{var}\{X_1\} + \mathbf{var}\{X_2\}}. \qquad (10)$$

  Variances being positive (were one of them zero, we would know the true characteristic precisely and there would be no reason to engage in estimator composition) and finite, for any combination of $\mathbf{var}\{X_1\}$ and $\mathbf{var}\{X_2\}$ there exists exactly one value of α^*, which satisfies the condition and this value is always greater than zero and smaller than one. Being between 0 and 1 is neccessary for calling α^* weights.

  Because $\mathbf{var}\{C(\alpha)\}$ is a continuous function of α and for $\alpha < \alpha^*$, the function is decreaing, and for $\alpha^* < \alpha$ the function is increasing (which can be easily seen from its derivative), the α^* represent a global minumum. Thus, variance at boundary points $\alpha = 0$ and $\alpha = 1$ must be neccessarily greater. Boundary points are thus irrelevant.

  The Variance Rule. We have shown the optimal weight (labeled with asterisk) on a more precise estimator X_1 is

$$\alpha^* = \frac{\mathbf{var}\{X_2\}}{\mathbf{var}\{X_1\} + \mathbf{var}\{X_2\}}, \qquad (11)$$

[2] Although $\alpha \in \mathbb{R}$ without the restriction can in some cases improve the resulting precision, it renders one of the weights negative, which is difficult to interpret. Fortunately, as we will see, in most situations the optimal α will lie in the zero-one interval.

and the optimal weight on the less precise estimator X_2 is

$$1-\alpha^* = \frac{\mathbf{var}\left\{X_1\right\}}{\mathbf{var}\left\{X_1\right\} + \mathbf{var}\left\{X_2\right\}}, \tag{12}$$

It means both estimators X_1 and X_2 will be used to construct the statistically optimal composite indicator, even if one was individually much less advantageous (exhibiting much higher variance) than the other.

The alpha condition presents a rigorous rule how to attribute weights. If X_1 and X_2 are (or are thought to be) be equally reliable, the rule sets $\alpha = 0.5$. This correspons to the simple average option presented in the introduction. The less reliable X_2 is relative to X_1, in other words, the greater it's variance is relative to the variance of X_1, the less weight X_2 should receive.

If variances are uknown. In some cases it is not possible to estimate variances of the estimator any further beyond a qualified guess. For example, consider valuing a stock when knowing that typical P/E in the relevant industry is 6.5 and expected earings on our stock is USD 2.0, leading to an estimate of USD 13. What is the variance of such estimator? It is obvious there *is* some variance stemming from the fact the ratio 6.5 is taken from certain sample of companies which might differ from each other and also from our company; it is only unknown. Nevertheless, when an analyst arbitrarily assigns weight to this estimate, he or she is implicitly making an assumption on relative variances of the estimators. The analyst should be aware of the fact and check if these implied differences in variances are reasonable.

If variances are known. In other cases, sample variances are observable. Typically, the estimator value (estimate) is based on certain number of independent observations from a same distribution. For example, the analyst can use P/E ratios of 8 comparable companies and MV/BV ratios of 6 comparable companies. One thus can calculate sample variances of both estimators and derive weights in accord with the optimality condition on α.

It should be noted the variances (or their sample counterparts) must be calculated at the level of estimates and not at the level of particularities in their determination. For example, in the P/E case, one must calculate variance from the values of the assessed stock which are implied by each comparable firm's P/E ratio and *not* variance of the P/E ratios. More on this in Chapters 3 and 4.

Variance reduction techniques. While keeping the estimators unbiased (or at the expense of a small bias), variance of individual estimators can be reduced by using techniques of robust estimation. This also lowers the composite variance and also affects the choice of optimal α. The most common methods are these.

- **Median estimation.** If the estimate is based on more observations, take their median instead of mean.
- **Ignoring unrealistic observations.** If we believe some observation is not realistic, we disregard it. Obviously, as extreme values have dispro-

portionate impact (provided we do not use median estimation), it is of special interest to check plausibility of the extreme observations.

- **Trimming observations** according to some objective criterion. Say, values which lie more than 3 standard deviations from the median, are disregarded and the point estimate and variance is calculated from the remaining observations.

Although these methods are more robust than means, in the sense if data are bad[3], they perform much better. But higher robustness comes at the expense of precision when the data are well-behaved. Unless we know exactly the data flaws, it is not easy to devise the best correction strategy. As Greene (2012) states, alternatives to means might be favorable expecially when the sample size is small (which is often our case), though their setting is usually quite arbitrary and it is not easy to verify if these measures fare better than their non-robust counterparts in a particular case.[4] For this reason, variance-enhancement measures are left aside in this article.

3 Standard deviation and coefficent of variation in the alpha condition

In the previous chapter, it was shown the optimal α is based on variances. This corresponds to Vasicek (1973)[5] and it is a straightforward piece of probability theory. One could ask if other commonly used measures of dispersion, standard deviations or coefficients of variation can be plugged into similar formulae.

Standard deviations. It is apparent that

$$\tilde{\alpha}^* = \frac{\mathbf{std}\left\{X_2\right\}}{\mathbf{std}\left\{X_1\right\} + \mathbf{std}\left\{X_2\right\}}, \tag{13}$$

where $\mathbf{std}\left\{X_1\right\} = (\mathbf{var}\left\{X_1\right\})^{\frac{1}{2}}$ and $\mathbf{std}\left\{X_2\right\} = (\mathbf{var}\left\{X_2\right\})^{\frac{1}{2}}$, deviates from the optimality condition presented in the previous chapter ($\alpha^* \neq \tilde{\alpha}^*$), because if $\mathbf{var}\left\{X_1\right\} \leq \mathbf{var}\left\{X_2\right\}$,

$$\tilde{\alpha}^* = \frac{\mathbf{std}\left\{X_2\right\}}{\mathbf{std}\left\{X_1\right\} + \mathbf{std}\left\{X_2\right\}} \leq \frac{\mathbf{var}\left\{X_2\right\}}{\mathbf{var}\left\{X_1\right\} + \mathbf{var}\left\{X_2\right\}} = \alpha^* \tag{14}$$

and $\alpha^* = \tilde{\alpha}^*$ if and only if variances (and thus also standard deviations) are equal. In any other case, the standard-deviation based $\tilde{\alpha}^*$ attributes lower-than-optimal weight to the better indicator X_1, thus producing a supoptimal composite indicator.[6]

[3] In some applications, "badness" is caused by errors, for example in measurement. Nevertheless, in a great bulk of applications, "badness" means the values in the sample are heterodox and thus do not well describe the qualities of the item whose characteristic we are trying to infer, e.g. we are valuing Microsoft stock using Exxon Mobil P/E ratio. But extreme observation is not *ipso facto* bad, it might convey new information. For this reason, automatically disregarding extreme datapoints is a dangerous practice.

[4] Robust measures usually have only known asymptotical (for unlimited dataset) distributions and finite-sample variance estimates are either approximated or modelled using bootstrapping, which is downgraded if the sample size is small.

[5] Who implicitly used this argument to show prior information in Bayesian framework can enhance security beta estimates

[6] The proof that $C(\tilde{\alpha}^*)$ has greater variance/standard deviation/MSE/RMSE than $C(\alpha^*)$ arrives to the inequality between arithmetic and geometric averages of two numbers.

Thus, if one works with standard deviations, their square powers must be taken when calculating the weights.

Coefficients of variation. Another idea is to weight estimators in accordance with their coefficients of variation (the ratios of standard deviations to their means):

$$\tilde{\tilde{\alpha}}^* = \frac{\mathbf{cv}\left\{X_2\right\}}{\mathbf{cv}\left\{X_1\right\} + \mathbf{cv}\left\{X_2\right\}}, \tag{15}$$

where $\mathbf{cv}\left\{X_1\right\} = \frac{\mathbf{std}\{X_1\}}{\mathbb{E}\{X_1\}}$ and $\mathbf{cv}\left\{X_2\right\} = \frac{\mathbf{std}\{X_2\}}{\mathbb{E}\{X_1\}}$. The logic of coefficients of variation is intuitively appealing. They measure average dispersion as a percentage of mean values. Because most variables are assumed to be heteroscedastic (i.e. their absolute dispersion rises as mean rises), comparisons of absolute dispersion measures as variance and standard deviation is biased in favor of variables with lower means. To have dispersion measure comparable across variables with various means, standard deviation should be scaled by mean. In other words, if the dispersion is approximately proportional to mean, coefficient of variation stays the same.

One thus can tend to apply this "enhanced" measure to our two-estimator scenarios. But, what are the means here? As we said earlier, both estimators are unbiased, i.e. their expected values equal the unobserved true parameter μ. Then, their expected values are the same. Therefore, there is absolutely no reason to augment the standard deviation measure by scaling it with a constant. Were their expected values *not* equal to μ and different from each other, the bias (the difference between their expected value and μ) would be rather small and scaling would not make much sense either.

In our applications we don't see the expected values, but only means (of many observations) or a single value (if there estimate is based on only one observation). Using these means to construct the coefficients of variation adds another error to the coefficents. In other words, we use the same value in the estimates(X) and once again in determining their precision (**cv**). This is not very good.[7] [8]

From this reason, standard deviation should be used instead of coefficient of variation when determining weights. Standard deviations have been discussed earlier.

Finally, it is worth reminding that to have MSE measures comparable, it is critical to determine variances (or standard deviations) of estimates of the unknown characteristic μ and not variance of particularities in their determination (see the example in Chapter 4). This alleviates the need for using variation coefficients at all.

[7] Strictly speaking, even when computing variance or standard deviation we utilize the mean value. Nevertheless, the effect of the mean is much more pronounced in the coefficient of variation.

[8] There are statistical tests to reject that both means come from distributions with identical expected values. They might be used to decide (with some probability) if at least one of the indicators is biased. Such testing has 3 drawbacks. Firstly, the conclusion is never certain. Secondly, even the knowlegde that *some* indicator *is* biased, is not helpful in determining *which one* and *by how much*. Thirdly, even the existence of the bias does not justify using coefficients of variation as means of improving precision.

4 Numerical illustration

Let's illustate the ideas presented in the previous two chapter on a simple example. We have two indicators, price-to-earnings ratio (P/E) and market-value-to-book-value ratio (MV/BV) for 5 comparable companies labeled Bravo, Charlie, Delta, Echo and Foxtrot to determine the value of company Golf's common stock. Company Golf's earnings per share are USD 2.0 and its book value per share is USD 8.0. P/E and MV/BV multiples for similar companies are in the table below. The "implied price" is calculated as the product of Golf's dollar item (earnings per share or book value per share) and reference company's multiple. For simplicity, sample variances are computed as

$$\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}, \tag{16}$$

where x_i is i-th observation out of n and $\bar{x}$ is their arithmetic average, and standard deviations and coefficients of variance are derived from it. A more proper approach should be to use $n-1$ in the denominator to account for the fact the average is calculated from the very same observations which tends to artifically reduce the variance.

	P/E indicator		MV/BV indicator	
Comparison company	Ratio	Implied price	Ratio	Implied price
Bravo	5.5	11	1.2	9.6
Charlie	6	12	3.1	24.8
Delta	6.9	13.8	1.4	11.2
Echo	7.4	14.8	2.1	16.8
Foxtrot	8.5	17	1.6	12.8
Average	6.86	13.72	1.88	15.04

The ratios and the price of Golf's stock the ratios imply have the following descriptive statistics.

Level	Indicator	Mean	Variance	Standard deviation	Coefficient of variation
Ratio	P/E indicator	6.86	1.11	1.06	0.15
Ratio	MV/BV indicator	1.88	0.46	0.68	0.36
Implied price	P/E indicator	13.72	4.46	2.11	0.15
Implied price	MV/BV indicator	15.04	29.54	5.44	0.36

As we can see from the comparison of variances of both indicators on the implied price level, P/E is here the more precise measure (X_1). If optimal α is determined by plugging variances, standard deviations and coefficients of variation in the above-mentioned formulae, the following alphas are derived.

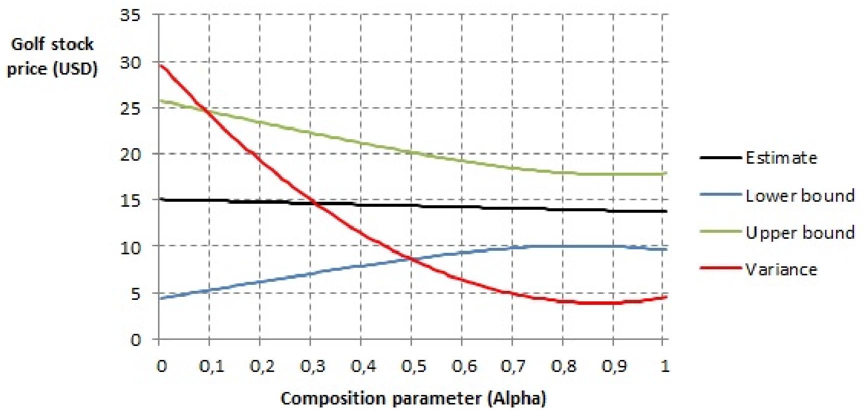
Measure	Implied price of Golf's stock	Ratios
Variance (α^*)	**0.8689**	0.2929
Standard deviation ($\tilde{\alpha}^*$)	0.7202	0.3916
Coefficient of variation ($\tilde{\tilde{\alpha}}^*$)	0.7014	0.7014

As mentioned in the previous chapters, the correct procedure is to use variances of the estimators, which are here the variances of implied prices (the result is in bold). From the table we can see the all the other options produce different weights. This necessarily means all other weightings are suboptimal. Especially, when we falsely compute dispersion of ratios rather than prices, terrible mistakes occur: the less precise measure receives much heavier weight than the more precise one! In this respect, only the coefficient of variation is fitted to work with ratios, though it still differs sligtly[9] from the weight when standard deviation of prices is employed, and it is always suboptimal vis-a-vis the Variance Rule.

Finally, the composite estimate of Golf's stock price is

$$C(\alpha^*) = \alpha^* \cdot X_1 + (1-\alpha^*) \cdot X_2 = \tag{17}$$
$$= 0.8689 \cdot 13.72 + (1 - 0.8689) \cdot 15.04 = 13.89 \text{ USD}. \tag{18}$$

If we plot point estimate, confidence bands of the final estimate of Golf stock price (based on 95% confidence using normal quantile 1.96) and composite variance[10] ($C(\alpha)$) for $0 \leq \alpha \leq 1$, we obtain the following chart.



Note that for $\alpha = 0$ we use MV/BV estimator only and for $\alpha = 1$ we use P/E estimator only. Lower variance and narrower bands for the optimal value of $\alpha^* \approx 0.87$ indicates the Variance Rule is an improvement over using only one of these two estimators.[11]

5 Correlation of estimators

In the baseline analysis we assumed there is no correlation between the estimators. Here we relax this simplifying assumption. Relative to the no-correlation case, the emergence of correlation

[9]This is not surprising. As expected values of both estimators are theoretically equal (or close to) the unobservable characteristic (μ), there is no theoretical distinction between standard deviation and coefficient of variation weights (of prices). Sampling fluctuations of means around these expected values force $\tilde{\tilde{\alpha}}^*$ to fluctuate (in both directions) around $\tilde{\alpha}^*$

[10]Although the actual correlation between the estimates (of prices) is –0.135, the variance was calculated under zero correlation to show the composition effects described in the baseline case where zero correlation was assumed. As shown in Chapter 5, negative correlation would lead to even better composite variances and X_2 would receive little heavier weight (i.e., optimal α would be lower).

[11]The confidence bands are constructed using a same methodology for all values of α and thus their widths are comparable. Nevertheless, its dangerous to claim they really represent the declared 95% reliability.

- **improves** the maximal attainable precision of the composite estimator $C(\alpha)$ by reducing the composite variance for *any* α, if the correlation between X_1 and X_2 is negative. This is because their standalone imperfections have the tendency to cancel out.
- **worsens** the maximal attainable precision of the composite estimator $C(\alpha)$ by increasing the composite variance for *any* α, if the correlation between X_1 and X_2 is positive. This is because similar standalone imperfections have the tendency to occur for both estimators simultaneously, degrading their diversification ability.

If both estimators are based on the same sample (say, same reference group of firms) correlation stems from several reasons, including the following four.

- **Estimators use same items**. If both indicators involve the same builing blocks, they will be correlated. It is not surprising P/E and Market capitalization/Net income will be the same and thus perfectly correlated. Similarly, valuations using DCF models under slightly different setups (for example, perpetuity and two-stage models) provide correlated estimates, as both use many simular inputs.
- **Estimator items affect each other**. For example, when using Price-to-sales and P/E, it is likely for all firms than earnings are high when sales are high, both affecting share prices in a similar way.
- **Estimator items are jointly affected by a third variable**. If a common third factor, for example, good management, positively affects both P/E and MV/BV for a firm, correlation occurs.
- **Sample is improperly selected**. If the sample includes certain subjects which provide high estimates under both estimators and other subjects provide low estimates for both estimators, correlation occurs.

This short list illustrates that correlation is likely to occur and will probably take the (unbeneficial) form of positive relationship and not a (beneficial) reversed one. We will address only the Variance Rule, as the results for standard deviation and coefficient of variation cases are analogical to the results in the Chapter 4. When the estimators are correlated, their variance is ($\mathbf{cov}\left\{X_1, X_2\right\} \neq 0$)

$$\mathbf{var}\left\{C(a)\right\} = \alpha^2 \cdot \mathbf{var}\left\{X_1\right\} + (1-\alpha)^2 \cdot \mathbf{var}\left\{X_2\right\} + \tag{19}$$

$$+ 2 \cdot \alpha \cdot (1-\alpha) \cdot \mathbf{cov}\left\{X_1, X_2\right\}. \tag{20}$$

Setting optimal weights again requires to differentiate $\mathbf{var}\left\{C(a)\right\}$ with respect to α and set the derivative to zero. It yields

$$2 \cdot \alpha^* \cdot \mathbf{var}\left\{X_1\right\} - 2 \cdot (1-\alpha^*) \cdot \mathbf{var}\left\{X_2\right\} + \tag{21}$$

$$+2 \cdot (1 - 2 \cdot \alpha^*) \cdot \mathbf{cov}\left\{X_1, X_2\right\} = 0 \tag{22}$$

$$\alpha^* = \frac{\mathbf{var}\left\{X_2\right\} - \mathbf{cov}\left\{X_1, X_2\right\}}{\mathbf{var}\left\{X_1\right\} + \mathbf{var}\left\{X_2\right\} - 2 \cdot \mathbf{cov}\left\{X_1, X_2\right\}}. \tag{23}$$

As the derivative is a linear function, we have only to check the local minimum (α^*) fulfills the condition of being $0 \leq \alpha \leq 1$. If yes, then it represents global minimum. If not, one of the boundary solutions ($\alpha = 0$ or $\alpha = 1$) is optimal. First we will check if $\alpha^* \geq 0$. We will show the α^* fraction has both the nominator and the numerator positive. Let's start with the numerator. The variance of a variable $X_1 - X_2$ is, as all variances, nonnegative. If this variable had zero variance, it must mean one of them is always by a constant higher (we said variances of both indicators are positive), which would violate the unbiasedness condition $\mathbb{E}\{X_1\} = \mathbb{E}\{X_2\} = \mu$. Because $\mathbf{var}\{X_1\} \leq \mathbf{var}\{X_2\}$, it implies

$$0 = \frac{0}{2} < \frac{\mathbf{var}\{X_1 - X_2\}}{2} = \frac{\mathbf{var}\{X_1\} + \mathbf{var}\{X_2\} - 2 \cdot \mathbf{cov}\{X_1, X_2\}}{2} \leq \tag{24}$$

$$\leq \frac{\mathbf{var}\{X_2\} + \mathbf{var}\{X_2\} - 2 \cdot \mathbf{cov}\{X_1, X_2\}}{2} = \tag{25}$$

$$= \mathbf{var}\{X_2\} - \mathbf{cov}\{X_1, X_2\}, \tag{26}$$

which proves the numerator is always positive. Now let's move to the denominator. The denominator equals the variance of a variable $X_1 - X_2$ and thus is again positive. As a result, α^* satisfies the positivity condition.

Second, we will check if $\alpha^* \leq 1$. This requires

$$\alpha^* = \frac{\mathbf{var}\{X_2\} - \mathbf{cov}\{X_1, X_2\}}{\mathbf{var}\{X_1\} + \mathbf{var}\{X_2\} - 2 \cdot \mathbf{cov}\{X_1, X_2\}} \leq 1 \tag{27}$$

$$\mathbf{var}\{X_2\} - \mathbf{cov}\{X_1, X_2\} \leq \mathbf{var}\{X_1\} + \mathbf{var}\{X_2\} - 2 \cdot \mathbf{cov}\{X_1, X_2\} \tag{28}$$

$$\mathbf{cov}\{X_1, X_2\} \leq \mathbf{var}\{X_1\}. \tag{29}$$

If correlation between X_1 and X_2 is defined as

$$\mathbf{cor}\{X_1, X_2\} = \frac{\mathbf{cov}\{X_1, X_2\}}{(\mathbf{var}\{X_1\} \cdot \mathbf{var}\{X_2\})^{\frac{1}{2}}}, \tag{30}$$

the less-than-unity condition can be simplified to

$$\mathbf{cor}\{X_1, X_2\} \leq \frac{\mathbf{std}\{X_1\}}{\mathbf{std}\{X_2\}}. \tag{31}$$

This time, the condition is not generally fulfilled. If the covariance is positive and higher than the variance of the more precise estimator, the interior solution does not exist and it is better to set $\alpha = 1$ and use the better estimator only.

Restated in terms of correlation, in means the correlation coefficent must be lower than the ratio of estimator standard deviations. As X_1 is more precise, the ratio is no greater than one. In the case when both are equally precise, the correlation has to be under 1 to use both estimators, which is almost granted. Nevertheless, given the fact positive correlations are likely to occur and estimator variances might substantially differ, in many applications it will be optimal to omit the less precise estimator.

The case under correlation can be summarized as follows. If the estimator variances are the same, correlation does not alter weight allocation as both receive equal weights again. If estimator variances differ, correlation does plays a role. If the correlation is negative, the optimality condition delivers lower α^*

and the better estimator is weighted less heavily (in some cases even less than 0.5 to fully exploit the diversification potential). Contrarily, if the correlation is positive, the better estimator receives heavier weight. If the correlation is highly positive and the less precise estimator is much less precise, the final estimate should be based on the more precise estimator only.

In many applications, including the example in Chapter 4, it is easy to measure sample correlations between the estimators and reflect this information in the optimal weight choice. As we have seen, the effect of correlation can be dramatic. Analysts are therefore strongly recommended to do so. In other applications (especially when there is only one value for the estimator) it is more difficult or even impossible to come up with correlation estimates as this is more difficult than making a qualified guess about variances. In such situations the users should stick to the no-correlation case.[12]

6 Biased estimators

In the last excursion we will depart from the baseline case of no-correlation and no-bias by investigating what happens if one of the estimators exhibits a bias. Again, we will assume it is X_2, which is biased ($\mathbb{E}\{X_2\} = \mu + \delta$, $\delta \neq 0$) and X_1 remains unbiased ($\mathbb{E}\{X_1\} = \mu$, $\delta \neq 0$). Here we make no assumption about standalone variances or standalone MSE. If there is a bias in one of the estimators, the composite estimator $C(\alpha)$ will generally be also biased, as

$$\mathbb{E}\{C(\alpha)\} = \mathbb{E}\{\alpha \cdot X_1 + (1-\alpha) \cdot X_2\} = \tag{32}$$

$$= \alpha \cdot \mathbb{E}\{X_1\} + (1-\alpha) \cdot \mathbb{E}\{X_2\} = \tag{33}$$

$$= \alpha \cdot \mu + (1-\alpha) \cdot (\mu + \delta) = \mu + (1-\alpha) \cdot \delta \neq \mu. \tag{34}$$

If biases are present, considering variances is no more equivalent and we have to resort to MSE to determine the optimal weights or to decide if some estimator should be completely disregarded. MSE of the composite indicator equals its bias plus its variance:

$$\mathbf{MSE}(C(\alpha)) = ((1-\alpha) \cdot \delta)^2 + \alpha^2 \cdot \mathbf{var}\{X_1\} + (1-\alpha)^2 \cdot \mathbf{var}\{X_2\}, \tag{35}$$

when we assume again zero correlation. After differentiating $\mathbf{MSE}(C(\alpha))$ with respect to α and setting the derivative equal to zero, we obtain

$$-2 \cdot \delta^2 + 2 \cdot \alpha^* \delta^2 + 2 \cdot \alpha^* \cdot \mathbf{var}\{X_1\} - 2 \cdot (1-\alpha^*) \cdot \mathbf{var}\{X_2\} = 0 \tag{36}$$

$$\alpha^* = \frac{\mathbf{var}\{X_2\} + \delta^2}{\mathbf{var}\{X_1\} + \mathbf{var}\{X_2\} + \delta^2}. \tag{37}$$

Now we will test if such α^* satisfies $0 \leq \alpha \leq 1$ to become the globally optimal weight (again, we can see the derivative of MSE is a linear function of α). First the positivity condition. As variances are nonnegative and δ^2 is positive, α^* is positive, too. Second the no-greater-than-one condition. As all terms are positive and $\gamma \equiv \mathbf{var}\{X_2\} + \delta^2$ is positive too, the condition requires $\frac{\gamma}{\mathbf{var}\{X_1\}+\gamma} \leq 1$, which is always fulfilled as inequality.

[12] Resorting to zero correlation is not just convenient simplification. From Bayesian perspective, if there are no clues what the correlation should be (if positive or negative), choosing zero correlation is the best option.

The presence of bias in X_2 means the unbiased indicator X_1 is, all else equal, weighted more heavily than in the no-bias case; the higher the bias, the heavier the weighting. Nevertheles, no matter how great the bias of the second indicator should be, it is still preferrable to use it. Why? Firstly, if the biased estimator has substantial variance relative to the unbiased one, under zero correlation the dispersion tends to cancel out (were there significant positive correlation, the biased estimator would be ommited). Secondly, if the biased estimator has low variance relative to the unbiased one, attributing some weights to the biased one reduces the volatility in the unbiased one. Although the composite mean value suffers a bit, the moderation is strong enough to lower the MSE.

Because $\mathbf{MSE}(X_1) = \mathbf{var}\{X_1\}$ and $\mathbf{MSE}(X_2) = \delta^2 + \mathbf{var}\{X_2\}$, it is possible to express the optimal weighting in terms of Mean Squared Errors of the indicators, thus stating a more general proposition.

$$\alpha^* = \frac{\mathbf{MSE}(X_2)}{\mathbf{MSE}(X_1) + \mathbf{MSE}(X_2)} \tag{38}$$

Nevertheless, the alpha conditions here are not very useful. In almost all applications, the bias δ and the bias squared are unobservable, because the bias is derived relative to the true unobservable characteristic μ, which is unobservable by name. Thus, we can't evaluate the optimal weight formula (α^*). Indeed, if we knew the bias in terms of its direction and magnitude, we would be able to add/subtract δ to/from X_2 to obtain an unbiased estimator (as $\mathbb{E}\{X_2 - \delta\} = \mu + \delta - \delta = \mu$) and apply the optimal weights from Chapter 2 or Chapter 5.

This analysis only showed that biasedness does not mean the estimator should be thrown away.

7 Conclusion

Returning to the three options how the compose a final estimate out of two indicators which were mentioned in the introduction, their general viability can be summarized as follows.

1. **Weighted average.** Assign weights to each of these estimates and calculate their weighted average. This is optimal solution in most cases. The weights should be set in accordance with the Variance Rule.

2. **Simple average.** Take a simple average of these estimates. This is the optimal solution if variances of both indicator estimates are (or are thought to be) identical. It is also the preferred choice if there is no information about them.

3. **Single indicator.** Take the estimate given by the indicator we believe it is the more precise one. This is the optimal solution if there is strong positive correlation between the indicators and their variances differ from each other. Counter-intuitively, if variance of one of the estimators is large or the estimator contains a bias, it is still better to use both estimators and reflecting this in the weighting scheme.

The optimality rule is the Variance Rule, if one uses standard deviations, they should be adjusted accordingly. Basing the weights on coeffcients of variation is erroneous.

As the optimal choice utilizes variances and does not rely on any other aspects of estimator distributions, it has very general validity and can be immediately employed to address a plethora of cases. The Variance Rule can be used in two ways. First, if sample variances and correlations are observable, the analyst should use them to obtain rigorous statistically-optimal weights. Correlations should not be automatically assumed away. In many practical situations there is considerable correlation, affecting the optimal weights allocation, or even recommending to disregard one of the indicators at all. Second, if sample variances and correlations are not observable, the analyst should verify, using the Variance Rule, if the relation of variances tactitly implied by the choice of weights is realistic.

References

[1] Greene, W. H. (2012). *Econometric Analysis.* 7th Edition. Harlow: Pearson.

[2] Mařík, M. (2011). *Metody oceňování podniku: Proces ocenění, základní metody a postupy.* 3. upravené a rozšířené vydání. Praha: Ekopress.

[3] Vasicek, O. A. (1973). A Note on Using Cross-Sectional Information in Bayesian Estimation of Security Betas. *Journal of Finance*, Vol 28, No. 5, pp. 1233-1239