國立臺灣大學社會科學院經濟學系

碩士論文

Department of Economics

College of Social Sciences

National Taiwan University

Master Thesis

預測組合的演變及各學科中的組合概念

On the evolution of the forecast combinations and the pooling ideas

宋育芳

Yu-Fang Sung

指導教授：梁國源 博士

Advisor: Kuo- Yuan Liang, Ph.D.

中華民國 104 年 1 月

January, 2015

致謝

本論文立基於重視開創性的價值觀，批判性地回顧諾貝爾經濟學獎得主 C.Granger 提出預測組合的功勞。

感謝梁國源老師(元大寶華綜合經濟研究院院長)讓我能在經濟系裡做這種思想、歷史類的批判文章，並有這一分思辨自由。當初我因為推崇經濟學用模型分析處理議題的能力，抱著熱情進入經濟系，但隨著時日推移，我進一步接觸到來自其他學科質疑挑戰的聲音。我曾經嘗試將這些意見彙整到課堂上報告，期待透過討論，得到更深刻的看法，但卻遇到冷漠的對待、對於我帶來的挑戰嗤之以鼻。如今，因為有梁老師的這份學術雅量，我得到比較好的機會能好好的梳理文獻、根據證據發展較為縝密的論述，並藉此表達我的一些困惑和懷疑。梁老師多年深耕於預測組合的專業能力、熟稔相關文獻的深厚功底，提供本論文許多重要的觀點和資料，才避免這個作品流於膚淺的批判。

感謝簡錦漢老師(台灣中央研究院經濟所所長)提出的三點挑戰，讓我反省自己論述上的缺失。第一，預測組合中集思廣益的精神老早在平均值這個敘述統計量中就展現了，我的文獻範圍需要修正。第二，預測組合的流行，若果真是因為統計軟體的開發促成的，我應該要先證明計算預測組合是一件費神費力的事。第三，不應該只是為了找到論文吸引人的亮點，在摘要的撰寫上措辭過度激烈，還是應當以理服人。

感謝周文林老師(前香港中文大學經濟系教授，現任香港冠域經濟研究中心資深研究員)提醒我寫作論文應該明確指出自己的發現或貢獻。周老師秉持專業的精神，在口試會上直言不諱的挑戰本論文的貢獻，促使我反省自己的侷限，要求自己在內容和表述上進一步精煉。另外，周老師也費心地一一指點我本文內容游離、文法錯誤的地方，並提出他的修正意見，老師細心勤勉的精神，讓人感動。

感謝林惠玲老師(台灣大學社會科學院院長)辛勤辦理院務、照顧學子，促成獎學金有效的分配到需要的學生手上，我也蒙受其惠。老師當天在人事招待的提點，也讓我有所學習。

摘要

C. Granger 因為提出 Bates and Granger (1969)、Newbold and Granger (1974)及 Granger and Ramanathan (1984)奠定了他預測組合理論立基者的地位。但 Newbold and Granger (1974)的公式與 Markowitz (1952) 的公式在數理上完全相同，而且早於 Granger and Ramanathan (1984)已有 Crane and Crotty (1967)做過相似的事，以迴歸的方式綜合模型。另外，預測組合提出之後到 1974 年仍廣受批評，但之後卻又備受稱譽，這也顯示預測組合的得勢，可能有其創新貢獻之外的原因，其中的要因是電腦科技的發展。本文梳理這些脈絡，以較批判的角度，重新檢視 Granger 對預測組合的貢獻。

集思廣益、綜合資訊的觀念及做法也被運用在氣象學中的集合預報、心理學中的判斷預測及聚合分析、地震學中的地震預測及震測分析、生態學中的環分析。對組合概念的生發動機及流變有興趣的學者，可以參考本文已知組合概念的蓬勃發展。

關鍵字 預測組合、投資組合、預測涵蓋檢定、判斷預測、聚合分析、集合預報、地震預測組合、震測分析、環分析

Abstract

Though Granger is well recognized as a theoretical founding father of forecast combination, two evidences go against the general attribution of his contribution. First, the forecast combination formula and the derivation of its weights are exact the same as Markowitz (1952). Second, before Granger and Ramanathan (1984), Crane and Crotty (1967) had already combined forecasts through multiple regression. Furthermore, forecasts combination was heavily criticized before 1974, but was widely accepted after that. All these factors indicates the rising importance of forecast combination is due to reasons other than its innovative contribution. Studies shows the development of computer technology has played a role.

Other than the reposition of Granger's status, after 1969, various combination practices are developed as well, such as ensemble forecast in meteorology, earthquake forecast and seismic analysis in seismology, and loop analysis in ecology. Researchers who interest in the idea of combination may consult this thesis to have an understanding of the various findings.

Keywords *forecast combinations; portfolio selection; forecast encompassing; judgmental forecast; meta-analysis; ensemble forecast; combining earthquake forecasts; seismic analysis; loop analysis*

Contents

Exhibits

Tables

Chapter 1　Introduction

1.1 Theoretical development of forecast combination

Forecast combination is a simple and pragmatic way to possibly produce better forecasts. For example, in many cases, just the individual forecasts are available, rather than the information they are based on, and so combining is appropriate. The method was generally founded by Granger. Forecast combination method generally fall in to three catalog: variance-covariance method, regression-based method and Bayesian-based method (Diebold 2007; Liang and Shih 1994).

Begin with variance-covariance method (Bates and Granger 1969; Newbold and Granger 1974). Suppose there are M unbiased forecasts $\boldsymbol{F'}_T = (F_{1,T}, \dots, F_{M,T})$ of some quantity X_T. Then the linear combination

$$C_T = \boldsymbol{w'}_T \boldsymbol{F}_T, \qquad \boldsymbol{w'}_T \mathbf{1} = \mathbf{1}, \qquad 0 \le w_{i,T} \le 1 \; for\ all\ i,$$

where C_T is the combined forecast, $\boldsymbol{w'}_T = (w_{1,T}, \dots, w_{M,T})$ is weight, and $\mathbf{1}' = (1, \dots, 1)$.

Minimizing the variance of the combined forecast error will result in

$$\boldsymbol{w}_T = (\Sigma^{-1}\mathbf{1})/(\mathbf{1}'\Sigma^{-1}\mathbf{1}),$$

where

$$\Sigma = \boldsymbol{E}(\boldsymbol{e}_T \, \boldsymbol{e}_T{}') \text{ and } \boldsymbol{e}_T = X_T\mathbf{1} - \boldsymbol{F}_T.$$

In the theoretical perspective, Newbold and Granger (1974) is an extension of Bates and Granger (1969), which extend the a 2 by 2 variance covariance matrix to k by k one. The k by k variance covariance matrix require intensive computation, fortunately, Granger and Ramanathan (1984) propose regression-base method, and therefore popularize the use of forecast combination. Suppose there are m forecasts $y_{t+h,t}$ which made at time t for t+h,

$$y_{t+h} = \sum_{i=1}^{M} w_i \, y_{t+h,t,i} + e_{t+h,t},$$

by simply regressing realizations y_{t+h} on forecasts, one derive the weight w_i for each forecast. In fact, the optimal variance-covariance combining weights have a regression interpretation as w_i subject to $\sum w_i = 1$ and the intercept is excluded, which is known as method B. In practice, it is usually preferable not to constrain weights sum to 1 (but still exclude the intercept, method A) or more flexibly, lifting both constraints (method C) (Liang and Shih 1994). Summary of theoretical development of forecast combination is in Exhibit 1.

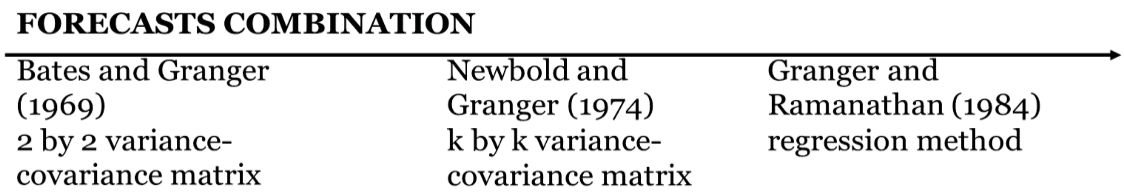


Exhibit 1 Evolution of forecast combinations

1.2 Impacts of forecast combination

Today, forecast combination have been applied in diverse area and have influence on the theoretical development of economics. According to Clemen (1989), the techniques of pooling forecasts was used as common practice in institutions such as the business outlook surveys of ASA/NBER since 1968, consensus macroeconomics forecasts of Blue Chip Economic Enterprises since 1976, and economic forecasts of The Financial Times at least before 1986. Successfully been applied in forecasting economics such as inflation, money supply, exchange rates, stock prices and sales forecasting. Among them, forecast combinations is used heavily in financial engineering especially. Application of combined forecasting has not been limited to economics, outcomes of football games, wilderness area use, check volume and many other things were all included.

Also, forecast combinations influences the development of economic theory. Rational expectations and efficiency theory have taken root in the spirit of forecast combinations (Holden and Peel 1989; Holden et al. 1985). For they all share the same theoretical position of maximum information (Bunn 1989).

1.3 Contributions of this thesis

The contributions of the thesis are three.

First, though Granger is well recognized as a theoretical founding father of forecast combination, two evidences go against the general attribution of his contribution. First, the forecast combination formula and the derivation of its weights are exact the same as Markowitz (1952). Second, before Granger and Ramanathan (1984), Crane and Crotty (1967) had already combined forecasts through multiple regression. Furthermore, forecasts combination was heavily criticized before 1974, but was widely accepted after that. All these factors indicates the rising importance of forecast combination is due to reasons other than its innovative contribution.

This thesis find that the development of computer technology has played a role. To support the explanation, this paper surveys early pooling approaches that may inspired Bates and Granger (1969). A brief introduction to the subsequent evolution of forecast combinations is included.

Second, the ideas of combination beyond economic forecasting and their inter-disciplines relations were presented, such as ensemble forecast in meteorology, earthquake forecast and seismic analysis in seismology, and loop analysis in ecology. Researchers interested in the idea of combination may consult this thesis to have an understanding of the various findings.

Third, the thesis adjusts and expands the exhibit in “historical development of combining forecasts literature” (Clemen 1989), which gave an influential[1] review of forecast combination.

[1] Cited 1531 times on Google scholar. Updated on January 17, 2015.

Chapter 2 From pooling ideas to the innovation of forecast combinations

Pooling techniques has been our daily practice from long time age. The simplest example, mean is a basic descriptive statistic for a distribution. Beyond mathematic, combining forecasts subjectively is also a simple example for the application of pooling ideas. People knew a sounder forecasting estimate can be obtained by combining and averaging estimates (Board 1963). In the suggestion of National Industrial Conference, to forecast sales well “[o]ne of the oldest and simplest methods of forecasting” (12) is to pool and average[2] the views of managers.

People employ combining techniques in forecasting because it can “broaden the base of forecasting” (12) and thus “obtain a sounder forecast of sales than could be made by a single estimator” (12). These reasons are just the same as Bates and Granger (1969). More applications of pooling ideas are demonstrated below, which are likely to inspire forecast combination method.

2.1 Portfolio selections (Markowitz 1999)─ pooling assets and mathematical equivalence

Risk diversification (pooling assets) has already been a common practice among business investors far before Markowitz proposed his “Portfolio selection” (1952). For example, in the Merchant of Venice, Shakespeare has the businessman Antonio say: “My ventures are not in one bottom trusted, nor to one place; nor is my whole estate upon the fortune of this present year; therefore, my merchandise makes me not sad.” (Act I, Scene 1). Same behavior can be found on modern businessman, investment trusts of Scotland and England in the middle of the 19th century provided diversification for their customers, their practices influenced modern investment companies. According to (Markowitz 1999), Wiesenberger's annual reports (since 1941) showed that these firms held large numbers of securities to diversify risks.

To an investor who wants to minimize the risks in his investment, it is not an asset's own risk that is important, but rather the contribution the asset makes to the variance of his entire asset allocation package. But before Markowitz (1952) mathematical tools only facilitated investor to calculate individual asset variance,

[2] Not in the numerical sense. In psychology, this is called “clinical method”. The decision-maker combines or processes information in his or her mind.

Markowitz innovates in calculating the variance of a weighted sum of the asset allocation package:

suppose one has M assets, the returns of assets is $\mathbf{R}' = (R_1, \ldots, R_M)$ and the expected return is $\boldsymbol{\mu} = \mathrm{E}\,(\mathbf{R})$, the expected return of a portfolio is

$$\mathrm{E} = \mathbf{x}'\boldsymbol{\mu}, \quad \text{where } \mathbf{x}' = (x_1, \ldots, x_M) \textit{ and } \mathbf{1}' = (1, \ldots, 1).$$

The variance of a portfolio is $\mathrm{V} = \mathrm{E}(\mathbf{ee}'), \quad \mathbf{e} = \mathbf{R} - \boldsymbol{\mu}$.

And the constraint is $\mathbf{x}'\mathbf{1} = 1, \quad x_i \geq 0$.

By Lagrange formula, the optimal investment amount is

$$\mathbf{x} = (\Sigma^{-1}\mathbf{1})/(\mathbf{1}'\Sigma^{-1}\mathbf{1}),$$

which is exactly the optimal weight in Newbold and Granger (1974). The mathematical equivalence indicates that Granger may learn from Markowitz.

Some famous scholars had already aware of the analogues between portfolio selections and forecast combination. Timmermann (2006) had said "the portfolio is the combination of forecasts and the source of risk reflects incomplete information about the target variable and model misspecification possibly due to non-stationarities in the underlying data generating process." Winkler (1989) had also said "[j]ust as investors create diversified portfolios to reduce risk, a combined forecast can be thought of as having a smaller risk of an extremely large error than an individual forecast." (Table 1)

Table 1 Similarities between portfolio and combination of forecasts

Portfolio selection	Vs	Combination of forecasts	Sources
Source of risk	Vs	Incomplete information about the target variable and model misspecification	Timmermann (2006, p. 139)
Diversified portfolios to reduce risk	Vs	Combine forecasts to have a smaller risk of an extremely large error than an individual forecast	Winkler (1989, p. 606)

Because the mathematical equivalence, **Markowitz's risk reduction effect can explain the improvement in the accuracy of forecast combinations under the minimum error-variance criterion**. Examples, such as: (1) Armstrong (2006) examined numerous forecasting methods for reducing forecast error and summarized that one of the advantages of forecast combinations is "spread risk". Arguably, his comment has taken root in Markowitz (1952). The method of portfolio is efficient in minimizing the overall risk within a portfolio by spreading risk through combining techniques. As a forerunner of forecast combinations, the method of portfolio may inspire Armstrong to describe the advantage of forecast combinations as risk spreading.

(2) Hibon and Evgeniou (2005) used the idea of risk reduction to explain the practical strength of forecast combination as well. They improved forecasting ability by firstly using a simple model-selection criterion to select among forecasts and gained a significant improvement in the accuracy of the selected combined forecast over that of the selected individual forecasts. Regarding this result, Hibon and Evgeniou (2005) said "[t]hese results indicate that the advantage of combining forecasts is not that the best possible combinations perform better than the best possible individual forecasts, but that it is less risky in practice to combine forecasts than to select an individual forecasting method." Like forecast combinations, we can say that "the advantage of a portfolio is not a portfolio will perform better than the possible best return and least risk asset, but that it is less risky or more practical than to put all your money on the possible-not-existed ideal asset."

(3) To spread risks, one would better to diversify assets across industries with different economic characteristics. The more distinct in the economic characteristics, the better. Same rule applies to forecast combination as well. To improve further the forecasting ability, component in the combined formula should be chose from models or people with distinct information sets (Granger 1989; Wallis 2011).All three examples illustrated here show the theoretical similarity between portfolio and forecast combinations (or to be critical, the inherited nature of forecast combinations from portfolio).

2.2 Two-stage forecasting model (Crane and Crotty 1967)--combining forecasts through multiple regression technique

Crane and Crotty (1967) is suspected as another forerunner of Granger's forecast combination. They combine forecasts through multiple regression technique with a regressor produced by exponential smoothing model (time series model) and other regressors (multiple regression model). The two-stage forecasting model took the advantage on the complement characteristics of time series analysis model and multiple regression model. Since time series models uses information contained in the historical movement pattern and is good at detecting and adjusting to changes in the forecast series, but does not use the information of independent variables and failed in the prediction of major changes in the trend, combining one into the other will aid in the forecasting ability. One thing noteworthy is that the two-stage forecasting model was even not an invention of Crane and Crotty (1967), it 'has been successfully applied to a problem important to asset management in banks, the forecasting of demand deposit' (505).

2.3 Combining multiple estimates before Bates and Granger (1969)

Besides, in the field of statistics, combining multiple estimates has been used since 1936 (Clemen 1989). Early proposition of was made by Edgerton and Kolbe (1936) and Horst (1936). The authors took a minimization of the sum of squares of the differences of the standard scores for the estimates, and maximization of the pairwise separation among the sample points respectively. Both techniques were similar in essence to least squares though were not used by modern researchers anymore. Not long before 1969, minimum squared-error combination of estimates were provided by Halperin (1961).

2.4 Granger's remark

Granger remarks that the idea of forecast combinations as his own idea. In the obituary of Granger published in the International Journal of Forecasting, Elliott (2009) described that the idea of Granger causality is likely inspired Granger himself to study further on the issue of forecasting. That is, the question of whether or not one variable results in another turned into the question of whether or not the variable is valuable for forecasting another variable. Granger himself attributed the finding to his observation based on the work of Barnard (1963). Granger found the result may be improved by taking a simple average. In his words, Bates and him then "just developed the idea" (Teräsvirta 1995, 587) that the predicting result is quite possibly better if not throwing one of the forecasts away but combining. The same words had been appeared several years ago when Granger was invited to review on the topic of forecast combination (Granger 1989). He referred the theoretical finding to his observation on Barnard (1963) as well.

But the mathematical equivalence between Markowitz (1952) and Newbold and

Granger (1974), the method similarity between Crane and Crotty (1967) and Granger and Ramanathan (1984), both indicate Granger may learn from others.

Chapter 3 The change in economists and statistician's attitude

Despite forecast combination received wide acceptances and induced numerous applications nowadays, it was heavily criticized by economists and statistician when the idea was first introduced (Newbold and Granger 1974). The main concern of the mainstream econometricians is to understand economic structural relationship by modelling so they put most of their efforts on aggregating information and trying to construct a robust model. In their view, if the model is "good", then a "good" forecast is guaranteed. In contrast, forecast combination combines forecast rather than combines information which those forecasts were based. This practice was at odds with the forecasting climate then.

3.1 Early appreciator

Although forecast combination is inferior to combining information directly, it can deal "with short, dirty time series with tools that managerial users of the forecast can understand" (Newbold and Granger 1974, 152). The comment above was done by Mr. Stern who worked in industry.

Early appreciators are practitioners who work in industry. Mr. Craddock who worked in Meteorological Office mentioned his own experience to support for forecast combination. He said:

> Whatever the views held on the combining of forecasts of time series obtained by different methods, there is no doubt that combined long-range weather forecasts, each based on several predictions founded on different physical principles, are better on average than the predictions given by any single method (156).

These evidence shows practitioners tend to value the pragmatic value of pooling approach. Moreover, Bates and Granger (1969) was published on operational organization journal instead of economic or econometric journals.

3.2 Late appreciator—the development of computer technology

The change in attitude reflects increasing appreciation on practical value. The popularization of forecast combination may attribute to the development of computer technology[3]. **Before the regression-based method was proposed (Granger and Ramanathan 1984), calculate forecast combination is computational intensive**. Because researchers have to calculate the k by k variance-covariance matrix.

It was until 1966, SAS was developed at North Carolina State University; SPSS was developed at SPSS Incorporate in 1968; Minitab was developed at the Pennsylvania State University in 1972. And it was not until then regression analysis became easier to use, for previously sometimes it took up to 24 hours to receive the result from one regression. The high time cost may be one of the reasons prevented experts to study further analytical tools for forecast combination. For the immature of computer statistics technology, Newbold and Granger (1974) which promotes the 'fully automatic' value of forecast combinations received heavy attacks from econometricians. It is likely to be the reason that led to the overall ignorance upon Crane and Crotty (1967) and other early forerunners as well.

Things began to change after 1975. Clemen (1989)'s exhibit showed an upward trend of cumulative number of articles of forecast combinations after 1975 (Exhibit 2). **The development of computer technology facilitates empirical researchers to do large scale computation which was in favor of combining approach**. M competition, a study that utilizes 1001 time series data to evaluate and compare the accuracy of different forecasting methods, shows the outperformance of the simple average method over all the individual methods (Makridakis et al. 1982).A further investigation on the combination issue vindicated the robustness of averaging (combining) approach (Makridakis and Winkler 1983). Numerous empirical results in support of forecast combination approaches and the timely publication of Granger and Ramanathan (1984) both led to popularize the use of forecast combination.

---

[3] Development in techniques or methods facilitate the development or popularization of theories. For example, Newton Raphson method were used to solve non-linear systems of equations, but because it required intensive computation, macroeconomic theories were hardly used it. It was until the popularization of an easier computational techniques, Gauss-Seidel method, in 1970s, macroeconomic theorists were facilitated to develop more complex models (Evans 1969).

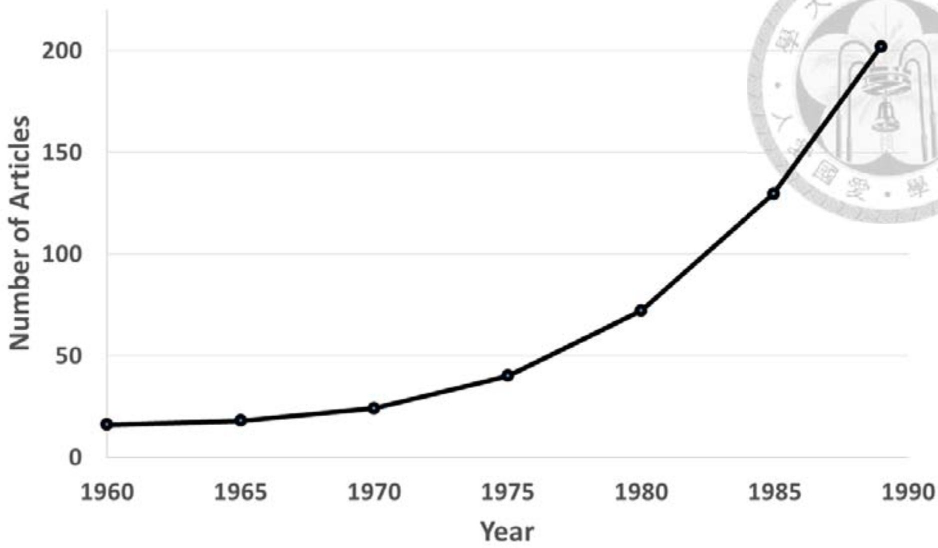


Exhibit 2 Cumulative number of articles published on combined forecasts (adapted from Clemen 1989)

Chapter 4 Evolution of forecast combinations and forecast encompassing

Forecast combinations was generally not accepted by econometricians in the beginning, therefore the development began in the field of operational researches where Bates and Granger published their paper.

4.1 Interaction between operational research and forecast combinations

Following Bates and Granger (1969) in the journal Operational Research Quarterly [ORQ] came a stream of articles in the same journal, including articles by Bunn (1975, 1977), Öller (1978) and Dickinson (1973, 1975). Dickinson (1973, 1975) investigated on the estimation of weight and looked further into the sampling distribution of weights. **Dickinson (1973)** used the minimum-variance criterion to analyze the sampling distributions of the weights, deriving the confidence limits for the estimates of the weights and of the variance of a combined forecast. The theoretical analysis showed the unreliability of the weight estimates, indicating a limited improvement of accuracy. Dickinson (1975)latter continued this study to show a minimum variance criterion will at least result in error variance that is no greater than that of any of the component forecasts. Recently, Liang et al. (2006) refined Dickinson (1973) on the distribution of the optimal combining weights by establishing a model of an inverted linear combination of two dependent F-variates. Moreover, they generalized the combining model to the case of combining three independent competing forecasts.

Dickinson (1975) also discussed the statistical properties of the weight estimators for the occurrence of negative weights. Bunn (1985) enriched the study of sign of weight by examining sign conditions under various weighting schemes, not only the error variance minimizing method but also equal weighting, optimal weighting with independence assumption, and three variations of a Bayesian combination. While Bunn’s work give a prototype of combining pair forecasts, Liang (1992) derived a general framework of multiple forecast combinations and contributed to provide a practical framework for quick check of the sign of weights. A brief summary is provided in Exhibit 3.

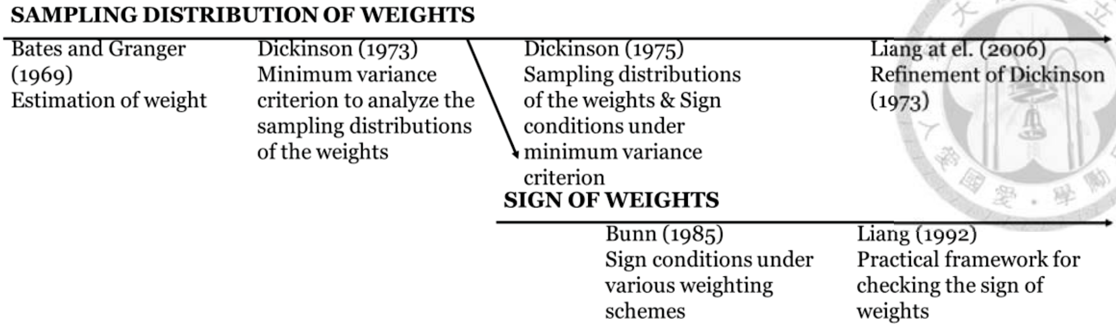


Exhibit 3 Theoretical development made by Dickinson (1973, 1975)

Bunn (1975, 1977) who worked on the topic of the forecast combination and who appeared in ORQ as well. He suggested a Bayesian outranking approach to enhance performance over small samples. **Bunn (1975)** showed a decision maker can meaningfully assigned subjective probabilities over a set of forecasting models and updated, according to a Bayesian process, so that one forecast will outperform another. **Bunn (1977)** compared the method utilizing subjective probabilities on the relative forecasting ability of each predictor (or said 'outperformance method') and the minimum-variance method by simulating experiments and found the former will outperform the latter if there is little prior information (less than 10 observations and possibly less than 30).

Öller (1978) developed the Bayesian framework with a set of self-scoring weights derived from the experts themselves. Each expert was asked to rate subjectively a given sum of confidence weights over his own forecasts. When the sum of the confidence weights is limited, these weights could function as weights for the computation of combined forecasts. According to a Bayesian process, records of the experts' previous performance can be used to adjust the confidence weights attached to the individual forecasts.

Operational research model is closely related to forecast combinations (Exhibit 4). **The proposal and the publication of the forecast combination method prospers the study of operational research, meanwhile, operational research models aid in forecast combinations for dealing with multiple objectives.** According to Clemen (1989),Lawrence and Reeves (1981), Reeves and Lawrence (1982) and Gulledge Jr et al. (1986) utilized multiple objective linear programming to minimize composite of various error statistics; Wall and Correia (1989) programmed a preferences optimization approach.

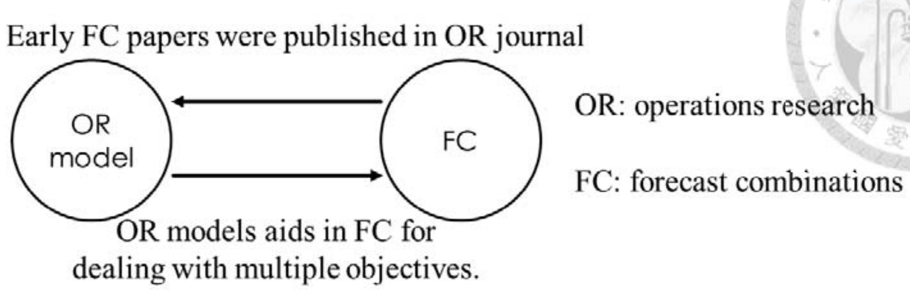


Exhibit 4 Operational research model closely related to forecast combinations

4.2 Late development

The development after Bates and Granger (1969) is enormous. Materials selected in this chapter are mainly based on Granger's reviewing work "Combining forecasts——twenty years later" (1989). More late reviewing articles can be found (Wallis 2011) but they have essentially taken root in Granger (1989).

4.2.1 Different information set

In the original Bates and Granger (1969) settings, the two forecasts were based on the same information set. Separating individual's information from common information, Granger concludes in the case forecasters share common information, equal weight combinations is useful. Moreover, it is useful to include more forecasts in the combination, even if the forecast is based on same information set, because a new forecast can improve the combination in the sense of adding individual information upon common sense. Also, Granger illustrate the usefulness of negative weights. Equivalent setting had been set up by Kim, 2001 #100@@author-year} who works independently in accounting and finance Wallis (2011).

4.2.2 Simple extensions

In the original Bates and Granger (1969) settings, **past values** of the forecast were not used efficiently. Researchers are encourged to include past values in combination. And if the data is not a stationary series, Granger recommanded to see Hallman and Kamstra (1989) as well as other generalizations. Moreover, multiple step forecasts are worth to try.

Another extension come with the idea of **time-varing weights**. Granger suggests a time-varying parameter regression using the Kalman filter. Also, one may refer to Engle et al. (1984).

Although Granger encourages include as more forecasts as possible, it is complicated when there are many forecasts available for combination. **Selecting data** is therefore important. Ranking all the forecasts on performance in terms of squared forecast error and leaving only the best is suggested as a trimmed way. Another decomposing way is proposed (Figlewski 1983) that forecast error was decomposed into two parts (commom error and individual error) and the wegihts was decided based on the relative sizes of the variances of the two components.

Combining forecasts with horizons longer than one period causes problems. Because one forecast may performs better over the other in a short-term but becomes worse in a long term and therefore the weights can vary with time horizon. Based on co-integration ideas, forecasts consistent with both short-term and long-term models at all horizons has been suggested (Engle et al. 1989). Transforming the data scale by taking log may help. Also, Granger has encouraged to develop the use of Batyesian updating schemes.

4.2.3 Combining probability distributions

More complicated extensions are associated with testing the conditions of encompassing. One of the necessary conditions for model P to encompass Q is that the economically relevant features (variance, confidence interval, quartile, etc) of the one-step forecast from P has to dominates the corresponding forecast from Q. How to test encomapsing of a combination need the knowledge of combining probability distributions. A relavant question is the combination of quantiles. By combining a pair of quantiles forecasts, for example the first quarter and the third quarter, to form a forecast interquantile range (Granger et al. 1989).

Combining probability distributions is more pertinent to the problem, which is, yet, basically ignored by economists but developed well by business and management school (Clemen 1989).In regards of no completely satisfactory combining technique in the literature, Granger proposed to a possible method. First, find the corresponding quantile function of each distribution function. And second, by inversing the combination of two quantile functions, to find a sensible combined function (Granger 1989).

4.2.3.1 Axiom approaches

However, on the topic of mathematical combination of probability distribution, it is inevitable to discuss axiom approaches, which focused on axiom-based aggregation formulas (Clemen and Winkler 1999). Aggregation of probability distribution has been long developed in management science and risk analysis journals, two common approaches are 'linear opinion pool' and 'logarithmic opinion pool'.

'Linear opinion pool' was proposed by Stone (1961) in the article "The opinion pool", in which a weighted linear combination of the forecasters' probabilities had been proposed. It combines subjective probability distribution to get group consensus in a mathematical approach. Let $fi(\theta)$ represent the probability distribution for a parameter θ of subject i. A consensus of probability distribution, denoted as a single distribution $f(\theta)$, can be written in a weighted average form. That is

$$f(\theta) = \sum_i w_i f_i(\theta),\ where\ w_i \geq 0\ and\ \sum_i w_i = 1.$$

Several weighting schemes were proposed for the method, including: simple average, weighted by ranking, weighted by self-rating, weighted according to the previous performance. Simply put, the weighs are determined subjectively (Clemen and Winkler 1999).

Another axiom approach is the logarithmic approach. The 'logarithmic opinion pool' is usually written using the geometric form, as

$$\frac{\prod f_i^{w_i}(y)}{\int \prod f_i^{w_i}(y)\ dy}.$$

With its own strength, logarithmic combination attracts scholars' attention. Logarithmic pooling method is convenient to manipulate. No matter first combine individual distributions, then update the combined distribution following Bayesian, or update individual distributions first, then combine, if with logarithmic pooling method, same results are derived; this property is said to satisfy the principle of external Bayesianity (Clemen and Winkler 1999, Wallis 2011).

4.2.3.2 Bayesian Approaches

Around 1980s, rising concerns about Bayesian approach shift attention from the axiomatic approach to the development of Bayesian combination models (Bunn 1989)[4].

---

[4] According to Bunn (1989), at the time he published Bunn (1975, 1977), there was an increasing acceptability of the "Bayesian approach to using multiple experts and different sources of evidence" (162), and this trend "reinforced the alternative idea of using multiple models for forecasting" (162).

Winkler (1968) and Morris (1974) have proposed a general Bayesian updating scheme to combine information and assess differential weights. Though some people give credit to Morris (1974) as the first establisher of Bayesian consensus model (Hall and Mitchell 2007), **Winkler (1968)** is probably the first researcher who proposed the primary framework. The Bayesian formwork was called "nature conjugate method", investigating the consensus of subjective probability distribution. He assumed that $f(\theta)$ represents a prior distribution, θ is the uncertain variable and i is information, and defines Bayesian theorem in the form

$$f(\theta|i) = \frac{f(\theta)\, l(i|\theta)}{\int\ f(\theta)\, l(i|\theta)\ d\theta},$$

where $l(i|\theta)$ is primitively interpreted as a sampling distribution or a likelihood function.

(Morris (1974)) enriched the Bayes' interpretation by decomposing the components of information z into two parts: one is from expert (denoted as e) and another from decision maker (denoted as d). Decision makers' prior probability assessment on θ, $f(\theta|d)$, will be altered upon reception of expert's probability assessment on θ, $g(\theta|e)$. The likelihood function $l(.)$ therefore explains how the decision maker subjectively feels about the credibility of the expert's probability assessment. The posterior probability distribution of decision maker can be write as

$$f(\theta|g(\theta|e), d) = \frac{f(\theta|d)\ l[g(\theta|e)|d]}{\int\ f(\theta|d)\ l[g(\theta|e)|d]\ d\theta},$$

where $\int\ f(\theta|d)\ l[g(\theta|e)|d]\ d\theta$ is the aggregation of the probability assessment of both decision maker and expert. Due to this sophisticated reinterpretation of Winkler (1968), (Morris (1974)) was credited as the first theoretical paper which is wholly consistent with the Bayesian view of probability. One thing notably is that Morris (1974)was published in the same journal, Management Science, as Winkler (1968), which indicates their inheriting relation.

Decisions in the face of uncertainty should be based on all available information, requiring combination of information obtained from models and experts; however, in the real world, due to common training and experiences of experts, the fact that experts have some sort of dependence is inevitable. With regard to the issue, **Winkler (1981)** presents a theoretical model which formally allows dependence among experts without requiring a prior for particular form of consensus density function. Normal results were presented and the sensitivity to the degree of dependence was found in the consensus distribution.

Inspired by Winkler (1981), **Agnew (1985)** extended further the Bayesian consensus model to the case in which dependent experts provide probability assessment on multiple unknown parameters. Moreover, it developed Bayesian sequential updating procedure, which uses experts' past performance to determine weights in each period.

The literature extended but frustrates from practical difficulties to find the likelihood function. Because of this, effort has gone into the practical models for aggregating single probabilities and probability distributions (Clemen and Winkler 1999).

4.2.3.2.1 *Bayesian* combinations of event probabilities

For Bayesian combinations of event probabilities there are independence approach, Genest and Schervish approach, Bernoulli approach, and Normal approach (Clemen and Winkler 1999).

4.2.3.2.2 Bayesian models for combining probability distributions

On the other front, there are Bayesian models which have been developed for combining probability distributions for continuous occurrence probability of a certain event.

The normal model has been important in this field. According to Liang and Shih (1994), the typical minimum-variance model for combining forecasts (Bates and Granger 1969, Newbold and Granger 1974) is consistent with the normal model (Winkler 1981, Bordley 1982). Moreover, a rewritten regression model (Granger and Ramanathan 1984) is equivalent to the normal model as well (Bordley 1986).In brief, I show the relations in Exhibit 5.

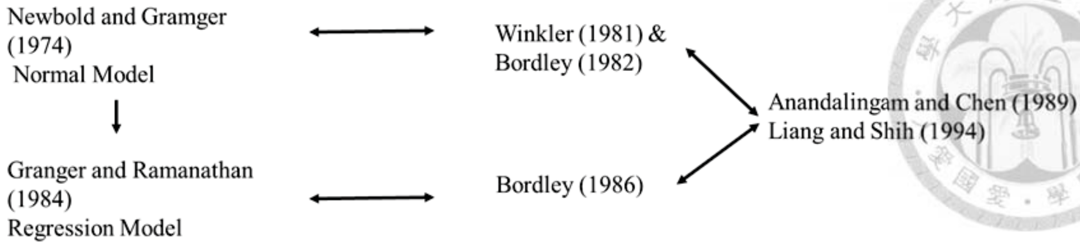


Exhibit 5 Connection between Bayesian Models for Combining Probability Distributions in the Normal settings and forecast combinations

By setting up the some necessary assumption, Bayesian combinations of probability are equivalent to forecast combination (Liang and Shih 1994). The Bayesian model in Winkler (1981) is equivalent to Newbold and Granger (1974) by assuming Normal distributed prior, Normal distributed likelihood and location invariant. Following Winkler (1981), the Bayesian model in Bordley (1982) also is in equivalence to Newbold and Granger (1974) by assuming uniform distributed prior, Normal distributed likelihood, known variance-covariance matrix, location invariant. Although the prior in Bordley is not normal, but since the prior is unimodal and symmetric[5], this is generally not a problem (Cleman and Winkler 1999). Besides, the mean of posterior density in (Bordley 1986) can be equivalent to the rewritten regression model in Granger and Ramanathan (1984) by assuming Normal distributed prior and Normal distributed likelihood. Summary is provided in Table 2.

Table 2 Bayesian equivalence of the theories of forecast combination

	Assumption of Bayesian interpretation	Result of Bayesian interpretation
Winkler (1981)	normal distributed prior, normal distributed likelihood and location invariant	equivalent to Normal model
Bordley (1982)	uniform distributed prior, normal distributed likelihood, known variance-covariance matrix, location invariant	equivalent to Normal model
Bordley (1986)	normal distributed prior and normal distributed likelihood	mean of posterior density is equivalent to the rewritten regression model

The theoretical evolution continues. Anandalingam and Chen (1989) generalized results of Winkler (1981), Bordley (1982, 1986), deriving their models respectively under different conditions. Liang and Shih (1994) relaxed further the assumption of unbiased decision maker's prior.

5 Even if the unimodal prior is just roughly symmetric, that would not be a problem (Clemen and Winkler 1999).

Although the normal model has been popular, it has some shortcomings, the obvious one is that a normal prior is required. As a consequence, several extensions are proposed (Clemen and Winkler 1999).

4.3 Forecasts encompassing

Nelson (1972) and Cooper and Nelson (1975) arouse an issue of forecast encompassing by using exactly the same formula as Bates and Granger (1969). The similarity of forecast encompassing and forecast combinations at first appearance had once induced me to categorize Nelson and Cooper's work as just an extension of Granger's work, but the distinguish idea in essence prevents me to do so. In contrast to Bates and Granger (1969) who skipped the evaluation steps and were satisfied with combining information just through combining multiple forecasts, Nelson (1972) and Cooper and Nelson (1975) evaluated the informational increment of an econometric model to the time series model, intending to synthesize model with combined information set. Nelson (1972) concludes that significant weights of both models in the combining regression is due to the inability of a model to include all available information. Cooper and Nelson (1975) followed the previous study, looking into the decreasing prediction errors in a post-sample test through the significance level of t-statistics in the combining regression.

4.3.1 Similar to Bates and Granger (1969) at first appearance

Nelson had written the formula of composite forecasts (Nelson 1972) to evaluate the prediction performance of the FRBMIT-PENN (FMP) econometric model of the U.S. economy by using the simple time-series models, an empirical representations of individual endogenous variables as stochastic processes of integrated autoregressive moving average (ARIMA) form, to establish standards of accuracy. The formula is,

$$A_t = \beta\,(FMP)_t + (1-\beta)(ARIMA)_t + \varepsilon_t.$$

Derived $\hat{\beta}$ as

$$\hat{\beta} = \frac{\sum[(FMP)_t - (ARIMA)_t][A_t - (ARIMA)_{t]}}{\sum[(FMP)_t - (ARIMA)_t]^2},$$

which after transformation was exactly same as the product of variance-covariance method in Bates and Granger (1969),

$$\hat{\beta} = \frac{\sum u_{2t}^2 - \sum u_{1t}u_{2t}}{\sum u_{1t}^2 + \sum u_{2t}^2 - 2\sum u_{1t}u_{2t}}.$$

However, he was not using the $\hat{\beta}$ as weight to optimize the forecasting ability, but rather he interpreted results in some other way.

4.3.2 Inherited the spirit of Markowitz (1952)

Nelson (1972) drew analogues between his work and Markowitz (1952)'s, saying that the $\hat{\beta}$ in encompassing formula is just **"as the weight for a minimum variance two-asset portfolio depends on the covariance of returns as well as on return variances" (Nelson 1972, 911).** In brief, both of them measures the influence of one thing (forecasting model or portfolio) by its impact to the whole, but not by its individual properties.

Rather than measure individual errors of each model, they use a composite forecast as a benchmark to measure the expected loss reduction in associated with a combined formula. Evidences showed that composite models were largely more accurate than FMP models but only accurate than ARIMA models in some cases, reflecting the inefficiency of certain FMP models in a sense that combining the FMP models with an ARIMA models significantly reduced the forecasting error of FMP.

Nelson (1972)'s motivation is standing from the viewpoint of the decision maker, making an overall evaluation of information contained in models. The question of whether one model or the other is more accurate is irrelevant, for a decision maker, his objective is to minimize expected loss, the contribution of one model should therefore measure by its comparison with the composite model. He had said (Nelson 1972):

> [F]rom the viewpoint of the decision maker the question of whether one set of predictions or the other is more accurate is irrelevant. Since his objective is to minimize expected loss, he will purchase any piece of information which reduces expected loss by more than its cost. Thus, the value of the ARIMA predictions, for example, is not measured by their individual errors, but rather by the contribution which they are able to make to the reduction in expected loss associated with a composite prediction or a set of composite predictions (913).

The idea underlying is just the same as Markowitz (1952). Markowitz's work showed that it is trivial to look at a security's own risk, to an investor, the important thing is the contribution the security brings to the variance of his entire portfolio. Comparable in idea to Markowitz, Nelson knew the important thing to a decision maker is the overall contribution one model makes to the forecasting ability of the composite forecast rather than individually evaluate the accuracy of each model.

4.3.3 Forecast combinations and encompassing

The connection of forecast combinations and encompassing was discussed by several studies. Note that, however, nether Nelson (1972) nor Cooper and Nelson (1975) used the term 'forecast encompassing'. It was until 1986, **Mizon (1984)** coined the term. This model evaluation technique which essentially coincides with the forecast combination formula hereafter arouse concerns about the connection between it (forecast encompassing) and forecast combinations.

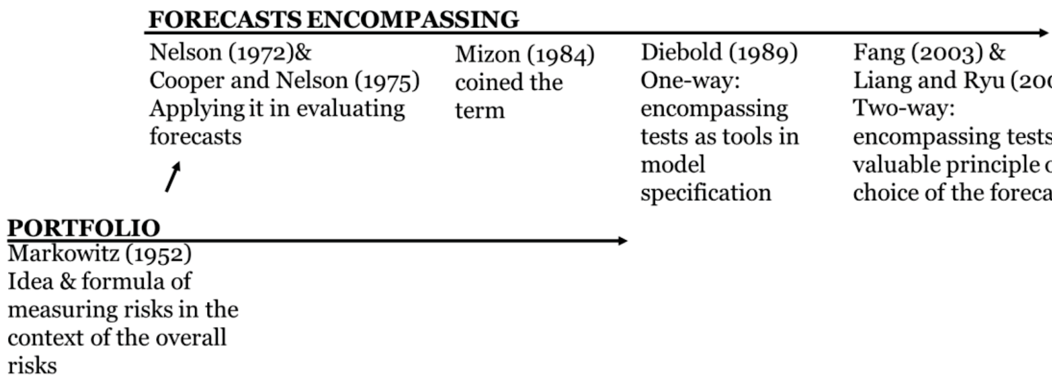
After Mizon (1984), regarding the changing characteristics of economy system and the consequence insufficient validity of each individual model to be passable overall performance, **Chong and Hendry (1986)** were motivated to investigate the suitable situation of using a system evaluation techniques. Their motivation indicating this paper was in line with Nelson's work, essentially, efforts to improve the model specification was encouraged. They investigated in 4 methods, among these system evaluation techniques, one of the approaches was forecast encompassing. Comparing the empirical and theoretical fitness of forecast encompassing with that of conventional methods, they concluded the forecast encompassing is both feasible and more promising than others.

Following Chong and Hendry (1986), **Diebold (1989)** inherited the viewpoint of using combining regression as information encompassing test. In his work, though the pragmatic virtues of forecast combinations was argued, the efficiency of combining forecast was still in doubt to Diebold. Eventually, using combining regression to facilitate the combination of information set was emphasized.

Fang (2003) extended Diebold (1989) not only demonstrated encompassing tests as tools in model specification but reversely demonstrated encompassing tests as tools to explain the accuracy improvement of forecast combinations. **Liang and Ryu (2003)** showed further encompassing tests as a valuable principle on the choice of the forecasts in the combining regression. Though without reference to Fang (2003), Liang and Ryu (2003) also established a two-way interaction between forecast combinations and forecasts encompassing.

Recently, more complex econometrics models are connected to the study of forecasts encompassing such as nested model, quintile forecasts, probability forecasts (Clements and Harvey 2010; Clements and Harvey 2011).

Brief summary is provided in Exhibit 6



.

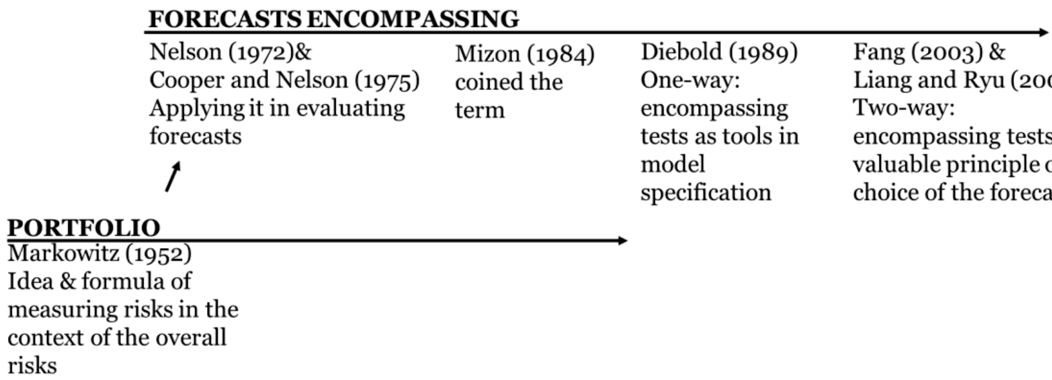


Exhibit 6 Evolution of forecasts encompassing

4.4 A modified exhibit

From the perspective of theoretical development, the evolution of forecast combination is provided in Chapter 1. From perspective of early application of pooling ideas, the impact of portfolio selection and two-stage forecasting model on forecast combination are suggested (Chapter 2). The late development of forecast combination is briefly introduced in this chapter. Also, the impact of portfolio selection on forecast encompassing is suggested.

In sum, a modified exhibit based on Clemen (1989)'s review is provided in Exhibit 7.

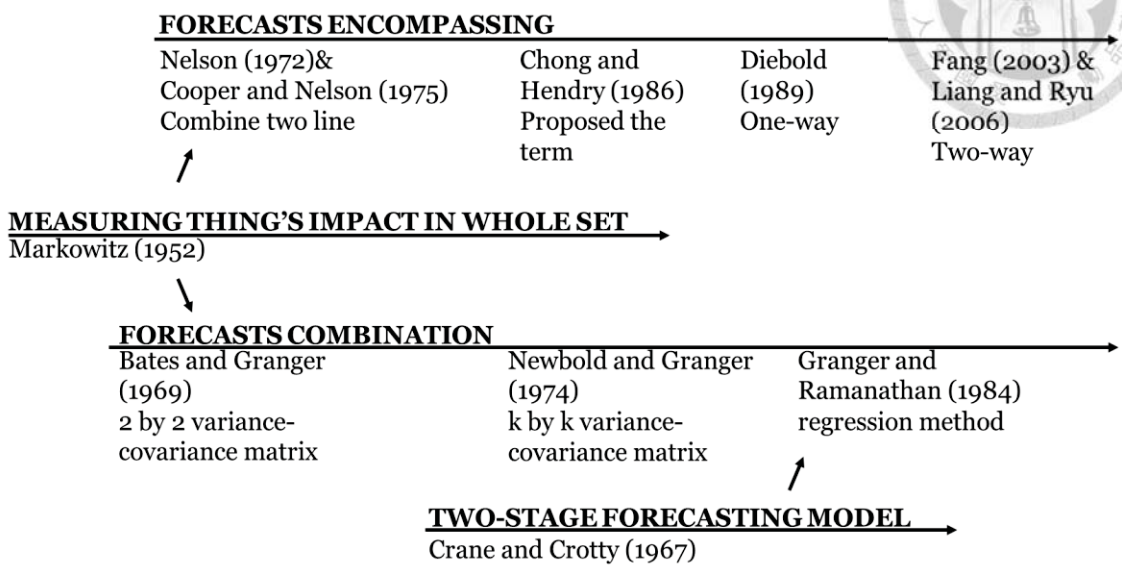


Exhibit 7 A modified exhibit based on Clemen (1989)

Chapter 5 The ensemble ideas in natural science

Natural science suffers no less for uncertainty problem than social science. Such as chaos in the fluid dynamics, no always best model for earthquake prediction and the complexity in ecology system all deeply bother scientists who pursue accuracy in prediction. Combination techniques are therefore employed by the nature scientists. In a sense, they are just like a business man who faces the uncertainty in business and address problem by seeking a better prediction method as well.

Researchers interested in the theory of combination may consult this thesis to have a brief understanding of how other researchers come up with the combination idea, and how does it connect to Granger's forecast combinations. Put simply, motivation of using combination ideas and its evolution or its connection to forecasts combinations will be stressed here.

5.1 Ensemble forecasting in meteorology

Forecasting weather in numerical way has been developed long ago but the innate chaotic nature of climate system made precise prediction impossible. A bit earlier than 1969, average forecasts was made to get a more precise results. Soon the concept of combining has taken root in meteorologists' mind.

The motivation for meteorologists to develop the techniques of forecast combination is due to uncertainty. This is just as same as Markowitz's reason. The main spirit of Markowitz's theory of portfolio is the underlying uncertainty in portfolios. Regard to uncertainty, they take a same processes, averaging estimates to generate a representative estimation.

This field may not be familiar by economists so I will first introduce its history and operational method.

5.1.1 Development of numerical weather prediction

It was until the mid of 19^{th} century, scientists came up with the idea that it should be possible to forecast weather from calculations based upon natural laws (FitzRoy 1863). Latter, accompanying with the development in physics (Bjerknes 1904) and computer technology, Charney (1951) achieve to apply his barotropic equation set (Richardson 2007). Weather forecasting has thereby come to dominate meteorology through its application in numerical weather prediction (NWP).

The quality of forecasting accuracy was poor in 1950s even just for a two-day forecast (Kalnay et al. 1998). However, the forecasting error has been reduced remarkably in the decade following. Technology has brought breakthrough, taking the leading forecasting agent, Met Office in United Kingdom, for example. From 1954 to 1966, the first operational system was established; from 1967 to 1978, computer has already facilitate scientists to deal with 10-level model that solved the Navier–Stokes equations of several weather character (including: fluid motion, the thermodynamic, heat transfer and continuity equations, etc.); from 1976 to 1992, NWP advanced in Mesoscale which is an intermediate scale between the scales of weather systems and of microclimates, meanwhile, a new 15-level model was developed to replace the 10-level model in 1982 for use in global aviation; more recently, efforts are put in the development of a unified climate–forecast model (Golding et al. 2004). The improvement of NWP basically follows after the improvement of computer equipment. No matter D. Hendry once had complained British government that "when the official weather forecasting service missed correctly forecasting a particularly damaging storm the response was to buy larger computers for the forecasters; when the economic forecasters failed to predict a major economic event it was decided to substantially reduce support for research in our area"(Granger 2001, 478).

5.1.2 Introduction of ensemble forecasting

Ensemble forecasting is one of the branches of numerical weather prediction (NWP) that allow us to estimate the uncertainty in a weather forecast as well as the most likely outcome. Instead of running the NWP model once, the model runs many times from very slightly different initial conditions to deal with the chaotic nature in weather forecasting. Two procedures are often used to modify the model settings within an ensemble. One is multi-model ensemble which use more than one model within the ensemble. The other is multi-physics ensemble which use the same model but with different combinations of physical parameterization settings(Organization 2012).

It follows that, rather than a deterministic forecast, ensemble forecast produces a probability which evaluate the probability an event to occur at a particular location. In fact ensemble forecast produce more than probability forecast, which will be covered latter. Exhibit 8 illustrates how a multi-model ensemble samples the uncertainty of the forecast.

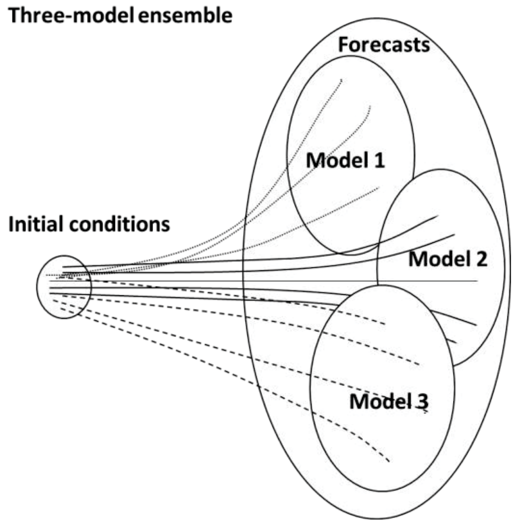


Exhibit 8 Concept of ensemble forecast (adapted from Fritsch et al. 2000, 572).

6.1.3 Operational method

Ensemble forecast applied to differing range of weather forecast, the types of ensemble forecast include global, regional and convective-scale ensemble forecast with various observation domains from 70 kilometers to 1 kilometers. The standard ensemble forecast products includes not only the probability forecast mentioned before but also ensemble mean, ensemble spread, quantiles, spaghetti maps, postage stamp maps and site-specific meteograms (Organization 2012).

Assigning weights in a combination is most critical in ensemble forecast. The process generally starts from sampling uncertainties and assign more weight on forecasts which have high resolution. Examples of the capabilities of numerical high resolution forecast are obvious differing weather characteristic in a location, such as the differing thermal, wind and precipitation patterns in a valley or slope; or the different weather patterns on the opposed side of a weather barriers (e.g. Alps, Andes, Appalachians) and distinction of resulting thermal, wind and precipitation patterns. Because forecast in these are is high resolution and therefore produces high precision, forecasters call them “high resolution” or “high control”. Comparing the relative capabilities of ensemble members with high resolution/control, forecasters assign weights in ensemble (EPS 2006).

Subjective modification of weight in an attempt to improve the distribution is accepted but not encouraged. For certain very short-period forecasts and for local forecasts over a small area subjective modification may help but not for longer-period forecasts or forecasts over a large area. Using the whole ensemble distribution in a probabilistic approach is strongly recommended (Organization 2012).

5.1.3 Motivation and evolution of ensemble forecast

5.1.3.1 Chaos

Lorenz (1963) mathematically described the unstable characteristic of the fluid dynamics. Given a finite systems of deterministic ordinary nonlinear differential equations which is designed to represent the hydrodynamic flow, the solution is sensitive to the initial conditions. A slightly differing initial states will quite frequently end up in considerably different states, which correspond to calm or stormy weather ('Edward Lorenz'). However, the chaotic nature of the fluid dynamics and the limited observation data certainly lead to uncertainty in the estimation of true initial state. As a consequence, it is hard for a single set of initial states to properly predict the final state. This influential paper established the field of chaos theory and have led to the theoretical development of ensemble forecasting.

Incidentally, **Sanders (1963)** investigate the uncertain nature of subjective probability forecasts by averaging subjective weather forecasts, illustrating the superiority of the combined procedures. Latter, Bosart (1975) also tested on the performance of the average subjective weather forecasts and confirm Sanders (1963)'s results.

5.1.3.2 Early development

Stochastic method was applied to "assess the value of new or improved data by considering their to assess the value of new or improved data"(Epstein 1969, 739). The Monte Carlo experiment results shows stochastic dynamic predictions have outperformed traditional deterministic procedure significantly in terms of mean square errors.

The ensemble conduct continued, as I mentioned before, in 1974, Craddock who worked in Meteorological Office was invited to the discussion of Newbold and Granger (1974) has commented on forecast combination in the perspective of ensemble weather forecast. In a consensus of combined ideas, professors from economic and business cooperated with meteorologists. Winkler and Murphy (1977) showed the outperformance of the average of subjective precipitation probability forecasts over individual opinion made by numerical weather forecaster. Clemen and Murphy (1986) average the subjective weather forecasts and model output statistics, finding again the superiority of the combined product over the individual one. Inspired by meteorology, professors from economic and business applied forecast combination techniques in the context of weather prediction. Clemen (1985) based on the Bayesian framework outlined by Morris (1974) to discuss whether human weather forecaster can bring new information to the mechanical guidance forecast and whether on human weather forecaster bring incremental information over the other, in the context of precipitation probability forecasts.

To sum up, meteorologists has tried to model weather system since 1951 but the innate chaotic nature prevent them to easily do so. Ensemble forecast has been put into practice since 1963 but was mature till 1999 which has to thank to the idea from forecast combinations (Newbold and Granger 1974) and the advanced computing power. Professors in economic and business were also inspired by meteorologists, doing forecast combinations researches in the context of weather forecast, such as applying forecast combinations technique (Winkler and Murphy 1977) or investigating the informational contribution in the context of weather forecast combinations (Clemen 1985). A brief summary is provided in Exhibit 9.

Ensemble forecasting provides more information to weather forecasters. In the process of producing ensemble forecast, researchers gain some benefit from comparisons. If the estimated forecasts vary a lot, researchers can thereby calculating a probability of the uncertainty for the final product. On the other hand, if the estimated forecasts are all very similar, forecasters may therefore have more confident on the accuracy of the final product. ('Ensemble Forecasting'). The application of ensemble forecasts has not been restricted to weather forecasting, it also applies to flood forecasting and bunches of projects (Cloke and Pappenberger 2009).

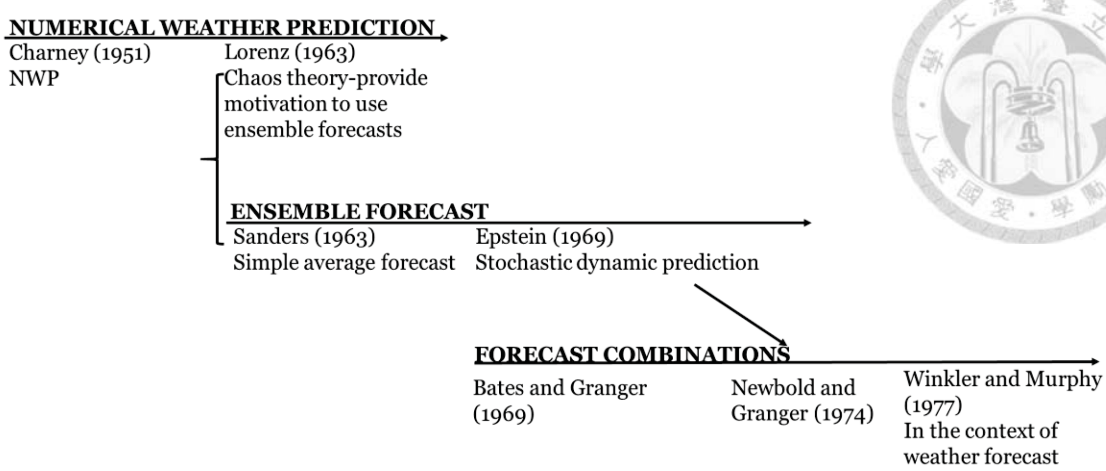


Exhibit 9 Evolution of ensemble forecasts

5.2 Seismology and Ecology

To enrich the findings in the application of combined ideas, embedded combined ideas in seismology and ecology are illustrated in this thesis as well, which potentially enlarge the modified exhibit in Chapter 4.4.

5.2.1 Seismology

The well-known fact that “[e]arthquake prediction research has been conducted for over 100 years with no obvious successes” (Geller 1997, 425) provide me an impetus to dive into this branch of study. No matter to address uncertainty in the choice of proper prediction models (Cooper and Nelson 1975; Nelson 1972) or to improve forecasting accuracy (Sanders 1963; Bates and Granger 1969), combination techniques were widely applied as a nature corresponding answer. **However, it’s until late 1990s, seismologists start to use combination techniques in papers**. Most of the review of seismology still ignores the developing combination techniques (Ben-Menahem 1995; Agnew 2002). Seismologists still mainly used the term ‘combination’ in physical meaning, such as the ‘combination of point forces’.

Without doubt, **the practical concern is the most important spirit that motivate seismologists to use combination techniques**. After all, the more precise we can avoid the destruction of earthquake, the more safety we enjoy. Combination techniques has applied in earthquake forecasting and also applied in the earthquake building engineering such as seismic response spectra and nonlinear dynamic analysis to reach reliable estimate.

Marzocchi et al. (2012) illustrated the uncertainty problem seismologists

have met with and clearly pointed out the practical value of merging models, saying that:

> [I]n practice, we never know which of the candidate models will be the best in a long testing phase. We also note that the best candidate model may capture one important part of the earthquake generation process well, while others might suitably represent secondary, or at least more subtle, features (2577).

The spirit and the response of the cited paragraph are just the same as that in economic forecasting. Both of them used the combination techniques in address of uncertainty problems (Winkler 1989, 608):

> In our uncertain and rapidly changing world, I think that adhering strictly to this [the development of 'true' model] ideal is counterproductive in most important forecasting situations. I prefer to view forecasts as information and the combining of forecasts as the aggregation of information. The key question is how best to accomplish this aggregation (606).

Regarding the uncertain problem, Neither Marzocchi et al. (2012) agree with the practice that "simply adopt the model that has performed best so far and disregard all others" (2577) nor do several scholars who do economic forecast. They all reason the advantage of forecast combination for some of its component models may outperform over the up-to-now best model in other period.

5.2.1.1 Earthquake forecasting

Combination techniques in earthquake forecasting had been happened relatively late. Fedotov et al. (1977) mentioned it while compared two earthquake statistics method. And just one sentence did he said about the combination techniques in the middle of his paper: "Thus, a combined use of various methods seems to be one of the hopeful ways of increasing efficiency of prediction" (320). No more in introduction and no more in conclusion as if nobody would care about his suggestion.

To my knowledge[6], the next time that seismologists came up with the idea of combination had to wait until 1990s. Sobolev et al. (1991) followed Fedotov et al. (1977) drawing compiling maps of expected earthquakes and both of them found positive forecasting ability in real time. And more pertinent to this paper, Sobolev et al. (1991)

[6] Though Sobolev et al. (1991) mentioned Aki (1981) as a forerunner of precursors combinations as well, I found that Aki (1981)'s work is more pertinent to an universal measurement that would be useful for unifying areas of earthquake prediction research.

discussed further the issue of combination. By using **Bayesian formula**, they combined the probability of expectation of a large earthquake with prognostic precursors. Up to 1989, three earthquakes occur within the areas of expected earthquakes but outside the center. But there were to areas indicated high possibility of earthquake but no strong earthquakes have been reported yet. Criticism though admitted that this paper pushed forward the concept of using combination techniques to improve forecasting ability, pointed out that the interdependence of the combined elements may be a potential problem (Shebalin et al. 2014).

Other than the Bayesian method, because "[i]t is well known that combining many models ……may yield higher performances than any individual member" (37), seismologists are working on their way to combine forecasts in their method. Shebalin et al. (2014) proposed a **rate combinations method**. It transforms different model outcome (for example: some measures in level, some measure in number) into one base and then multiplies the parameter derived with earthquake occurrence rate to get a new combined model.

After 2000s, more papers are aiming at investigating **combined short-term and long-term models** and correspondingly exhibited a 'suspected' close relation between earthquake forecasting and forecast combinations. I use the word 'suspected' because though no directly citation from Granger or other scholars in economic and business, the similarity in the tool and the spirit indicate a strong tight between the two subjects. Rhoades and Gerstenberger (2009) combined short-term earthquake probability (STEP) forecasting model with long-range earthquake forecasting model EEPAS in their paper. Each model typically based on time, density and location and a Poisson function to generate, a prediction, an earthquake occurrence rate. The authors combine the rate-based model by using the relative performance of the model as measurement to choose weights in the weighted average formula. Just as same as forecast combinations. Moreover, they also evaluate model performance by some conventional statistical tools, such as AIC. Technically, the difference between forecast combinations and their mixture models is just that seismologists tend to write down likelihood function and use simplex method, a popular algorithm for linear programming, to solve the optimization problem. Rhoades (2013) extended the study to a more wide ranging.

5.2.1.2 Seismic analysis

Seismic analysis is an earthquake building engineering and is a subset of structural analysis. A basic method of structural analysis methods is Seismic Response Spectra, a subset of Response-spectrum analysis (RSA) which measures the contribution from each natural mode of vibration to indicate the likely maximum seismic response of a structure. Therefore, the total response of a structure can be defined as a combination of natural mode of vibration. As a note, the term 'spectra' in the method name means that, for each mode, a response is read from the spectrum.

Seismic response spectra is the application of response spectra in earthquake engineering. By finding out the natural frequency of a structure and the peak response of that, the forces that a structure must be designed to resist can be derived. (Gupta 1992)

Regarding the complex response of buildings to earthquakes, more complex models were proposed, such as **nonlinear dynamic analysis**. The response of detailed structural model to ground motion was recorded and combined to lower the uncertainty of estimation. Similar to the case in weather forecast, the outcome forecast is sensitive to the initial condition. The response of buildings to earthquakes is very sensitive to the characteristics of the ground motion, the input, as well. As a consequence, a combination technique is needed to derive the final reliable estimate (Wilson et al. 1972).

Today, precise and timely earthquake prediction is still difficult if not impossible but scientists have already reflect on the uncertain property of the natural phenomenon. More and more seismic techniques that used combined skill have already put into use, not only for earthquake forecasting but also for understanding the earth structure deeply, such as the combination skills in seismic tomography (Valentine and Woodhouse 2010).

5.2.2 Ecology

The complexity of ecology has put biologists in a great debate about how to model it. In the beginning of 20th century, **Clements (1916) proposed the formation of a plant community as a complex and integrated organism**. Phillips (1931) approved with Clements view. And a politian coined the term 'holism' with a plethora explanation on the integrated biotic community (Lefkaditou and Stamou 2006). **In disagreement, Tansley (1935) advocated a mechanical interpretation on the ecology system called ecosystem**. Besides, in rejection of the downward causation proposed in holism, **Gleason (1926)** advocated for the idea that each community is interacted with its own fickle environment and therefore to study the temporary and fluctuating community, a population-centered view of ecology is encouraged.

The debate continued to 1960s. **Eugene Odum and Howard Odum adopted the view of holism with that of reductionism, which was noted as systems ecology**. While they are promoting the concept of an integrated community, they also stress on the physical interpretation of the ecosystem (Odum 1969). But Levins and Lewontin (1980) criticized them as just forming a large-scale computer models and "do not fit under the heading of 'holism' at all" (50). **On the other hand, MacArthur and Wilson (1967) held a Newtonian worldview promoting to form the knowledge of the complex ecosystem by investigating the basic components**. The debate has continued half a century; however, the debate is trivial in a sense. Because holism has never put their ideal in practice, there is actually only one modelling method in use, that is the method of reductionism (Lefkaditou and Stamou 2006).

Richard Levins' critiques toward system ecology made him seemingly properly to be categorized as a reductionists in the dichotomy; however, philosophically he held a holistic perspective (Levins 1974):

> The most difficult general problem of contemporary science is how to deal with complex systems as wholes. Most of the training of scientists, especially in the United States and Great Britain, is in the opposite direction. We are taught to isolate parts of a problem and to answer the question "What is this system?" by telling what it is made of (123).

But there was no proper holistic method. To Levins, modelling a one-to-one reflection of the complex ecology is difficult in practice if not impossible (Levins 1966). More importantly, he criticizes that system ecology is just a form of large-scale reductionism (Levins and Lewontin 1980). By addressing the problem that there is no proper holistic approach available, he brought breakthrough.

5.2.2.1 Model the complexity

To deal with complexity and "work with manageable models which maximize generality, realism, and precision" (Levins 1966, 422), **Levins proposed to build on several models that trade-off generality, realism, and precision**. Though he did not explicitly mention that information should be collected through model combinations but in the end of his paper he concluded that "a satisfactory theory is usually a cluster of models" (431). The underlying spirit is that building up a universally applicable model does not guarantee its validity but "generat[ing] good testable hypotheses relevant to important problems" (430) practically enables researchers to do validation test and only a cluster of verified models validates a theory.

In response to his 1966 work, **Levins then proposed 'loop analysis' to partially specify the system or, say, combine some of the models** (Levins 1974). Different from forecast combinations, loop analysis is a qualitative method and is known as 'qualitative model' in ecology also. Loop analysis is typically drawn as diagrams, called 'signed digraphs', with circles and lines that represent the relation between two variables. For example, if there is a direct positive effect of one variable upon another, the line will be ended in a pointed arrow; if a negative interaction, the line will end with a solid circle; otherwise, will be an absence of line. The three signed representation can be converted to three numbers -1, 0, 1 and form community matrix. The next step is usually using some matrix algebra functions to test the system or predict the behavior of system response to a disturbance. (Puccia and Levins 1991). This innovation approach is intended to correct the "one-sided analytical quantitative approach" then (Levins 1974, 137). The trial is rewarded. Because its simplicity, loop analysis is now accepted as a standard method in biology.

Levins does not put more emphasis on combined model after Levins (1974), but stresses more on the use of dialectic materialism to merge the gap between the mechanistic materialism and idealism. He encourages the exploration of the complex interrelationship in ecology by qualitative method, a dialectic way, but not numerically synthesizing results of models (Levins and Lewontin 1980; Levins 2006). His thought was not noticed in the 1960s and 1970s, but then a bunch of papers dealing with his strategy of modelling as well as his philosophical positions has appeared.

Chapter 6 The ideas of combination in psychology

Psychologists are interested in the property of group consensus so they combine individual's prediction in both subjective and mechanical ways. The latter approach finally transforms itself as one type of forecast combinations. Numerous findings were available in psychological studies but the volatile nature of the human-related science gave a piles of conflicting results. Psychologists are therefore renewed a technique of aggregation (combination) to reach consensus.

This chapter will still stress on the **motivation** of using combination ideas and **its evolution** or its **connection** to forecasts combinations.

6.1 Meta-analysis

Combined idea is used in Meta-analysis and the development of meta-analysis produces meta-regression which shares some similarity with forecast combination. And therefore a noteworthy branch. Meta-analysis is a procedure to establish guidelines for reliable reviews and integrate studies of similar research questions, designing to coordinate conflicting results produced by studies of social and behavioral sciences. For in these sciences, the research environment is hard to control and the subject's behavior is complex to explain, which result in enormous discordant test outcomes of the same hypothesis.

6.1.1 Psychotherapy can not help patients…

It was a raging debate in psychology that prompted the method of meta-analysis. In 1952, Hans Eysenck argued that psychotherapy can not help patients. Years passed by, array of positive, null or negative results have been produced but failed to resolve the debate. Smith and Glass (1977), however, provided an answer to it by standardizing and averaging the treatment-control differences of 375 psychotherapy. The debate did not subside immediately but the noun has been remained since then (Wilson and Lipsey 2001).

Some traced the origin of meta-analysis backed to statisticians' works in 1930s, such as Tippett ([1931] 1952), Fisher (1932), Pearson (1933) and Cochran (1937). Early studies was mainly originated in agricultural investigation and was extended to the research of meta-mathematics, meta-psychology and meta-evaluation. But no matter where the method of meta-analysis had taken root in, the fact that Glass (1976) reawakened the importance of combining various topics was agreed unanimously (Wolf 1986).

To address the problem of piles of conflicting studied in educational psychology, Glass coined the term meta-analysis. In his word, "[Meta-analysis] refers to the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings" (Glass 1976, 3). An early quantification method is vote counting but it suffers from statistical problems. Other than that, two approaches are influential in meta-analysis. (1) One combines research through composition of statistical significance of tests. While this would only produce a statistic to test the hypothesis, (2) the other one measures the magnitude of the experimental effect across studies, providing the information of strength of the interest. Despite the small differences in appearances, the various procedures used to code different forms of quantitative study findings are based on the concept of standardization.

The first one is known as **combined tests**. Methods of combined tests range from various simple counting procedures to a verity of summation techniques involving either logarithmic transformations of significance levels or t-test statistics or z-test statistics. Results of the procedures lead to a certain asymptotic distribution and thus enable the conduct of hypothesis test. To avoid confusing, I will list three of them below, according to Wolf (1986).

Fisher combined test (Fisher 1932),

$$\chi^2 = -2 \sum log_e\, p.$$

p is the one-tailed probability associated with each test. The degrees of freedom equal to two times the number of test.

Stouffer combined test (Stouffer et al. 1949),

$$Z_c = \frac{\sum z}{\sqrt{n}}.$$

z is the z-statistics associated with each test. n is the number of tests combined.

Similar to forecast combination, weighted Stouffer combined test have been developed (Mosteller and Bush 1954),

$$\text{weighted } Z_c = \frac{\sum df * z}{\sqrt{df^2}}.$$

The results of each combined test are basically consistent to each other.

Winer combined test (Winer 1971),

$$Z_c = \frac{\sum t}{\sqrt{\sum \frac{df}{df-2}}}.$$

t is the t-statistics associated with each test. df is the degrees of freedom associated with each test. And thus the ease of calculation can be regard as a prime consideration, depending on whether studies combined report p, t or z.

While the 'combined test' reports the summary index of the statistical significance of results relating to a hypothesis, 'measure of effect size' reveals further the degree of deviation from the null hypothesis to the alternative hypothesis. The former produces test statistics while the latter produces estimates.

The second one is known as measure of **effect size**. The phrase 'effect size' is meant to avoid the implication of causality and refer to the degree to which the null hypothesis is false. Two ways are available for measure including group difference and weighted. One method, '**group differences**', standardizes the difference of the control-experimental group mean.

$$d_c = \frac{\sum d}{n}, \qquad where \text{ d} = \frac{|\,\overline{x}_1 - \overline{x}_2\,|}{Sd}.$$

d_c can be interpreted either as a revelation of the degree of deviation from the null hypothesis to the alternative hypothesis. A thumb rule defines that d = 0.2 is small effect; 0.5 medium effect; 0.8 large effect (Cohen 1977). n is the number of tests combined. $\overline{x}_1$ and $\overline{x}_2$ is the sample mean of the control group and experimental group respectively. Sd is the sample standard deviation of either the control group or the experimental group. Alternatively, the pooled sample standard deviation can be used.

Weighted schemes for effect size of group differences have been developed (Hedges 1982),

$$d_c = \frac{\sum w * d}{\sum w}, \qquad where \text{ w} = \frac{2N}{8 + d^2}$$

N is the total sample size of both experimental and control groups (Rosenthal and Rubin 1982).

The other method, '**correlational relationships**' takes a simple average of the Pearson correlation coefficients (r) or the Fisher Z (Z_r) of each studies. The correlation measures the relation between dependent variable and control variable, functioning similarly as the coefficient in a regression model.

$$\bar{\mathrm{r}} = \frac{\sum r}{n},$$

$$\text{or} \quad \bar{Z}_r = \frac{\sum Z_r}{n}.$$

n is the number of correlation coefficients combined. Practically speaking the results of $\bar{\mathrm{r}}$ and $\bar{Z}_r$ are consistent to each other (Wolf 1986). Whatever method is employed, if null hypothesis is true, both the difference of two group means and the average of correlation are zero. Put differently, null hypothesis always indicates the effect size is zero, so the greater the result of either method, the higher the possibility of rejecting the null hypothesis. Cohen (1977) again provides a thumb rule that $\bar{\mathrm{r}}$ = 0.1 is small effect; 0.3 medium effect; 0.5 large effect. Researchers used thumb rule to interpret the results effect size is due to its lack of asymptotic distribution. Other than that, various way have been tried. Some researchers assume a normal distribution to construct confidence interval of group differences to examine whether the effect size of group differences encompasses zero or not. Some used a 'binominal effect size display (BESD)', estimating the differences in success rates between experimental group and control group, to interpret the effect size of correlation coefficients (Rosenthal and Rubin 1982). Overall, interpretation of the results of effect size is more difficult than that of combined test.

6.1.2 Compare with forecast combination

Both meta-analysis and forecast combination utilized the idea of combining; however, they do not share much similarity, except both of them suffer from the disadvantages of aggregating. Meta-analysis **are invented in persuasion for goals different from forecast combination**. Forecast combination pursues accuracy in a practical manner and addresses the inability problem of synthesizing model by combining the estimates (or forecasts) of models. Meta-analysis facilitates researchers to review or interpret literature from a numerical ground by combining statistics or standardized measurements.

Criticisms of meta-analysis on the disadvantages of aggregating are similar to that of forecast combination. Such as (1) if the input is dissimilar in subjects, no logical conclusion can be drawn; (2) if some of the input quality are poor, results are uninterpretable (Wilson and Lipsey 2001). To sum up, both combining procedures suffer from wiping out the individuality of each research and thus hinder respective comparisons and refinements.

Technically, they have several minor differences and similarities but I have mentioned before. To avoid repeats, Table 3 simply summarizes the comparison of meta-analysis and forecast combination. This table is simple but tells part of the story of the wide application of combined idea.

Table 3 Compariaon between meta-analysis and forecast combinations

	Combined test	Effect size	Forecast combinations
Goal	Provides numerical ground to review		Pursues accuracy in a practical manner and addresses the inability problem of synthesizing model
Input	Statistics (p, t or z)	Standardized measurements (r or Z_r)	Estimates or forecasts
Pooling way	Aggregation and divide by some special denominator	Simple or weighted averaging	
Output	Statistic (χ^2or Z)	Degree to which the H_0 is false	Estimate or forecast
Asymptotic distribution	Lead to certain asymptotic distribution	Still unknown	Follows an exact quadratic form (梁國源 and 高志祥 1994)
Defects	Wiping out the individuality		

6.1.3 No strong inter-development

Meta-regression analysis have been proposed to coordinate empirical research in economics by Stanley and Jarrell (1989). It was then widely applied to economic studies and developed complex theoretical form such as fixed effect meta-regression and random effects meta-regression. Despite the connection of meta-regression to economics and statistics, the connection between meta-analysis and forecast combination is weak.

As an heir of meta-analysis, Stanley and Jarrell (1989) proposed meta-regression aiming to harmonize the dissonance in empirical economic literature. The technique was more reliable than standard meta-analysis in a sense of less involvement in subjective judgment. As a consequence, **the technique is very popular in subjects that has been troubled by no unanimous evaluation procedures**, such as psychology (Winsper et al. 2013; Kline et al. 2013; Jackson and Dishman 2006; Haby et al. 2006; Cheng et al. 2007; Berger et al. 2012) and medicine tests (Valachis et al. 2009; Sgourakis et al. 2011; Riboh et al. 2013; Lesko et al. 2013; Garg et al. 2003; Garg et al. 2006; Florescu et al. 2012; Drewes et al. 2012; Chan et al. 2012; Briel et al. 2009; Bexkens et al. 2014). The function form is

$$b_j = \beta + \sum_k \alpha_k Z_{jk} + e_j \quad (j = 1, \dots L).$$

β is the parameter of interest. b_j is the estimation of β of the jth study. Z_{jk} systematically characterize the feature of each study, evaluating questions such as (Stanley and Jarrell 1989),

> whether the data was time series or cross-sectional,
>
> whether single equation or simultaneous systems were used,
>
> whether certain particularly important variables (e.g. lagged consumption) were included. (165)

α_k can be interpreted as average biases introduced by misspecification. Stanley and Jarrell (1989) claimed that meta-regression analysis is more objective than traditional review, for example the specification of meta-regression analysis model can rely on statistical specification test, preventing the reliance on subjective judgments.

Meta-regression analysis was widely applied to economic studies, such as union wage premiums (Jarrell and Stanley 1990); minimum wage with a stress on the objective characteristic of meta-regression analysis (Doucouliagos and Stanley 2009); the value of a statistical life with a more complex mixed effects regression model (Bellavance et al. 2009); cost-benefit analysis of a policy on the privatization issue (Bel et al. 2010). Numerous searching results can be found by keying in 'Meta-regression analysis' and 'Economics' on Google Scholar.

Not only be applied to the study of economics, **meta-analysis regression also took root in statistics and became more elaborated in model design**. In addition to the basic meta-regression analysis model, there are fixed-effect meta-regression and random effects meta-regression. The former one assumes the existence of a true effect size which is shared by all studies combined. The estimate of interest represents this common effect size. By contrast, under the random effects meta-regression model, one assumes the variation of true effect from study to study. The effect size in the study combined is just a sample of the relevant distribution of effects. And thus the estimate of interest represents the mean effect in this distribution (Borenstein et al. 2011).

Despite the strong connection of meta-regression to economics and statistics, **the connection between meta-analysis and forecast combination, to my knowledge, is weakly established by J. Scott Armstrong**. Armstrong specialized in forecasting methods and had studied both meta-analysis and forecast combination. He had tried, but not completed, to investigate forecast combination by using meta-analysis (Armstrong 1989). Other than that, the connection of meta-analysis and forecasting had built by him. He heavily used meta-analysis as evidence to evaluate forecasting methods (Armstrong 2006). Moreover, he suggested forecasters to use meta-analysis as a guide of forecasting (Armstrong 2007).

6.2 Mechanical combination of individual judgments

The review of mechanical combination of individual judgments was expanded in detail base on Clemen (1989). The topics of **subjective group consensus** are innate related to the nature of combination. Investigating them may bring abundant reward of appreciation and understanding the application of combination ideas, but for the limit of space, the topics will be skipped. All the articles reviewed here combine individual judgments by mechanical, or mathematical way. Therefore, this part is named as 'Mechanical combination of individual judgments'.

One of the branches of mechanical combination of individual judgments is "concocted group judgments" which is coined for referring the group judgments that is formed by calculation rather than by subjective discussion within a group.

6.2.1 Does beauty have a standard?

> Are beautiful things really beautiful, or do we only think them so, because they give us a certain kind of pleasurable feeling, feeling which we have been taught to call 'disinterested,' 'immediate,' 'universal,' etc.? (Gordon 1923, 36)

Psychologists are motivated to investigate concocted group consensus by their curiosity about the process of group consensus formation and its property.

Gordon adopted the following method: first, she formed concocted group orders. "For example a certain rug [the material used] would be given first place by one person, third place by another, twelfth place by another, etc. These numbers were added, and the rug which had the smallest resulting number was assigned first place in the composite order of merit. The rug which had the largest resulting number was called last […] in the composite order" (39). Second, she compared agreement of group with group. She sequentially divided total subjects into, for example, 20, 4 and 2 groups. The group order derived from each group were compared with other group orders. For example, if there were 200 subjects, Gordon may divided them into 4 groups and produce 4 concocted group orders. Comparing the group orders with each other produce 6 correlation coefficients.

Results shows group agree with group. For example, the average correlation coefficient is 0.76 if there is 20 persons a group; 0.87 if 50 persons; and 0.93 if 100 persons. The results is quite exciting. If one group's judgment can be accepted as standard, then we may randomly collect some people's judgments and form a group judgment that is quite similar to the standard.

The question is "[t]he group may agree but are they therefore right" (Gordon 1924, 398). Gordon was then inspired to investigate the accuracy problem by using weight as a standard. She asked subjects to assign orders on glass bottles from the lightest to the heaviest. Next, she formed the concocted groups of 5 persons, 10 persons, etc. each time and derived the composite order. The result was exciting, while there are only 5 persons in a group, the average coefficient of correlation between composite order and the true order is just 0.68. But while she put more members in a group, the composite order is approximate to the true order. The average coefficient of correlation for 10 persons is 0.79; for 20 persons is 0.86; and for 50 persons is 0.94. Gordon found that the estimated order gradually approximately to the true order as more individual's rankings are pooling. Notably, the average coefficient of correlation between individual's order and the true order is only 0.41, the results of the group are all superior to the results of the average single individuals. This may indicate the advantage of gathering information from people.

6.2.1.1 Critique to Gordon and to forecast combination

However, the upshot of Gordon's work was eventually shown to be a statistical result rather than psychological outcome in perspective of test theory. Stroop (1932) argued the belief on the good quality of group judgment was merely a statistical consequence of aggregating data. Gordon's experiment setting accidentally satisfied two required assumptions of Spearman-Brown formula. First, the standard deviations of each ranking are equal. Second, the correlations of all pairs of rankings are equal. The first requirement is satisfied because the components of each ranking is changelessly number one to ten. And because Gordon formed grouped ranking with random-picked individual rankings, by this system of average, the correlations of pairs of grouped rankings are approximately equal. And therefore what demonstrate by Gordon just match to the prediction that was produced by applying Spearman-Brown formula on Gordon's results (Kelley 1925). **Similarly, some argues the upshot of forecast combination is just a statistical result as well**. Manski (2011) noted that the loss of mean forecast shall certainly small or equal to the average loss of individual forecast by Jenson's inequality with convex loss function (Exhibit 10).

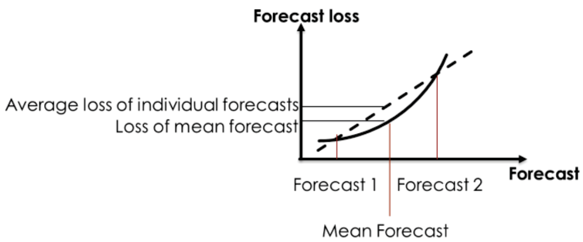


Exhibit 10 An illusion of the statistical upshot of forecast combination

6.2.1.2 Compare concocted group consensus with forecast combination

Gordon (1924)'s paper, to my knowledge, is the first paper focused on the accuracy of the combining estimations and thus a first paper employs similar concept in forecast combination in the field of psychology. But **the procedures are totally different from forecast combination**. First, concocted group consensus measures accuracy not by forecasting errors or other measurements used in forecast combination, it utilized Spearman's rank correlation coefficient as an evaluating tool,

$$\mathrm{r} = 1 - \frac{6\sum d_i^2}{n^3 - n},$$

where in the experiment, order of the boxes weights have to be assigned. d_i is the difference between the individual's order and the true order.

Second, concocted group consensus used a different way to combine estimates. For example, to make a concocted group including 5 persons, one have to (1) average the positions assigned to each weight by the first 5 persons and (2) rearrange the 'concocted' order, (3) continually repeating the concocted process itself to the next 5 persons till all members are assigned concocted group. The combining process is totally different to the simple average or weighted average used in forecast combination. An application of pooling idea independently developed by psychologists.

6.2.2 Combining multiple judgments

On the other front, psychologists manipulate regression techniques to combine clinical judgments. A brief review had already provided by Clemen (1989). Nowadays, the concern about human judgment in psychology has already became elements that can not be neglected in economic forecasting (Kuo and Liang 2004). To my knowledge, there was at least three papers published in International Journal of Forecasting, inheriting the tradition which has been initiated since Meehl (1954) and using forecast combinations as a main approach to explore their topics. **Fischer and Harvey (1999)** used techniques in group consensus giving feedback to judges to improve judges' performance. Results showed though the reception of the predicted outcome is helpful, judges' performance are still, in consistence with previous studies, inferior to that of simple averaging. Provision of mean absolute percentage errors which was updated each period to forecasters help further their performance but inconsistent is still a fatal wound of human being. More recently, **Jørgensen (2007)** reviewed empirical studies comparing the accuracy of expert judgments, model and the combination of the two. Among all sixteen relevant studies, ten of them indicates a higher average accuracy of expert judgments over that of model. The paper came with **(Hogarth (2007))**'s comment in the same periodical as well.

Chapter 7 Uncertainty and complexity

One of the econometricians' principle beliefs is that better understanding leads to better forecasts. But most of time, researchers are uncertain about what modelling strategies should be chose, what variables should be contain, and how to combine all the information to model the reality as real as possible. Moreover, combining information is usually time-costly and expense-costly. Rather than that, forecast combination combines forecasts directly and therefore save in time and expense. In addition, by combining forecasts, the technique indirectly combines the underlying information sets in each forecast figure. And thus the method takes advantage in broadening its information set, spreading risks of using certain modelling strategy, which are helpful to address the uncertain problem.

In address of the uncertainty in the market, portfolio selection diversifies portfolio (combines different assets) to reduce investment risks. By minimizing risks of a portfolio, investors assigns weights to each asset and form the optimal portfolio. Forecast encompassing method, similar to portfolio selection in mathematical sense, minimizes forecast error of the combination formula and produces coefficients which measure the informational contribution of each model.

In face of the chaos nature in weather, ensemble forecast method is proposed to simulate possible outcomes to give a more accurate forecast. On the other front, while a better modelling strategy is unsure, the technique of combining short-term and long-term earthquake forecast models is used.

Human behavior is complex and therefore the results of psychological studies are usually conflicting. To reach an agreeable result of certain type of psychological studies, meta-analysis method is suggested. Loop analysis method gathers a cluster of models which have different ecological concerns to facilitate researchers to partially model the complex ecology system and form a satisfactory theory.

Almost all the fields covered in this thesis, except group consensus, are in face of either uncertainty problem or complex difficult. The combination techniques bring us practically acceptable results.

Reference

Agnew, C. E. 'Bayesian Consensus Forecasts of Macroeconomic Variables'. *Journal of Forecasting* 4, no. 4 (1985): 363-76.
Agnew, D. C. 'History of Seismology'. *International handbook of earthquake and engineering seismology* 81 (2002): 3-11.
Aki, K. 'A Probabilistic Synthesis of Precursory Phenomena'. *Earthquake Prediction* (1981): 566-74.
Anandalingam, G., and L. Chen. 'Linear Combination of Forecasts: A General Bayesian Model'. *Journal of Forecasting* 8, no. 3 (1989): 199-214.
Armstrong, J. S. 'Combining Forecasts: The End of the Beginning or the Beginning of the End?'. *International Journal of Forecasting* 5, no. 4 (1989): 585-88.
———. 'Findings from Evidence-Based Forecasting: Methods for Reducing Forecast Error'. *International Journal of Forecasting* 22, no. 3 (2006): 583-98.
———. 'Significance Tests Harm Progress in Forecasting'. *International Journal of Forecasting* 23, no. 2 (2007): 321-27.
Barnard, G. 'New Methods of Quality Control'. *Journal of the Royal Statistical Society. Series A (General)* (1963): 255-58.
Bates, J. M., and C. W. J. Granger. 'Combination of Forecasts'. *Operational Research Quarterly* 20, no. 4 (1969): 451-68.
Bel, G., X. Fageda, and M. E. Warner. 'Is Private Production of Public Services Cheaper Than Public Production? A Meta-Regression Analysis of Solid Waste and Water Services'. *Journal of Policy Analysis and Management* 29, no. 3 (2010): 553-77.
Bellavance, F., G. Dionne, and M. Lebeau. 'The Value of a Statistical Life: A Meta-Analysis with a Mixed Effects Regression Model'. *Journal of Health Economics* 28, no. 2 (2009): 444-64.
Ben-Menahem, A. 'A Concise History of Mainstream Seismology: Origins, Legacy, and Perspectives'. *Bulletin of the Seismological Society of America* 85, no. 4 (1995): 1202-25.
Berger, W., E. Coutinho, I. Figueira, C. Marques-Portella, M. Luz, T. Neylan, C. Marmar, and M. Mendlowicz. 'Rescuers at Risk: A Systematic Review and Meta-Regression Analysis of the Worldwide Current Prevalence and Correlates of Ptsd in Rescue Workers'. *Social Psychiatry & Psychiatric Epidemiology* 47, no. 6 (2012): 1001-11.
Bexkens, A., L. Ruzzano, A. M. L. Collot d' Escury-Koenigs, M. W. Van der Molen, and H. M. Huizenga. 'Inhibition Deficits in Individuals with Intellectual Disability: A Meta-Regression Analysis'. *Journal of Intellectual Disability Research* 58, no. 1 (2014): 3-16.
Bjerknes, V. *Das Problem Der Wettervorhersage: Betrachtet Vom Standpunkte Der Mechanik Und Der Physik*, 1904.
Board, N. I. C. *Forecasting Sales*. New York, 1963.
Bordley, R. F. 'The Combination of Forecasts: A Bayesian Approach'. *Journal of the operational research society* (1982): 171-74.
———. 'Linear Combination of Forecasts with an Intercept: A Bayesian Approach'. *Journal of Forecasting* 5, no. 4 (1986): 243-49.
Borenstein, M., L. V. Hedges, J. P. Higgins, and H. R. Rothstein. *Introduction to Meta-Analysis*: John Wiley & Sons, 2011.
Bosart, L. F. 'Sunya Experimental Results in Forecasting Daily Temperature and Precipitation'. *Monthly Weather Review* 103, no. 11 (1975): 1013-20.
Briel, M., I. Ferreira-Gonzalez, J. J. You, P. J. Karanicolas, E. A. Akl, W. Ping, B. Blechacz, D. Bassler, W. Xinge, A. Sharman, I. Whitt, S. A. da Silva, Z. Khalid, A. J. Nordmann, Z. Qi, S. D. Walter, N. Vale, N. Bhatnagar, C. O'Regan, and E. J. Mills. 'Association between Change in High Density Lipoprotein Cholesterol and Cardiovascular Disease Morbidity and Mortality: Systematic Review and Meta-Regression Analysis'. *BMJ: British Medical Journal (Overseas & Retired Doctors Edition)* (2009): 522-26.

Bunn, D. W. 'A Bayesian Approach to the Linear Combination of Forecasts'. *Operational Research Quarterly* (1975): 325-29.

———. 'A Comparative Evaluation of the Outperformance and Minimum Variance Procedures for the Linear Synthesis of Forecasts'. *Operational Research Quarterly* (1977): 653-62.

———. 'Forecasting with More Than One Model'. *Journal of Forecasting* 8, no. 3 (1989): 161-66.

———. 'Statistical Efficiency in the Linear Combination of Forecasts'. *International Journal of Forecasting* 1, no. 2 (1985): 151-63.

Chan, R., R. Brooks, Z. Steel, T. Heung, J. Erlich, J. Chow, and M. Suranyi. 'The Psychosocial Correlates of Quality of Life in the Dialysis Population: A Systematic Review and Meta-Regression Analysis'. *Quality of Life Research* 21, no. 4 (2012): 563-80.

Charney, J. G. 'Dynamic Forecasting by Numerical Process'. *Compendium of Meteorology, American Meteorological Society, Boston* (1951): 470-82.

Cheng, J. Y. W., R. Y. L. Chen, J. S. N. Ko, and E. M. L. Ng. 'Efficacy and Safety of Atomoxetine for Attention-Deficit/Hyperactivity Disorder in Children and Adolescents—Meta-Analysis and Meta-Regression Analysis'. *Psychopharmacology* 194, no. 2 (2007): 197-209.

Chong, Y. Y., and D. F. Hendry. 'Econometric Evaluation of Linear Macro-Economic Models'. *The Review of Economic Studies* 53, no. 4 (1986): 671-90.

Clemen, R. T. 'Combining Forecasts: A Review and Annotated Bibliography'. *International journal of forecasting* 5, no. 4 (1989): 559-83.

———. 'Extraneous Expert Information'. *Journal of Forecasting* 4, no. 4 (1985): 329-48.

Clemen, R. T., and A. H. Murphy. 'Objective and Subjective Precipitation Probability Forecasts: Some Methods for Improving Forecast Quality'. *Weather and Forecasting* 1, no. 3 (1986): 213-18.

———. 'Objective and Subjective Precipitation Probability Forecasts: Statistical Analysis of Some Interrelationships'. *Weather and Forecasting* 1, no. 1 (1986): 56-65.

Clemen, R. T., and R. L. Winkler. 'Combining Probability Distributions from Experts in Risk Analysis'. *Risk analysis* 19, no. 2 (1999): 187-203.

Clements, F. E. *Plant Succession: An Analysis of the Development of Vegetation*: Carnegie Institution of Washington, 1916.

Clements, M. P., and D. I. Harvey. 'Combining Probability Forecasts'. *International Journal of Forecasting* 27, no. 2 (2011): 208-23.

———. 'Forecast Encompassing Tests and Probability Forecasts'. *Journal of Applied Econometrics* 25, no. 6 (2010): 1028-62.

Cochran, W. G. 'Problems Arising in the Analysis of a Series of Similar Experiments'. *Supplement to the Journal of the Royal Statistical Society* (1937): 102-18.

Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*: Lawrence Erlbaum Associates, Inc, 1977.

Cooper, J. P., and C. R. Nelson. 'The Ex Ante Prediction Performance of the St. Louis and Frb-Mit-Penn Econometric Models and Some Results on Composite Predictors'. *Journal of Money, Credit and Banking* (1975): 1-32.

Crane, D. B., and J. R. Crotty. 'A Two-Stage Forecasting Model: Exponential Smoothing and Multiple Regression'. *Management Science* 13, no. 8 (1967): B-501-B-07.

Dickinson, J. P. 'Some Comments on the Combination of Forecasts'. *Operational Research Quarterly* 26, no. 1 ii (1975): 205-10.

———. 'Some Statistical Results in the Combination of Forecasts'. *Operational Research Quarterly* 24, no. 2 (1973): 253-60.

Diebold, F. X. *Elements of Forecasting*. Mason, Ohio: Thomson/South-Western, 2007.

Diebold, F. X. 'Forecast Combination and Encompassing: Reconciling Two Divergent Literatures'. *International Journal of Forecasting* 5, no. 4 (1989): 589-92.

Doucouliagos, H., and T. D. Stanley. 'Publication Selection Bias in Minimum-Wage Research? A Meta-Regression Analysis'. *British Journal of Industrial Relations* 47, no. 2 (2009): 406-28.

Drewes, H. W., L. M. G. Steuten, L. C. Lemmens, C. A. Baan, H. C. Boshuizen, A. M. J. Elissen, K. M. M. Lemmens, J. A. C. Meeuwissen, and H. J. M. Vrijhoef. 'The Effectiveness of Chronic Care Management for Heart Failure: Meta-Regression Analyses to Explain the Heterogeneity in Outcomes'. *Health Services Research* 47, no. 5 (2012): 1926-59.
Edgerton, H. A., and L. E. Kolbe. 'The Method of Minimum Variation for the Combination of Criteria'. *Psychometrika* 1, no. 3 (1936): 183-87.
'Edward Lorenz'. http://www.britannica.com/.
Elliott, G. 'Sir Clive W. J. Granger (1934–2009)'. *International Journal of Forecasting* 25, no. 4 (2009): 639-41.
Engle, R. F., C. W. Granger, and J. J. Hallman. 'Merging Short-and Long-Run Forecasts: An Application of Seasonal Cointegration to Monthly Electricity Sales Forecasting'. *Journal of Econometrics* 40, no. 1 (1989): 45-62.
Engle, R. F., C. W. Granger, and D. Kraft. 'Combining Competing Forecasts of Inflation Using a Bivariate Arch Model'. *Journal of economic dynamics and control* 8, no. 2 (1984): 151-65.
'Ensemble Forecasting'. http://www.metoffice.gov.uk/research/areas/data-assimilation-and-nsembles/ensemble-forecasting/explanation.
EPS, C. E. T. o. 'Guidelines on Using Information from Eps in Combination with Single Higher Resolution Nwp Forecasts'. (2006).
Epstein, E. S. 'Stochastic Dynamic Prediction'. *Tellus A* 21, no. 6 (1969).
Evans, M. K. *Macroeconomic Activity : Theory, Forecasting, and Control; an Econometric Approach*. New York: Harper & Row, 1969.
Fang, Y. 'Forecasting Combination and Encompassing Tests'. *International Journal of Forecasting* 19, no. 1 (2003): 87-94.
Fedotov, S., G. Sobolev, S. Boldyrev, A. Gusev, A. Kondratenko, O. Potapova, L. Slavina, V. Theophylaktov, A. Khramov, and V. Shirokov. 'Long-and Short-Term Earthquake Prediction in Kamchatka'. *Tectonophysics* 37, no. 4 (1977): 305-21.
Figlewski, S. 'Optimal Price Forecasting Using Survey Data'. *The Review of Economics and Statistics* (1983): 13-21.
Fischer, I., and N. Harvey. 'Combining Forecasts: What Information Do Judges Need to Outperform the Simple Average?'. *International journal of forecasting* 15, no. 3 (1999): 227-46.
Fisher, R. *Statistical Methods for Research Workers*, 1932.
FitzRoy, R. *The Weather Book*. London, 1863.
Florescu, D. F., F. Qiu, M. A. McCartan, C. Mindru, P. D. Fey, and A. C. Kalil. 'What Is the Efficacy and Safety of Colistin for the Treatment of Ventilator-Associated Pneumonia? A Systematic Review and Meta-Regression'. *Clinical Infectious Diseases* 54, no. 5 (2012): 670-80.
Fritsch, J., J. Hilliker, J. Ross, and R. Vislocky. 'Model Consensus'. *Weather and forecasting* 15, no. 5 (2000): 571-82.
Garg, A. X., N. Muirhead, G. Knoll, R. C. Yang, G. V. R. Prasad, H. Thiessen-Philbrook, M. P. Rosas-Arellano, A. Housawi, and N. Boudville. 'Proteinuria and Reduced Kidney Function in Living Kidney Donors: A Systematic Review, Meta-Analysis, and Meta-Regression'. *Kidney International* 70, no. 10 (2006): 1801-10.
Garg, A. X., R. S. Suri, N. Barrowman, F. Rehman, D. Matsell, M. P. Rosas-Arellano, M. Salvadori, R. B. Haynes, and W. F. Clark. 'Long-Term Renal Prognosis of Diarrhea-Associated Hemolytic Uremic Syndrome: A Systematic Review, Meta-Analysis, and Meta-Regression'. *JAMA: Journal of the American Medical Association* 290, no. 10 (2003): 1360-70.
Geller, R. J. 'Earthquake Prediction: A Critical Review'. *Geophysical Journal International* 131, no. 3 (1997): 425-50.

Glass, G. V. 'Primary, Secondary, and Meta-Analysis of Research'. *Educational researcher* (1976): 3-8.

Gleason, H. A. 'The Individualistic Concept of the Plant Association'. *Bulletin of the Torrey Botanical Club* (1926): 7-26.

Golding, B., K. Mylne, and P. Clark. 'The History and Future of Numerical Weather Prediction in the Met Office'. *Weather* 59, no. 11 (2004): 299-306.

Gordon, K. 'Group Judgments in the Field of Lifted Weights'. *Journal of Experimental Psychology* 7, no. 5 (1924): 398-400.

———. 'A Study of Esthetic Judgments'. *Journal of Experimental Psychology* 6, no. 1 (1923): 36-43.

Granger, C. W. *Essays in Econometrics: Collected Papers of Clive Wj Granger*. Vol. 1: Cambridge University Press, 2001.

———. 'Invited Review Combining Forecasts—Twenty Years Later'. *Journal of Forecasting* 8, no. 3 (1989): 167-73.

Granger, C. W., and R. Ramanathan. 'Improved Methods of Combining Forecasts'. *Journal of Forecasting* 3, no. 2 (1984): 197-204.

Granger, C. W., H. White, and M. Kamstra. 'Interval Forecasting: An Analysis Based Upon Arch-Quantile Estimators'. *Journal of Econometrics* 40, no. 1 (1989): 87-96.

Gulledge Jr, T. R., J. L. Ringuest, and J. A. Richardson. 'Subjective Evaluation of Composite Econometric Policy Inputs'. *Socio-Economic Planning Sciences* 20, no. 1 (1986): 51-55.

Gupta, A. K. *Response Spectrum Method in Seismic Analysis and Design of Structures*. Vol. 4: CRC press, 1992.

Haby, M. M., M. Donnelly, J. Corry, and T. Vos. 'Cognitive Behavioural Therapy for Depression, Panic Disorder and Generalized Anxiety Disorder: A Meta-Regression of Factors That May Predict Outcome'. *Australian & New Zealand Journal of Psychiatry* 40, no. 1 (2006): 9-19.

Hall, S. G., and J. Mitchell. 'Combining Density Forecasts'. *International Journal of Forecasting* 23, no. 1 (2007): 1-13.

Hallman, J., and M. Kamstra. 'Combining Algorithms Based on Robust Estimation Techniques and Co-Integrating Restrictions'. *Journal of Forecasting* 8, no. 3 (1989): 189-98.

Halperin, M. 'Almost Linearly-Optimum Combination of Unbiased Estimates'. *Journal of the American Statistical Association* 56, no. 293 (1961): 36-43.

Hedges, L. V. 'Estimation of Effect Size from a Series of Independent Experiments'. *Psychological bulletin* 92, no. 2 (1982): 490-99.

Hibon, M., and T. Evgeniou. 'To Combine or Not to Combine: Selecting among Forecasts and Their Combinations'. *International Journal of Forecasting* 21, no. 1 (2005): 15-24.

Hogarth, R. M. 'Information Asymmetry and Aggregation Rules: A Comment on Jørgensen (2007)'. *International Journal of Forecasting* 23, no. 3 (2007): 465-67.

Holden, K., and D. Peel. 'Unbiasedness, Efficiency and the Combination of Economic Forecasts'. *Journal of Forecasting* 8, no. 3 (1989): 175-88.

Holden, K., D. Peel, and J. Thompson. *Expectations: Theory and Evidence*: Macmillan, 1985.

Horst, P. 'Obtaining a Composite Measure from a Number of Different Measures of the Same Attribute'. *Psychometrika* 1, no. 1 (1936): 53-60.

Jackson, E. M., and R. K. Dishman. 'Cardiorespiratory Fitness and Laboratory Stress: A Meta-Regression Analysis'. *Psychophysiology* 43, no. 1 (2006): 57-72.

Jarrell, S. B., and T. D. Stanley. 'A Meta-Analysis of the Union-Nonunion Wage Gap'. *Industrial and Labor Relations Review* (1990): 54-67.

Jørgensen, M. 'Forecasting of Software Development Work Effort: Evidence on Expert Judgement and Formal Models'. *International Journal of Forecasting* 23, no. 3 (2007): 449-62.

Kalnay, E., S. J. Lord, and R. D. McPherson. 'Maturity of Operational Numerical Weather Prediction: Medium Range'. *Bulletin of the American Meteorological Society* 79, no. 12 (1998): 2753-69.

Kelley, T. L. 'The Applicability of the Spearman-Brown Formula for the Measurement of Reliability'. *Journal of Educational Psychology* 16, no. 5 (1925): 300-03.
Kline, E., J. M. Gold, and J. Schiffman. 'Response to 'a Systematic Review and Meta-Regression Analysis of Aggression During First Episode of Psychosis''. *Acta Psychiatrica Scandinavica* 128, no. 6 (2013): 492-92.
Kuo, Y., and K. Liang. 'Human Judgments in New York State Sales and Use Tax Forecasting'. *Journal of Forecasting* 23, no. 4 (2004): 297-314.
Lawrence, K. D., and G. R. Reeves. 'Consensus Time Series Forecasting'. In *Organizations: Multiple Agents with Multiple Criteria*, 199-204: Springer, 1981.
Lefkaditou, A., and G. P. Stamou. 'Holism and Reductionism in Ecology: A Trivial Dichotomy and Levins' Non-Trivial Account'. *History and Philosophy of the Life Sciences* (2006): 313-36.
Lesko, C. R., S. R. Cole, A. Zinski, C. Poole, and M. J. Mugavero. 'A Systematic Review and Meta-Regression of Temporal Trends in Adult Cd4+ Cell Count at Presentation to Hiv Care, 1992–2011'. *Clinical Infectious Diseases* 57, no. 7 (2013): 1027-37.
Levins, R. 'Discussion Paper: The Qualitative Analysis of Partially Specified Systems'. *Annals of the New York Academy of Sciences* 231, no. 1 (1974): 123-38.
———. 'Strategies of Abstraction'. *Biology and Philosophy* 21, no. 5 (2006): 741-55.
———. 'The Strategy of Model Building in Population Biology'. *American scientist* (1966): 421-31.
Levins, R., and R. Lewontin. 'Dialectics and Reductionism in Ecology'. *Synthese* 43, no. 1 (1980): 47-78.
Liang, K. 'On the Sign of the Optimal Combining Weights under the Error-Variance Minimizing Criterion'. *Journal of Forecasting* 11, no. 8 (1992): 719-23.
Liang, K., J. C. Lee, and K. S. Shao. 'On the Distribution of the Inverted Linear Compound of Dependent F-Variates and Its Application to the Combination of Forecasts'. *Journal of Applied Statistics* 33, no. 9 (2006): 961-73.
Liang, K., and K. Ryu. 'Relationship of Forecast Encompassing to Composite Forecasts with Simulations and an Application'. *Seoul Journal of Economics* 16, no. 3 (2003): 363-86.
Liang, K., and Y. Shih. 'Bayesian Composite Forecasts : An Extension and a Clarification'. *Journal of Quantitative Economics* 10, no. 1 (1994): 105-22.
'Loop Analysis'. http://ipmnet.org/loop/default.aspx.
Lorenz, E. N. 'Deterministic Nonperiodic Flow'. *Journal of the Atmospheric Sciences* 20, no. 2 (1963): 130-41.
MacArthur, R. H., and E. O. Wilson. *The Theory of Island Biogeography*. Vol. 1: Princeton University Press, 1967.
Makridakis, S., A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, and R. Winkler. 'The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition'. *Journal of forecasting* 1, no. 2 (1982): 111-53.
Makridakis, S., and R. L. Winkler. 'Averages of Forecasts: Some Empirical Results'. *Management Science* 29, no. 9 (1983): 987-96.
Manski, C. F. 'Interpreting and Combining Heterogeneous Survey Forecasts'. In *Oxford Handbook on Economic Forecasting*, edited by M. C. a. D. Hendry, 457-72, 2011.
Markowitz, H. 'Portfolio Selection*'. *The journal of finance* 7, no. 1 (1952): 77-91.
Markowitz, H. M. 'The Early History of Portfolio Theory: 1600-1960'. *Financial Analysts Journal* (1999): 5-16.
Marzocchi, W., J. D. Zechar, and T. H. Jordan. 'Bayesian Forecast Evaluation and Ensemble Earthquake Forecasting'. *Bulletin of the Seismological Society of America* 102, no. 6 (2012): 2574-84.
Meehl, P. E. 'Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence'. (1954).
Mizon, G. E. *The Encompassing Approach in Econometrics*: Australian National University, Faculty of Economics and Research School of Social Sciences, 1984.

Morris, P. A. 'Combining Expert Judgments: A Bayesian Approach'. *Management Science* 23, no. 7 (1977): 679-93.
———. 'Decision Analysis Expert Use'. *Management Science* 20, no. 9 (1974): 1233-41.
Mosteller, F. M., and R. R. Bush. 'Selected Quantitative Techniques'. In *Handbook of Social Psychology*, edited by G. Lindzey. Cambridge, MA: Addison-Wesley, 1954.
Nelson, C. R. 'The Prediction Performance of the Frb-Mit-Penn Model of the Us Economy'. *The American Economic Review* (1972): 902-17.
Newbold, P., and C. W. Granger. 'Experience with Forecasting Univariate Time Series and the Combination of Forecasts'. *Journal of the Royal Statistical Society. Series A (General)* (1974): 131-65.
Odum, E. P. 'The Strategy of Ecosystem Development'. *Science* 164 (1969): 262-70.
Öller, L.-E. 'A Method for Pooling Forecasts'. *Journal of the Operational Research Society* (1978): 55-63.
Organization, W. M. *Guidelines on Ensemble Prediction Systems and Forecasting*, World Meteorological Organization, 2012.
Pearson, K. 'On a Method of Determining Whether a Sample of Size N Supposed to Have Been Drawn from a Parent Population Having a Known Probability Integral Has Probably Been Drawn at Random'. *Biometrika* (1933): 379-410.
Phillips, J. 'The Biotic Community'. *The Journal of Ecology* (1931): 1-24.
Puccia, C. J., and R. Levins. 'Qualitative Modeling in Ecology: Loop Analysis, Signed Digraphs, and Time Averaging'. In *Qualitative Simulation Modeling and Analysis*, 119-43: Springer, 1991.
Reeves, G. R., and K. D. Lawrence. 'Combining Multiple Forecasts Given Multiple Objectives'. *Journal of Forecasting* 1, no. 3 (1982): 271-79.
Rhoades, D. A. 'Mixture Models for Improved Earthquake Forecasting with Short-to-Medium Time Horizons'. *Bulletin of the Seismological Society of America* 103, no. 4 (2013): 2203-15.
Rhoades, D. A., and M. C. Gerstenberger. 'Mixture Models for Improved Short-Term Earthquake Forecasting'. *Bulletin of the Seismological Society of America* 99, no. 2A (2009): 636-46.
Riboh, J. C., V. Hasselblad, J. A. Godin, and R. C. Mather. 'Transtibial Versus Independent Drilling Techniques for Anterior Cruciate Ligament Reconstruction: A Systematic Review, Meta-Analysis, and Meta-Regression'. *American Journal of Sports Medicine* 41, no. 11 (2013): 2693-702.
Richardson, L. F. *Weather Prediction by Numerical Process*: Cambridge University Press, 2007.
Rosenthal, R., and D. B. Rubin. 'Further Meta-Analytic Procedures for Assessing Cognitive Gender Differences'. *Journal of Educational Psychology* 74, no. 5 (1982): 708-12.
———. 'A Simple, General Purpose Display of Magnitude of Experimental Effect'. *Journal of educational psychology* 74, no. 2 (1982): 166-69.
Sanders, F. 'On Subjective Probability Forecasting'. *Journal of Applied Meteorology* 2, no. 2 (1963): 191-201.
Sgourakis, G., S. Lanitis, I. Gockel, C. Kontovounisios, C. Karaliotas, K. Tsiftsi, A. Tsiamis, and C. C. Karaliotas. 'Transanal Endoscopic Microsurgery for T1 and T2 Rectal Cancers: A Meta-Analysis and Meta-Regression Analysis of Outcomes'. *American Surgeon* 77, no. 6 (2011): 761-72.
Shebalin, P. N., C. Narteau, J. D. Zechar, and M. Holschneider. 'Combining Earthquake Forecasts Using Differential Probability Gains'. *Earth, Planets and Space* 66, no. 1 (2014): 37-50.
Smith, M. L., and G. V. Glass. 'Meta-Analysis of Psychotherapy Outcome Studies'. *American psychologist* 32, no. 9 (1977): 752-60.
Sobolev, G., T. Chelidze, A. Zavyalov, L. Slavina, and V. Nikoladze. 'Maps of Expected Earthquakes Based on a Combination of Parameters'. *Tectonophysics* 193, no. 4 (1991): 255-65.

Stanley, T. D., and S. B. Jarrell. 'Meta-Regression Analysis: A Quantitative Method of Literature Surveys'. *Journal of Economic Surveys* 3, no. 2 (1989): 161-70.
Stone, M. 'The Opinion Pool'. *The Annals of Mathematical Statistics* 32, no. 4 (1961): 1339-42.
Stouffer, S. A., E. Suchman, L. DeVinney, S. Star, and R. Williams Jr. 'Tha American Soldier: Adjustment During Amy Life (Vol. 1). Princetom'. NJ: Princeton University Press, 1949.
Stroop, J. R. 'Is the Judgment of the Group Better Than That of the Average Member of the Group?'. *Journal of experimental Psychology* 15, no. 5 (1932): 550-62.
Tansley, A. G. 'The Use and Abuse of Vegetational Concepts and Terms'. *Ecology* 16, no. 3 (1935): 284-307.
Teräsvirta, T. 'Professor Clive W.J. Granger: An Interview for the International Journal of Forecasting'. *International Journal of Forecasting* 11, no. 4 (1995): 585-90.
Timmermann, A. 'Chapter 4 Forecast Combinations'. In *Handbook of Economic Forecasting*, edited by C. W. J. G. G. Elliott and A. Timmermann, 135-96: Elsevier, 2006.
Tippett, L. H. C. *The Methods of Statistics*. London: Williams & Norgate Ltd., [1931] 1952.
Valachis, A., D. Mauri, V. Karampoiki, N. P. Polyzos, I. Cortinovis, G. Koukourakis, G. Zacharias, A. Xilomenos, M. Tsappi, and G. Casazza. 'Time-Trend of Melanoma Screening Practice by Primary Care Physicians: A Meta-Regression Analysis'. *Upsala Journal of Medical Sciences* 114, no. 1 (2009): 32-40.
Valentine, A. P., and J. H. Woodhouse. 'Reducing Errors in Seismic Tomography: Combined Inversion for Sources and Structure'. *Geophysical Journal International* 180, no. 2 (2010): 847-57.
Wall, K. D., and C. Correia. 'A Preference-Based Method for Forecast Combination'. *Journal of Forecasting* 8, no. 3 (1989): 269-92.
Wallis, K. F. 'Combining Forecasts–Forty Years Later'. *Applied Financial Economics* 21, no. 1-2 (2011): 33-41.
Wilson, D. B., and M. Lipsey. *Practical Meta-Analysis*: Sage Publications, 2001.
Wilson, E., I. Farhoomand, and K. Bathe. 'Nonlinear Dynamic Analysis of Complex Structures'. *Earthquake Engineering & Structural Dynamics* 1, no. 3 (1972): 241-52.
Winer, B. *Statistical Principles in Experimental Design: 2d Ed*: McGraw-Hill, 1971.
Winkler, R. L. 'Combining Forecasts: A Philosophical Basis and Some Current Issues'. *International Journal of Forecasting* 5, no. 4 (1989): 605-09.
———. 'Combining Probability Distributions from Dependent Information Sources'. *Management Science* 27, no. 4 (1981): 479-88.
———. 'The Consensus of Subjective Probability Distributions'. *Management Science* 15, no. 2 (1968): B-61-B-75.
Winkler, R. L., and A. Murphy. 'Reliability of Subjective Probability Forecasts of Precipitation and Temperature'. *Applied Statistics* (1977): 41-47.
Winsper, C., R. Ganapathy, S. Marwaha, M. Large, M. Birchwood, and S. P. Singh. 'A Systematic Review and Meta-Regression Analysis of Aggression During the First Episode of Psychosis'. *Acta Psychiatrica Scandinavica* 128, no. 6 (2013): 413-21.
Wolf, F. M. *Meta-Analysis: Quantitative Methods for Research Synthesis*. Vol. 59: Sage, 1986.
梁國源、高志祥（1994）〈隨機性迴歸子下之預測組合理論〉，《中國統計學報》，32:3，367-388。