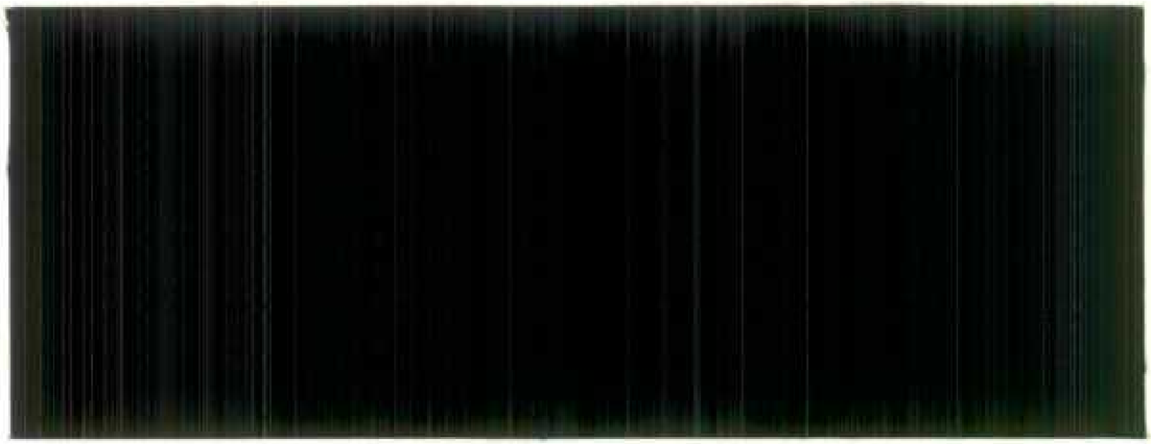
Statistics Canada   Statistique Canada

Agriculture Statistics Division



ALMV   OBN: 1870792

Statistics Canada
Agriculture Statistics Division
Farm Income and Prices Section

Working Paper

A description of Theil's RMSPE method in agricultural statistical forecasts

Number 01

by

Stuart Pursey

Published under the authority of the Minister of Supply and Services

September 1980
4-3104-535

TABLE OF CONTENTS

1.0 Introduction

The main purpose of this report is to present, explain, and show applications of Theil's Root Mean Square Prediction Error method. The method is an objective, numerical way of expressing the quality or "predictive performance" of an estimate of a parameter which is subject to later revision. Theil's method is based upon the concepts and methods used to describe the quality of an estimation procedure. These concepts, accuracy, precision, and bias, are explained below.

2.0 Part 1: Measurement of the Quality of an Estimation Procedure

The purpose of estimation is to assign a value to an unknown parameter. Hopefully, our method of estimation is such that it consistently provides estimates that are close to the value of the parameter. For illustration, let the parameter be Net Farm Income in 1975. This unknown value, in 1975, was estimated to be $4,327.9 million[1]. Conceptually, this estimate of $4,327.9 million is not unique - a different random selection of subjects in a probability survey or different personnel in a subjective estimation procedure are two of many examples that lead to differing values of an estimate. Let us suppose that an estimation procedure can yield n possible values for the parameter which we will represent by the symbol Z. Specifically, we denote the population or universe of possible estimates by z_1, z_2, ..., z_n. The survey yields one of the values , say z_{17}, but any of the other could conceivably occur. Objective, numerical measures of the quality of the z's are based on the notions of accuracy, bias and precision described in the next section.

---

[1] See reference 12.

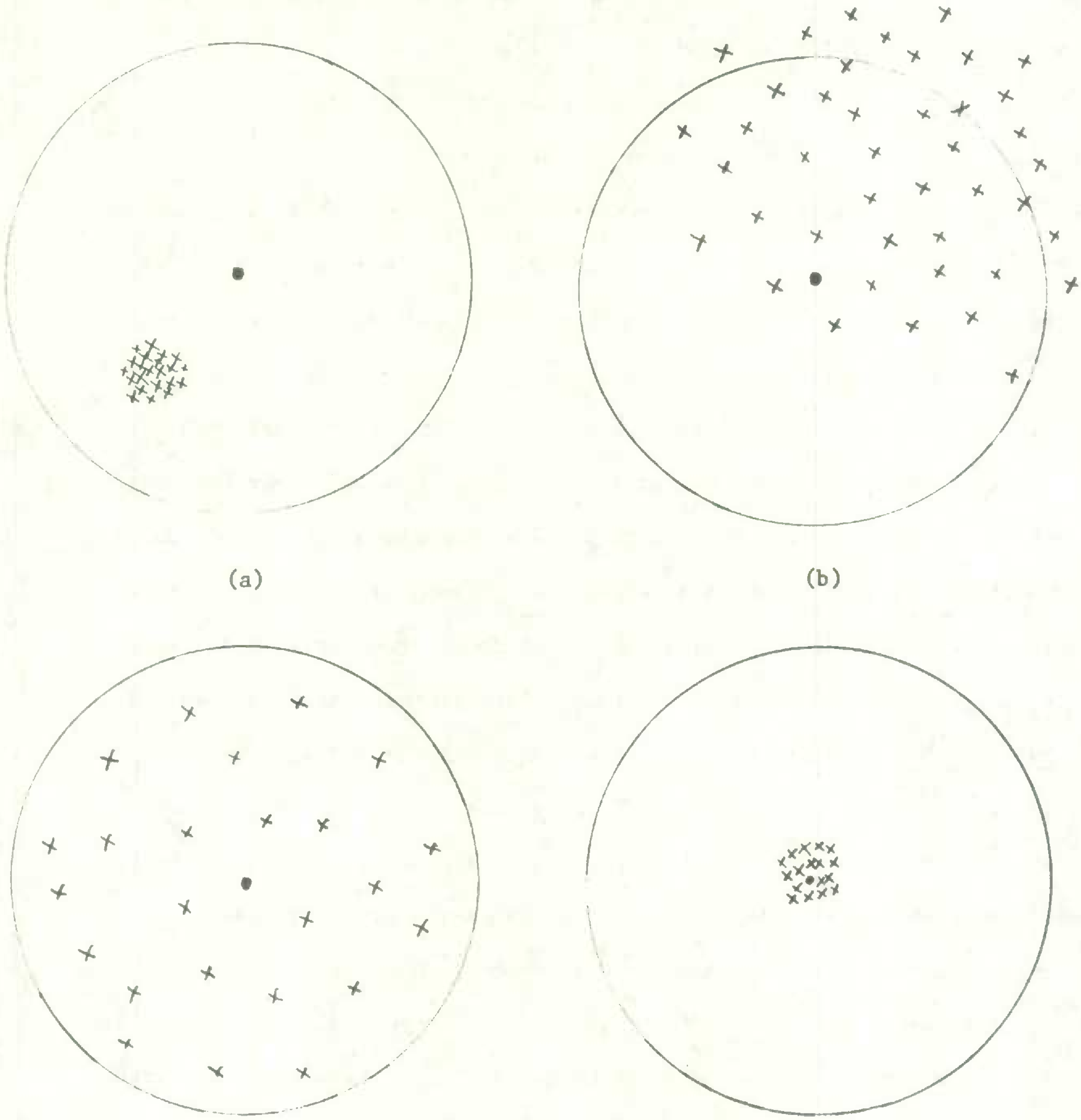
2.1 Three measures of quality – Accuracy, Bias and Precision

The concepts of accuracy, bias, and precision are often explained in terms of a marksman shooting at a target. In the analogy, the bull's eye represents the value of the parameter and the shots on the target represent the estimates. A perfect estimation procedure is analogous to the marksman hitting the bull's eye every time. This extract from Murphy (1961) is one such explanation.

> "It is, in fact, interesting to compare the measurement situation with that of a marksman aiming at a target. We would call him a precise marksman if, in firing a sequence of rounds, he were able to place all his shots in a rather small circle on the target. Any other rifleman unable to group his shots in such a small circle would naturally be regarded as less precise. Most people would accept this characterization whether either rifleman hits the bull's eye or not.
>
> Surely all would agree that if our man hits or nearly hits the bull's eye on all occasions, he should be called an accurate marksman. Unahppily, he may be a very precise marksman, but if his rifle is out of adjustment, perhaps the small circle of shots is centered at a point some distance from the bull's eye. In that case, we might regard him as an inaccurate marksman. Perhaps we should say that he is a potentially accurate marksman firing with a faulty rifle, but speaking categorically, we should have to say that the results were inaccurate."

Thus an estimation procedure that yields z_i's that are consistently close to each other (even if they are not close to the parameter Z) is deemed precise. Compare figures 2.1 a. and d. (precise) versus figures 2.1 b. and c. (imprecise). If the procedure yield's z_i's that "average out" to at, or near, the value of Z, then the procedure is deemed to be of low bias (even if individually the estimates are far away from Z). Compare figures 2.1 c. and d. (low bias) versus figures 2.1 a. and b. (high bias). Finally, if the procedure is both precise and of low bias, it is deemed accurate. Compare figures 2.1 a., b. and c. (inaccurate) versus figures 2.1 d. (accurate).

Figure 2.1: An Illustration of bias, precision and accuracy via the target shooting example.



(a) (b) (c) (d)

Low bias: (c) and (d)
Precise : (a) and (d)
Accurate: (d)

• — the true value
x — an estimate

Let us now consider the numerical definitions of accuracy, bias, and precision. For illustration, let us suppose we have two competing estimation procedures: A, which can yield only five possible estimates,

$$z_1 = 14,\ z_2 = 16,\ z_3 = 15,\ z_4 = 17,\ z_5 = 13;$$

and B, which can also yield only five possible estimates,

$$z_1 = 9,\ z_2 = 5,\ z_3 = 20,\ z_4 = 15,\ z_5 = 21.$$

Let us further suppose that the value of the parameter, Z, is equal to \$15. An estimation procedure is likely to yield many more than five possibilities. Differing estimates are the results of differing random selections of subjects, non-sampling errors, "seat of the pants" decisions, "rules of thumb", ... the list is endless. Not all the estimates are equally likely to be the estimate that we end up with as our representation of Z. If we estimate, for example, the average height of Canadians based on the heights of a random selection of 100 subjects, it is far more likely that the average height turns out to be around five and a half feet than around six and a half feet, although both estimates are quite possible. To explain the concepts of accuracy, bias, and precision, it is simpler to assume that each of the estimates $z_1, z_2, \ldots, z_n$ has the same chance of occurring. This simplification has no effect on understanding these concepts, but for completeness, a brief discussion of unequally likely z_i's is presented in Appendix 1.

(i) Accuracy

The numerical measure of accuracy is called the Mean Square Error (MSE). The MSE is calculated thus: for each of the n z_i's compute the quantity $(z_i - Z)^2$; then find the average of these n quantities.

The square root of this average is called the root mean square error of z.

In a formula, the MSE of z is:

$$\text{MSE }(z) = \frac{1}{n}((z_1 - Z)^2 + (z_2 - Z)^2 + \ldots + (z_n - Z)^2)$$

or using the sigma notation to denote summation,

$$\text{MSE }(z) = \frac{1}{n} \sum_{i=1}^{n} (z_i - Z)^2.$$

the root mean square error of z is

$$\text{RMSE }(z) = \sqrt{\text{MSE }(z)}.$$

Returning to estimation procedures A and B described above,

$$\begin{aligned}\text{MSE}(z \text{ for A}) &= 1/5((14-15)^2 + (16-15)^2 + (15-15)^2 + (17-15)^2 + (13-15)^2)\\ &= 2 \text{ square dollars}\end{aligned}$$

and

$$\begin{aligned}\text{MSE}(z \text{ for B}) &= 1/5((9-15)^2 + (5-15)^2 + (20-15)^2 + (15-15)^2 + (21-15)^2)\\ &= 39.4 \text{ square dollars.}\end{aligned}$$

Taking square roots,

RMSE(z for A) = 1.414 dollars;

and

RMSE(z for B) = 6.277 dollars.

The dispersion of the z's around Z is smaller for method A than for method B. Consequently, we judge that procedure A is more accurate than procedure B.

(ii) Bias

Let us now consider the quality of the estimation procedure "on the average" by using a quantity denoted E(z). E(z) is defined as the expected value of z; the value of z that we will get "on the average" or "in the long run". Under the simplifying assumption of equally occurring z's, the expected value is a simple average of all the z's,

$$E(z) = \frac{1}{n}(z_1 + z_2 + \dots + z_n)$$
$$= \frac{1}{n}\sum_{i=1}^{n} z_i.$$

For procedure A,

$$E(z \text{ for } A) = 1/5\ (14 + 16 + 15 + 17 + 13) = \$15$$

and for procedure B,

$$E(z \text{ for } B) = 1/5\ (9 + 5 + 20 + 15 + 21) = \$14.$$

Thus, "on the average" procedure A yields \$15 and procedure B yields \$14. Note that the expected value of procedure A, \$15, is equal to the value of Z. Procedure A is therefore called an unbiased estimation procedure. Bias, denoted B(z), is defined

$$B(z) = E(z) - Z.$$

Thus,

$$B(z \text{ for } A) = 15 - 15 = 0 \text{ dollars;}$$

and

$$B(z \text{ for } B) = 14 - 15 = 1 \text{ dollar.}$$

"On the average", procedure A is better than procedure B, if we measure quality by bias.

Overwhelming importance should not be applied to bias because experiments or surveys are never done "on the average". For example, suppose another estimation procedure yields z_1 = \$1 or z_2 = \$29. The expected value is \$15 = (1 + 29)/2 and so it is unbiased, but the procedure never yields an estimate close to \$15.

(iii) Precision

Thus far, two measures of the quality of an estimation procedure have been discussed, bias measured by $E(z) - Z$ and accuracy measured by the RMSE. Let us now consider a third quality called precision which is measured by a number called the variance of the estimation procedure. Under the simplifying convention of equally likely z's, the variance is equal to the average of the n quantities $(z - E(z))^2$. Thus,

$$\mathrm{Var}(z) = \frac{1}{n} \sum_{i=1}^{n} (z_i - E(z))^2.$$

This quantity is computationally similar to the mean square error. The only difference is that variance measures the dispersion about the expected value while the mean square error measures dispersion about the value of the parameter, Z. The standard error of Z is defined as the square root of the variance,

$$\mathrm{SE}(z) = \sqrt{\mathrm{Var}(z)}.$$

Noting from our discussion of bias that E(z for A) = \$15 and that E(z for B) = \$14, we find that,

$$\mathrm{Var}(z \text{ for A}) = 1/5\,\left((14-15)^2 + (16-15)^2 + (15-15)^2 + (17-15)^2 + (13-15)^2\right)$$
$$= 2 \text{ square dollars;}$$

and

$$\mathrm{Var}(z \text{ for B}) = 1/5\,\left((9-14)^2 + (5-14)^2 + (20-14)^2 + (15-14)^2 + (21-14)^2\right)$$
$$= 38.4 \text{ square dollars.}$$

Taking square roots,

SE(z for A) = \$1.414;

and

SE(z for B) = \$6.197.

Since the standard error for procedure A is less than that for procedure B, we say that procedure A is more precise than procedure B. Note that this measure of quality does not involve the parameter Z and consequently one might think that it is a poor measure of quality. However, for reasons discussed in Section 2.3, it is perhaps the most practical measure of quality.

2.2 The Coefficient of Variation

The coefficient of variation of z is defined,

$$C.V.(z) = 100\% \times SE(z)/E(z).$$

The definition of the C.V. is motivated by the following anomaly. Suppose, for example, that feed expenses in the Prairies are estimated to be \$250 million, that veterinary expenses are estimated to be \$30 million, and that the standard error of both these variables is \$10 million. It would thus not be surprising to find that the true value for veterinary expenses to be somewhere in the interval \$10 million to \$50 million* — quite a variability in plausible estimates. Conversely, we would quite reasonably consider the estimate for feed expenses, despite the same standard error, to be quite stable — the true value being somewhere in the interval \$230 million to \$280 million* — i.e. quite invariable. In order to distinguish numerically between these two "equivalent" standard errors, we define the coefficient of variation as a measure of variability with respect to the mean. Thus we get C.V. = 4.0% for the feed estimate and C.V. = 33.0% for the veterinary expense estimate. This number quite effectively demonstrates that relatively or proportionately, the feed estimate is the better estimate, despite the "equivalent" standard errors.

---

* These limits are (estimate — 2 S.E.) to (estimate + 2 S.E.)

The C.V. is a useful measure to compare the relative quality of estimates for disparate items (such as feed and veterinary expenses) whose totals are expected to differ widely — expense estimates for miscellaneous tools, fertilizer, insurance and heating fuel are several more examples. Also, since the C.V. is unitless, it is feasible to use it to compare the precision between items that have different units such as expenses, livestock numbers, and acreages.

Returning to procedures A and B,

$$C.V.(A) = 100 \times 2/15 = 9.43\%$$

and

$$C.V.(B) = 100 \times 38.4/14 = 44.26\%.$$

Thus relative to their means, Procedure A is more precise than Procedure B.

2.3 The Relationship between accuracy, bias, and precision and their estimation

This section has discussed three basic measures of quality:

1. accuracy measured by,

$$MSE(z) = \frac{1}{n} \sum_{i=1}^{n} (z_i - Z)^2;$$

2. bias measured by,

$$B(z) = E(z) - Z; \text{ and}$$

3. precision measured by

$$Var(z) = \frac{1}{n} \sum_{i=1}^{n} (z_i - E(z))^2.$$

Now it is known that these three measures of quality are related mathematically. Specifically,

$$MSE(z) = Var(z) + B(z)^2,$$

so accuracy is made up of two separate components: precision and bias.

Let us now ignore accuracy and concentrate on investigating precision and bias since accuracy can always be computed if we know the precision and bias.

The calculation of bias and precision, as presented in the formulae above, is impossible for two reasons. First, we do not know the value of Z (and if we did, we would have no reason to estimate it) and second, the estimation procedure yields only one z_i, say z_{17}, rather than all the z's. Let us instead attempt to estimate Var(z) and B(z) rather than compute them.

First we must examine the sources of error that cause bias and imprecision. The sources are divided into two broad classes — non-sampling errors and sampling errors.

Sampling errors exist when only a portion of the population is measured, rather than the complete population. Consequently the sample data, when it is manipulated to form the estimate, will not equal the true value, except by chance. The resulting error is called the sampling error.

In probability surveys, it is possible to estimate the average sampling error of all the possible samples based on the actual raw data of the one sample we happened to get in the survey. That is, not only does the raw data provide the estimate z, but it also provides an estimate of the sampling error portion of bias and precision. The formulae to compute the estimate of sampling error is dependent upon the type of probability sampling design chosen — it may be simple random sampling, cluster sampling, stratified sampling or any of a number of other schemes (or even combinations of sampling designs). The main point about probability sample surveys is that given the sampling design and given the subsequently obtained raw data one can calculate the estimate z and mathematically derive and compute good estimates of the sampling error.

In non-probability sample surveys or subjective estimation procedures, a specific probability sampling scheme is not used. From the point of view of delving into the procedure to find out the quality of the estimates it produced, non-probability surveys are extremely complex. Basically we are left with only z and no way of measuring the sampling error.

As its name suggests, non-sampling errors are all errors that are not due to a sample (rather than complete) portion being taken of the population. Several of an endless list of non-sampling errors are: calculation errors, non-response, mis-response, data capture errors, ... and so on. The detection and measurement of this type of error is very difficult. The best solution is to avoid the problem as much as possible. Consequently resources are spent detecting, minimizing and controlling these errors. Effective questionnaire design, training of enumerators, computer edit and imputation packages, post-survey quality checks and data capture verification are several examples of methods to control non-sampling errors. The undetected and uncontrolled errors, however, are most likely to be unmeasurable.

Under the worst of conditions – a non-probability survey and undetected and uncontrolled non-sampling error – data quality measurement is wholly unsatisfactory. Under the best of conditions – a probability survey and minimal non-sampling error – data quality is effectively but not perfectly measurable.

Suppose, however, that estimates are periodically revised. Theil (1954, 1963), has developed a method of estimating the quality of an estimate that has not yet been revised, relative to a later revision, based on original and revised estimates from previous years. Thus, contrary to the situations described in the foregoing, we will actually know the true value at a later time. The following sections deal with his method.

Part II. Theil's Root Mean Square Prediction Error Method

3.1 The calculation of the RMSPE

Table 3.1 shows estimates of Farm Realized Net Income for the years 1971 to 1978. The first column represents the original estimates or "predictions" and the second column represents the revised estimates. The quality or "prediction performance" of the original estimates for the years 1971 to 1977 is quite clear. For example, in 1971, the original estimate was off by \$1,167.4 − \$1,359.6 = − \$192.2 million or − 14.1%. For the year 1978, however, we cannot compute this. Perhaps the typical accuracy, as measured over the years 1971 to 1977, can be used to infer the accuracy of the estimate for year 1978. Theil's procedure is to calculate the average accuracy of the estimate over the years 1971 to 1977, and use it to infer the accuracy of the year 1978. Recalling from section 2.1 that the measure of accuracy is the mean square error

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (z_i - Z)^2,$$

we find that,

$$MSE(1971) = \frac{1}{1}(1{,}167.4 - 1{,}359.6)^2 = -192.2^2 = \$36{,}940.8$$

$$MSE(1972) = -141.3^2 = 19{,}965.7$$

$$MSE(1973) = 261.9^2 = 68{,}599.5$$

$$MSE(1974) = -372.1^2 = 138{,}437.6$$

$$MSE(1975) = -217.2^2 = 47{,}179.7$$

$$MSE(1976) = -378.0^2 = 142{,}874.9$$

$$MSE(1977) = -263.9^2 = 69{,}663.2.$$

The average MSE over these years is $\$273.5^2$ = \$74,808.8.

Table 3.1.1: Realized Farm Net Income Estimates, Canada, 1971-1978

	Original estimate (in November of year)	Revised estimate (in June of next year)
1971	1,167,392	1,359,579
1972	1,988,604	2,129,944
1973	2,968,266	2,706,351
1974	3,471,401	3,843,473
1975	3,959,277	4,176,486
1976	3,362,839	3,740,827
1977	3,264,703	3,528,641
1978	4,421,374	*

* Revised estimate expected in June 1979.

This number is called the Mean Square Prediction Error (MSPE). Its square root, \$273.5 million, is the Root Mean Square Prediction Error (RMSPE). Note that by its definition, Theil's procedure provides an estimate of accuracy, or equivalently, an estimate of precision plus bias squared. More formally, we define

$$\text{MSPE} = \frac{1}{n}\left((P_1 - A_1)^2 + (P_2 - A_2)^2 + \ldots + (P_n - A_n)^2\right)$$
$$= \frac{1}{n}\sum_{i=1}^{n} (P_i - A_i)^2,$$

where P_i is a prediction, A_i is the revised estimate, n is the number of years; and

$$\text{RMSPE} = \sqrt{\text{MSPE}}$$

The calculated RMSPE for realized net income is therefore,

$$\text{RMSPE} = \sqrt{\frac{1}{7}\left((1,167.4 - 1,359.6)^2 + (1,988.6 - 2,129.9)^2 + \ldots + (3,264.7 - 3,528.6)^2\right)}$$

$$= \$273.5 \text{ million.}$$

From this, we infer that the prediction performance or quality of the 1978 estimate is \$273.5 million.

The measurement of quality was performed above on the raw errors. A different approach to defining the error is to compute the percent error using the revised estimate as a base. This is 100% times $(P_i - A_i)/A_i$ and leads to a different RMSPE formula that computes the average percent error,

$$\text{RMSPE (on percents)} = 100\% \times \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{P_i - A_i}{A_i}\right)^2}$$

The effect of this approach is to make the raw error independent of the size of the revised estimate. For example, a one million dollar raw error on an item of size \$10 million (10% error) has the same percentage error as a much larger raw error of \$10 million on an item of size \$100 million (same 10% error). This characteristic of percent errors often makes good intuitive sense, and consequently percent errors are often used rather than raw errors.

Using percent errors, the RMSPE for original estimates of realized net income is

$$\text{RMSPE (on percents)} = \sqrt{\frac{1}{7}\left((-14.14)^2 + (-6.64)^2 + \ldots + (-7.48)^2\right)}$$
$$= 9.38\%$$

and consequently we infer the predictive performance of the 1978 estimate to be 9.38%.

This section has concentrated on the mechanics of calculating the RMSPE. The next sections deal with the applications of the RMSPE, the assumptions necessary to compute the RMSPE, and several miscellaneous points.

3.2 Valid Estimation of the RMSPE

Theil's method requires a stringent assumption. We must assume that the <u>true</u> RMSPE for <u>each</u> of the years is the same. If this is true, we can pool the information from all years to estimate the common true RMSPE and use it to infer the RMSPE for 1978 — <u>further</u> assuming that the <u>true</u> RMSPE for 1978 is equal to the common <u>true</u> RMSPE of the previous years.

Basically, the above paragraph is stating that for valid estimation everything must be truly representative of everything.

Since methods, personnel and the structure of data change, sometimes quite frequently, it may be difficult to find representative years to estimate the RMSPE. There seems no way out of this dilemna and consequently, valid estimation of the RMSPE will sometimes be impossible. The best approach seems to be to compute the RMSPE and examine the elements of it for at least some consistency from year to year. "Outlier" years – either far too good or far too bad in relation to the others – should probably be deleted and the RMSPE recomputed. In the next section, an example of this problem is demonstrated.

3.3 Applications of the RMSPE

There are two basic uses of the RMSPE: comparing predictive performances and computing confidence intervals.

(a) Comparing predictive performances

The RMSPE is a useful measure to compare the predictive performances of estimates between items (for example, cash receipts, total expenses and realized net farm income), between provinces and between types of estimates (for example, the forecast estimate, the projection estimate and the first published estimate).

Table 3.3.1 shows, for cash receipts, total expenses and realized income, RMSPE's for the forecast versus the first published estimate and the projection versus the first published estimate for each province and Canada, based on data for the year 1971 to 1977. Table 3.3.2 shows the raw data used to compute the total expense RMSPE's for P.E.I. and the subsequent calculations that lead to the RMSPE's for P.E.I. shown in Table 3.3.1.

Table 3.3.1: Root Mean Square Prediction Errors for the Forecast and the Projection versus the First Published Estimate on Total Expenses, Cash Receipts, and Realized Farm Net Income for Canada and Provinces (computed on percent errors and based on the years 1971-1977)

Province	Total Expenses		Cash Receipts		Realized Farm Net Income	
	I vs III	II vs III	I vs III	II vs III	I vs III	II vs III
	%		%		%	
Prince Edward Island	9.4	1.7	20.2	6.0	56.9	20.3
Nova Scotia	13.0	5.5	12.4	2.6	30.6	17.2
New Brunswick	11.8	4.5	16.7	4.8	46.6	17.9
Quebec	9.1	1.6	10.0	3.4	25.0	11.9
Ontario	8.6	2.4	10.7	3.6	26.0	14.5
Manitoba	13.0	3.1	16.6	4.5	35.7	13.3
Saskatchewan	10.3	3.3	18.7	3.9	38.2	11.3
Alberta	10.7	3.0	14.9	3.2	28.4	8.7
British Columbia	14.3	5.7	12.9	4.6	21.5	14.1
Canada	9.6	2.3	12.8	2.7	28.5	9.4

Source of raw data: Canadian Agricultural Outlook Conferences, 1971-1977.

I vs III — forecast versus first published estimate.
II vs III — projection versus first published estimate.

Table 3.3.2 Total Expenses for Prince Edward Island and the Calculations Leading to the RMSPE's Shown in Table 3.3.1

Year	Forecast	Projection	First published estimate
1971	39,000	37,005	37,517
1972	37,500	37,878	38,308
1973	37,963	44,341	45,819
1974	49,016	58,003	58,703
1975	65,339	66,929	65,592
1976	73,568	70,062	69,968
1977	73,915	72,565	71,816

Source: Canadian Agricultural Outlook Conferences, 1971-1977.

1. RMSPE for forecast versus first published estimate

$$\text{RMSPE} = \sqrt{\frac{1}{7}\left(\left(\frac{39{,}000 - 37{,}517}{37{,}517}\right)^2 + \left(\frac{37{,}500 - 38{,}308}{38{,}308}\right)^2 + \ldots + \left(\frac{73{,}915 - 71{,}816}{71{,}816}\right)^2\right)}$$

$$= 9.4\%$$

2. RMSPE for projection versus first published estimate

$$\text{RMSPE} = \sqrt{\frac{1}{7}\left(\left(\frac{37{,}005 - 37{,}517}{37{,}517}\right)^2 + \left(\frac{37{,}500 - 38{,}308}{38{,}308}\right)^2 + \ldots + \left(\frac{72{,}565 - 71{,}816}{71{,}816}\right)^2\right)}$$

$$= 1.7\%$$

The most immediate observation from Table 3.3.1 is that the projection is a much better predictor of the first published estimate than the forecast. There seems to be no vast differences between the provinces — although there is some suggestion that the estimates in the Maritimes are weakest. Finally, total expenses and cash receipts are measured with about equal quality — but with far more strength than the estimates for realized net income.

In Table 3.3.3, the second illustration, fifteen years of data were used to compute the RMSPE for hog production figures. But here, if we examine the trend of the percent errors over the past fifteen years, the necessary assumption that each year be representative of all the others seems to be broken. The last five years, 1972-1976, seem to form a definite group and it would seem that the RMSPE based on this group would provide a more realistic estimate of the current true RMSPE for hog production figures.

For this reason, the RMSPE was recalculated using the last five years. The RMSPE has changed from 14.5% to 1.9%, as shown in Table 3.3.4.

This illustration shows the usefulness of examining the trend of percent errors — first to pick out the correct years to estimate the RMSPE and second to provide a historical series of percent error which may prove to be of interest in its own right.

(b) Confidence intervals

A confidence interval is a lower limit and an upper limit created so that with a certain known probability (say 95%), the lower and upper limits will straddle the subsequently obtained revised estimate. If it is assumed that the predicted and revised estimates are both normally distributed, then 95% limits are

Table 3.3.3: An RMSPE for Hog Production Figures (,000's)

B.C. HOGS

YEAR	PREDICTION (1)	REVISED ESTIMATE (2)	ERROR	PERCENT ERROR
1961	47.0	41.6	5.4	12.981
1962	42.0	42.0	0.0	0.000
1963	37.0	38.0	-1.0	-2.632
1964	39.0	41.0	-2.0	-4.878
1965	36.0	39.0	-3.0	-7.692
1966	38.0	37.4	0.6	1.604
1967	44.0	50.0	-6.0	-12.000
1968	41.0	53.0	-12.0	-22.642
1969	38.0	51.0	-13.0	-25.490
1970	49.0	63.0	-14.0	-22.222
1971	47.0	73.5	-26.5	-36.054
1972	58.0	57.6	0.4	0.694
1973	52.0	51.2	0.8	1.563
1974	56.0	54.7	1.3	2.377
1975	56.0	54.3	1.7	3.131
1976	58.0	57.6	0.4	0.694

THEIL'S ROOT MEAN SQUARE PREDICTION ERROR

ON ERRORS : 9.001

ON PERCENT ERRORS : 14.534

Table 3.3.4: An RMSPE for Hog Production Figures (,000's)

B.C. HOGS

YEAR	PREDICTION (1)	REVISED ESTIMATE (2)	ERROR	PERCENT ERROR
1972	58.0	57.6	0.4	0.694
1973	52.0	51.2	0.8	1.563
1974	56.0	54.7	1.3	2.377
1975	56.0	54.3	1.7	3.131
1976	58.0	57.6	0.4	0.694

THEIL'S ROOT MEAN SQUARE PREDICTION ERROR

ON ERRORS : 1.053

ON PERCENT ERRORS : 1.942

(predicted − 2.00 times RMSPE) to (predicted + 2.00 times RMSPE). The number 2.00 is specifically associated with a normal distribution so if the distribution is not normal, then the use of 2.00 is not correct.

Each time confidence intervals are made, one should satisfy oneself that the distribution is normal or can reasonably be assumed to be normal. In probability suvveys, it is usually valid to assume a normal distribution (via the Central Limit Theorem of Statistics) but when estimates are subjectively obtained or obtained through non-probability surveys, the validity of such an assumption is open to much questioning. Consequently, for any specific set of data, the analyst must be prepared to think about this problem. In cases of doubt, it would be interesting to construct confidences and then see if the intervals cover the later-obtained revised estimates. If it is found, through experience, that this is the case (95% of the time), the analyst will have more confidence in making the intervals.

The major application and value, of the RMSPE, I feel, is its ability to objectively compare and evaluate the strengths and weakness of estimates between items, provinces and types of predictions.

3.4 Prediction of Changes Rather than Levels

Often percentage or level changes with respect to previous years are predicted rather than levels. That is, we may predict cash receipts to go up by (say) 11% or up by (say) $13 million rather than predicting the absolute level of cash receipts. The computation of RMSPE's based on changes is computationally identical to that for absolute levels − we simply replace the predicted level by the predicted change and the revised level estimate by the revised change estimate and then carry on computationally as before.*

Conceptually, it is possible to compute both an RMSPE based on changes and an RMSPE based on levels for each item. Consequently, the question may arise as to which type of RMSPE should be calculated. It seems reasonable to calculate the RMSPE based on changes when change estimates are published and to calculate the RMSPE based on levels when levels are published.

3.5 Other Methods of estimating prediction performance

The most interesting extension to Theil's method involves decomposing the RMSPE into a bias proportion, a variance proportion and a covariance proportion. Another method, not directly related to the RMSPE, is graphical in nature, and involves plotting the predictions versus the revised estimates.

These methods are discussed in Theil (1966) — an excellent source of information about forecasting accuracy.

3.6 Conclusion

Although the RMSPE method has limitations that the user must be continually aware of (a stringent assumption, a precise interpretation and applicability limited to estimates that are revised), the method does allow us to numerically express, relative to a later revision, the quality of a subjective estimation procedure in an objective manner. This, I feel, is most important because subjective statements of quality are nebulous and difficult to define and interpret consistently.

* Depending on the conceptual definition of a predicted change and a realized change, the RMSPE (on levels) can equal the RMSPE (on changes). No more about this will be said here, but see Tukey (1976) for his discussion on reasonable definitions of a realized change.

APPENDIX 1

Expected Values and Variances for Unequally Likely z_i's

In Section 2.1, it was assumed that each of the possible estimates,

$$z_1 = 14,\ z_2 = 16,\ z_3 = 15,\ z_4 = 17,\ z_5 = 13,$$

was equally likely to occur. Suppose, however, that z_2 occurs 96% of the time and the others occur 1% of the time each. It no longer appears that "in the long run" or "on the average" we will get \$15. Thus the expected value is properly defined for procedure A as,

$$E(z) = (.01 \times 14) + (.96 \times 16) + (.01 \times 15) + (.01 \times 17) + (.01 \times 13)$$
$$= \$15.95.$$

The expected value is actually a weighted average of the z's where the weights are the probabilities of occurance. Thus, $p_1 = .1$, $p_2 = .96$, $p_3 = .1$, $p_4 = .1$, and $p_5 = .1$ are the probability weights for procedure A. More formally, the expected value is defined

$$E(z) = p_1\ z_1 + p_2\ z_2 + \ldots + p_n\ z_n$$
$$= \sum_{i=1}^{n} p_i\ z_i.$$

Using the same reasoning, we define the variance of procedure A to be

$$Var(z) = p_1(z_1 - E(z))^2 + p_2(z_2 - E(z))^2 + \ldots + p_5(z_5 - E(z))^2$$
$$= .01(14 - 15.95)^2 + .96(16 - 15.95)^2 + \ldots + .01(13 - 15.95)^2$$
$$= 0.148 \text{ square dollars.}$$

Similarly,

$$MSE(z) = p_1(z_1 - 15)^2 + p_2(z_2 - 15)^2 + \ldots + p_5(z_5 - 15)^2$$
$$= .01(14 - 15)^2 + .96(16 - 15)^2 + \ldots + .01(13 - 15)^2$$
$$= 1.050 \text{ square dollars.}$$

...../2

From this we see that whatever we are measuring, whether squared deviations from the population mean or squared deviations from the value of the parameter, is a weighted average, where the weights are the probabilities of the z's. Thus, formal definitions of the variance and mean square error are,

$$\mathrm{Var}(z) = \sum_{i=1}^{n} p_i (z_i - E(z))^2$$

and

$$\mathrm{MSE}(z) = \sum_{i=1}^{n} p_i (z_i - Z)^2$$

respectively. Note that for equally likely z's, $p_i = \frac{1}{n}$ for all i.

It might be further inferred that estimates of these quantities through random samples also require a weighted average. However, this is not required because nature does the probability weighting for us. Just as a double six in dice throwing is rare to turn up, so also unlikely z's turn up rarely in random sampling. Consequently, the usual unweighted estimates of precision and the population mean, the sample variance and the sample mean respectively, provide correctly weighted estimates of these population parameters.

References

1. Andrusiak, G.W. (1978), Farm Income Situation and Outlook, Canadian Agricultural Outlook Conference Report (December, 1978).

2. Boyko, E.S. (1976), Farm Income Situation and Outlook, Canadian Agricultural Outlook Conference Report (December, 1976).

3. Boyko, E.S. (1977), Outlook for Canadian Farm Income, Canadian Agricultural Outlook Conference Report (December, 1977).

4. Gilchrist, V. (ed), (1971), Outlook: Farm Cash Income, Canadian Agricultural Outlook Conference Report (November, 1971).

5. Murphy, R.B. (1961). On the meaning of precision and accuracy. Statistical Concepts and Procedures; Vol. 1 of the Precision Measurement and Calibration (1969) series; edited by Terry H. Ku; National Bureau of Standards, Special Publication No. 300.

6. Porteous, W.L. (1972), Farm Income, Canadian Agricultural Outlook Conference Report (November, 1972).

7. Porteous, W.L. (1974), Farm Income, Operating Expenses, and Depreciation Charges, Canadian Agricultural Outlook Conference Report (January, 1974).

8. Porteous, W.L., J.W. Ross and E.S. Boyko (1975), Farm Income Situation and Outlook, Canadian Agricultural Outlook Conference Report (January, 1975).

9. Proud, R.B. and E.S. Boyko (1975), Farm Income Situation and Outlook, Canadian Agricultural Outlook Conference Report (December, 1975).

10. Purnell, G.R. (ed), Farm Cash Income Levels and Distribution 1969, 1970 and 1971, Canadian Agricultural Outlook Conference Report (November, 1970).

11. Shultz, W.M. (1978). On accuracy and precision in statistical estimates. Canadian Journal of Agricultural Economics 26, pp. 15-26.

12. CANADA. Statistics Canada. Farm Net Income/Revenu net agricole. Catalogue 21-202 (annual).

13. Theil, H. (1954). Who forecasts best? International Economic Papers, No. 5 (1955), edited by Alan T. Peacock, et al, The MacMillan Company, New York.

14. Theil, H., et al, (1966). Applied Economic Forecasting. Chapter 5, North Holland Publishing Co., Amsterdam.

15. Tukey, J (1976). Methodology, and the statistician's responsibility for both accuracy and relevance. Statistical Reporter, July 1976. pp. 253-262.