STAT 626: Outline of Lecture 11

Forecasting (§3.5)

1. **Best Linear Prediction for Stationary Processes**

2. **Forecasting an AR(2) Model: Finite Past**

3. **Durbin-Levinson Algorithm: PACF**

4. **Prediction, Error Variance (P_{n+1}^n) and Forecast Interval:**

$$x_{n+1}^n \pm 1.96\sqrt{P_{n+1}^n}$$

5. Prediction in the age of "Big Data".

Who do you think said the following:?

"It is difficult to make predictions, especially about the future"

Fun Reading

6. Graham Southorn (2016). **Great expectations: The past, present and future of prediction.** *Significance*, April Issue.

7. Philip Tetlock and Dan Gardner (2015). **Superforcasting: The Art and Science of Prediction.**

8. Michael Abramowicz (2008). **Predictocracy: Market Mechanisms for Public and Private Decision Making.**

   Predicting the future is serious business for virtually all public and private institutions, for they must often make important decisions based on such predictions. This visionary book explores how institutions from legislatures to corporations might improve their predictions and arrive at better decisions by means of prediction markets, a promising new tool with virtually unlimited potential applications.

9. James Surowiecki (2005). **The Wisdom of Crowds: Why the Many are Smarter than the Few**

FORECASTING

How is statistical or scientific forecast different from that of a **fortune teller or psychic?**

Given the time series data $x_1, \ldots, x_n$: What is a good way to forecast the next future value x_{n+1}?

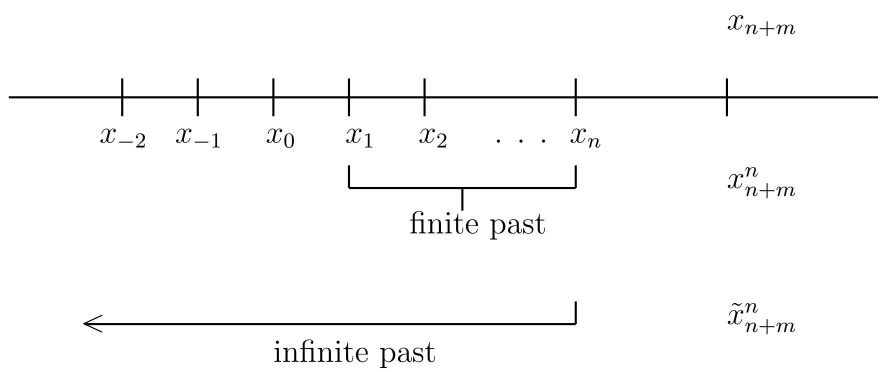
For the moment pretend the x_i are i.i.d. Then, their sample mean is the "best" predictor. Why?

In what sense is it the "best" predictor?

Forecasting a Stationary Series Based on the (In)finite Past

Assumption: The ACF or the parameters, and the past are known.

A pictorial setup for forecasting the future values $x_{n+m}, m = 1, 2, \ldots$:



1. What are their forecasts, forecast error, and forecast error variances?

   Forecast error: $x_{n+m} - x^n_{n+m}$

   Error variance: $P^n_{n+m} = \mathrm{Var}(x_{n+m} - x^n_{n+m})$

2. Their 95% forecast intervals?

$$x^n_{n+m} \pm 1.96\sqrt{P^n_{n+m}}.$$

Forecasting Based on the Finite Past:

A Review of Regression: Given the value of a random variable X, **find** β **to minimize the mean-square error (MSE) of predicting** Y **by** $\widehat{Y} = \beta X$:

$$\text{MSE}(\beta) = E(Y - \hat{Y})^2 = E(Y - \beta X)^2.$$

SOLUTION: The minimizer satisfies the *normal equation*:

$$\text{Var}(X) \;\; \widehat{\beta} = \text{Cov}(X, Y)$$

or

$$\widehat{\beta} = \frac{\text{Cov}(X, Y)}{\text{Var}(X)},$$

and

$$\text{MSE}(\widehat{\beta}) = E(Y - \widehat{Y})^2 = (1 - R^2)\text{Var}(Y).$$

What is R^2?

What is the form of the normal equations when X is a vector with p variables?

Time Series Forecast: Given the time series data $x_1, \ldots, x_n$ from a zero-mean stationary process $\{x_t\}$ with **known** autocovariance function,$\gamma(h)$, find the forecast of the next future value, x_{n+1}.
More precisely, find $\phi_{n1}, \ldots, \phi_n$ to minimize the MSE of the forecast:

$$E(x_{n+1} - \phi_{n1}x_n - \ldots - \phi_{nn}x_1)^2.$$

What are the Normal Equations?

The General Principle of Statistical Forecasting

1. Forecating: Begins when a good model is identified for the time series,

2. Given the time series data $x_1, \ldots, x_n$: **What are the principles for model-based forecasting ?**

   $$x_t = f(\beta, \text{Past of the Series}) + w_t.$$

   Example:

   $$x_t = \phi x_{t-1} + w_t.$$

   Principle: Replace the unknowns by their best ESTIMATES.
   Example:

   $$x_t = w_t + \theta w_{t-1}.$$

3. **Forecasting ARMA Models**

   Recall that causal ARMA models can be written as One-Sided MA(∞) of a white noise:

   $$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j},$$

   and **invertible** ARMA models can be written as One-Sided AR(∞);

   $$x_t = \sum_{j=0}^{\infty} \pi_j x_{t-j} + w_t.$$

~~**Table 3.1.** Behavior of the ACF and PACF for ARMA Models~~

	~~AR(p)~~	~~MA(q)~~	~~ARMA(p, q)~~
~~ACF~~	~~Tails off~~	~~Cuts off after lag q~~	~~Tails off~~
~~PACF~~	~~Cuts off after lag p~~	~~Tails off~~	~~Tails off~~

~~**Example 3.17 Preliminary Analysis of the Recruitment Series**~~

~~We consider the problem of modeling the Recruitment series shown in Figure 1.5. There are 453 months of observed recruitment ranging over the years 1950-1987. The ACF and the PACF given in Figure 3.5 are consistent with the behavior of an AR(2). The ACF has cycles corresponding roughly to a 12-month period, and the PACF has large values for $h = 1, 2$ and then is essentially zero for higher order lags. Based on Table 3.1, these results suggest that a second-order ($p = 2$) autoregressive model might provide a good fit. Although we will discuss estimation in detail in §3.6, we ran a regression (see §2.2) using the data triplets $\{(x; z_1, z_2) : (x_3; x_2, x_1), (x_4; x_3, x_2), \dots, (x_{453}; x_{452}, x_{451})\}$ to fit a model of the form~~

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$$

~~for $t = 3, 4, \dots, 453$. The values of the estimates were $\widehat{\phi}_0 = 6.74_{(1.11)}$, $\widehat{\phi}_1 = 1.35_{(.04)}$, $\widehat{\phi}_2 = -.46_{(.04)}$, and $\widehat{\sigma}^2_w = 89.72$, where the estimated standard errors are in parentheses.~~

~~The following R code can be used for this analysis. We use the script `acf2` to print and plot the ACF and PACF; see Appendix R for details.~~

```
acf2(rec, 48)       # will produce values and a graphic
(regr = ar.ols(rec, order=2, demean=FALSE, intercept=TRUE))
regr$asy.se.coef   # standard errors of the estimates
```

3.5 Forecasting

In forecasting, the goal is to predict future values of a time series, x_{n+m}, $m = 1, 2, \dots$, based on the data collected to the present, $\boldsymbol{x} = \{x_n, x_{n-1}, \dots, x_1\}$. Throughout this section, we will assume x_t is stationary and the model parameters are known. The problem of forecasting when the model parameters are unknown will be discussed in the next section; also, see Problem 3.26. The minimum mean square error predictor of x_{n+m} is

$$x^n_{n+m} = E(x_{n+m} \mid \boldsymbol{x}) \tag{3.57}$$

because the conditional expectation minimizes the mean square error

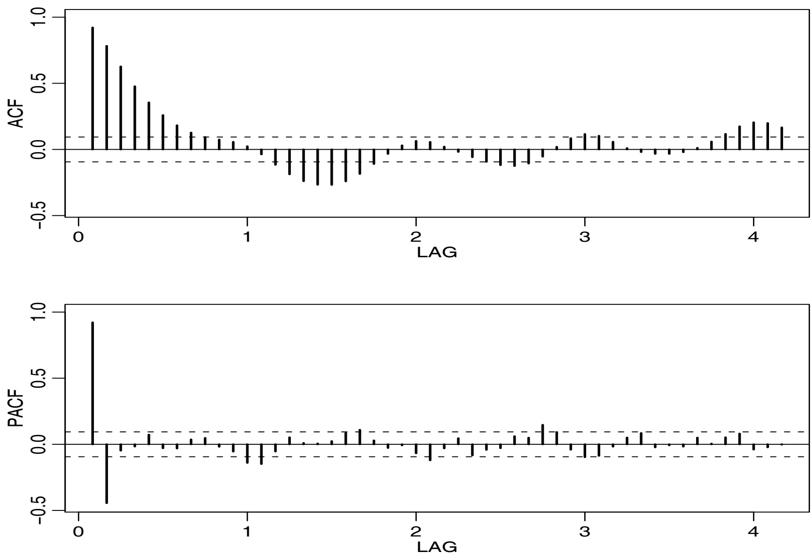


Fig. 3.5. ACF and PACF of the Recruitment series. Note that the lag axes are in terms of season (12 months in this case).

$$E\left[x_{n+m} - g(\boldsymbol{x})\right]^2, \tag{3.58}$$

where $g(\boldsymbol{x})$ is a function of the observations $\boldsymbol{x}$; see Problem 3.14.

First, we will restrict attention to predictors that are linear functions of the data, that is, predictors of the form

$$x_{n+m}^n = \alpha_0 + \sum_{k=1}^{n} \alpha_k x_k, \tag{3.59}$$

where $\alpha_0, \alpha_1, \ldots, \alpha_n$ are real numbers. Linear predictors of the form (3.59) that minimize the mean square prediction error (3.58) are called best linear predictors (BLPs). As we shall see, linear prediction depends only on the second-order moments of the process, which are easy to estimate from the data. Much of the material in this section is enhanced by the theoretical material presented in Appendix B. For example, Theorem B.3 states that if the process is Gaussian, minimum mean square error predictors and best linear predictors are the same. The following property, which is based on the Projection Theorem, Theorem B.1 of Appendix B, is a key result.

Property 3.3 Best Linear Prediction for Stationary Processes

Given data $x_1, \ldots, x_n$, the best linear predictor, $x_{n+m}^n = \alpha_0 + \sum_{k=1}^n \alpha_k x_k$, of x_{n+m}, for $m \geq 1$, is found by solving

$$E\left[\left(x_{n+m} - x_{n+m}^n\right) x_k\right] = 0, \quad k = 0, 1, \ldots, n, \tag{3.60}$$

where $x_0 = 1$, for $\alpha_0, \alpha_1, \ldots \alpha_n$.

Forecast error is UNCORRELTED with the predictors.

The equations specified in (3.60) are called the prediction equations, and they are used to solve for the coefficients $\{\alpha_0, \alpha_1, \dots, \alpha_n\}$. If $E(x_t) = \mu$, the first equation ($k = 0$) of (3.60) implies

$$E(x_{n+m}^n) = E(x_{n+m}) = \mu.$$

Thus, taking expectation in (3.59), we have

$$\mu = \alpha_0 + \sum_{k=1}^{n} \alpha_k \mu \qquad \text{or} \qquad \alpha_0 = \mu\Big(1 - \sum_{k=1}^{n} \alpha_k\Big).$$

Hence, the form of the BLP is

$$x_{n+m}^n = \mu + \sum_{k=1}^{n} \alpha_k (x_k - \mu).$$

Thus, until we discuss estimation, there is no loss of generality in considering the case that $\mu = 0$, in which case, $\alpha_0 = 0$.

First, consider one-step-ahead prediction. That is, given $\{x_1, \dots, x_n\}$, we wish to forecast the value of the time series at the next time point, x_{n+1}. The BLP of x_{n+1} is of the form

PACF

$$x_{n+1}^n = \phi_{n1} x_n + \phi_{n2} x_{n-1} + \cdots + \phi_{nn} x_1, \tag{3.61}$$

where, for purposes that will become clear shortly, we have written α_k in (3.59), as $\phi_{n,n+1-k}$ in (3.61), for $k = 1, \dots, n$. Using Property 3.3, the coefficients $\{\phi_{n1}, \phi_{n2}, \dots, \phi_{nn}\}$ satisfy

$$E\Big[\Big(x_{n+1} - \sum_{j=1}^{n} \phi_{nj} x_{n+1-j}\Big) x_{n+1-k}\Big] = 0, \quad k = 1, \dots, n,$$

or

$$\sum_{j=1}^{n} \phi_{nj} \gamma(k - j) = \gamma(k), \quad k = 1, \dots, n. \tag{3.62}$$

The prediction equations (3.62) can be written in matrix notation as

$$\Gamma_n \boldsymbol{\phi}_n = \boldsymbol{\gamma}_n, \tag{3.63}$$

where $\Gamma_n = \{\gamma(k - j)\}_{j,k=1}^n$ is an $n \times n$ matrix, $\boldsymbol{\phi}_n = (\phi_{n1}, \dots, \phi_{nn})'$ is an $n \times 1$ vector, and $\boldsymbol{\gamma}_n = (\gamma(1), \dots, \gamma(n))'$ is an $n \times 1$ vector.

The matrix Γ_n is nonnegative definite. If Γ_n is singular, there are many solutions to (3.63), but, by the Projection Theorem (Theorem B.1), x_{n+1}^n is unique. If Γ_n is nonsingular, the elements of $\boldsymbol{\phi}_n$ are unique, and are given by

$$\boldsymbol{\phi}_n = \Gamma_n^{-1} \boldsymbol{\gamma}_n. \tag{3.64}$$

For ARMA models, the fact that $\sigma_w^2 > 0$ and $\gamma(h) \to 0$ as $h \to \infty$ is enough to ensure that Γ_n is positive definite (Problem 3.12). It is sometimes convenient to write the one-step-ahead forecast in vector notation

$$\boxed{x_{n+1}^n = \boldsymbol{\phi}_n' \boldsymbol{x},} \tag{3.65}$$

where $\boldsymbol{x} = (x_n, x_{n-1}, \dots, x_1)'$.

The mean square one-step-ahead prediction error is

$$\boxed{P_{n+1}^n = E(x_{n+1} - x_{n+1}^n)^2 = \gamma(0) - \boldsymbol{\gamma}_n' \Gamma_n^{-1} \boldsymbol{\gamma}_n.} \tag{3.66}$$

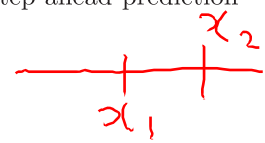
To verify (3.66) using (3.64) and (3.65),

$$\begin{aligned}
E(x_{n+1} - x_{n+1}^n)^2 &= E(x_{n+1} - \boldsymbol{\phi}_n' \boldsymbol{x})^2 = E(x_{n+1} - \boldsymbol{\gamma}_n' \Gamma_n^{-1} \boldsymbol{x})^2 \\
&= E(x_{n+1}^2 - 2\boldsymbol{\gamma}_n' \Gamma_n^{-1} \boldsymbol{x} x_{n+1} + \boldsymbol{\gamma}_n' \Gamma_n^{-1} \boldsymbol{x}\boldsymbol{x}' \Gamma_n^{-1} \boldsymbol{\gamma}_n) \\
&= \gamma(0) - 2\boldsymbol{\gamma}_n' \Gamma_n^{-1} \boldsymbol{\gamma}_n + \boldsymbol{\gamma}_n' \Gamma_n^{-1} \Gamma_n \Gamma_n^{-1} \boldsymbol{\gamma}_n \\
&= \gamma(0) - \boldsymbol{\gamma}_n' \Gamma_n^{-1} \boldsymbol{\gamma}_n.
\end{aligned}$$

Example 3.18 Prediction for an AR(2)

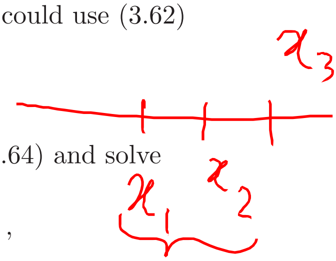
Suppose we have a causal AR(2) process $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$, and one observation x_1. Then, using equation (3.64), the one-step-ahead prediction of x_2 based on x_1 is

$$x_2^1 = \phi_{11} x_1 = \frac{\gamma(1)}{\gamma(0)} x_1 = \rho(1) x_1.$$



Now, suppose we want the one-step-ahead prediction of x_3 based on two observations x_1 and x_2; i.e., $x_3^2 = \phi_{21} x_2 + \phi_{22} x_1$. We could use (3.62)

$$\begin{aligned}
\phi_{21}\gamma(0) + \phi_{22}\gamma(1) &= \gamma(1) \\
\phi_{21}\gamma(1) + \phi_{22}\gamma(0) &= \gamma(2)
\end{aligned}$$



to solve for ϕ_{21} and ϕ_{22}, or use the matrix form in (3.64) and solve

$$\begin{pmatrix} \phi_{21} \\ \phi_{22} \end{pmatrix} = \begin{pmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{pmatrix}^{-1} \begin{pmatrix} \gamma(1) \\ \gamma(2) \end{pmatrix},$$

but, it should be apparent from the model that $x_3^2 = \phi_1 x_2 + \phi_2 x_1$. Because $\phi_1 x_2 + \phi_2 x_1$ satisfies the prediction equations (3.60),

$$E\{[x_3 - (\phi_1 x_2 + \phi_2 x_1)] x_1\} = E(w_3 x_1) = 0,$$

$$E\{[x_3 - (\phi_1 x_2 + \phi_2 x_1)] x_2\} = E(w_3 x_2) = 0,$$

it follows that, indeed, $x_3^2 = \phi_1 x_2 + \phi_2 x_1$, and by the uniqueness of the coefficients in this case, that $\phi_{21} = \phi_1$ and $\phi_{22} = \phi_2$. Continuing in this way, it is easy to verify that, for $n \geq 2$,

$$x_{n+1}^n = \phi_1 x_n + \phi_2 x_{n-1}.$$

That is, $\phi_{n1} = \phi_1, \phi_{n2} = \phi_2$, and $\phi_{nj} = 0$, for $j = 3, 4, \dots, n$.

From Example 3.18, it should be clear (Problem 3.40) that, if the time series is a causal AR(p) process, then, for $n \geq p$,

$$x_{n+1}^n = \phi_1 x_n + \phi_2 x_{n-1} + \cdots + \phi_p x_{n-p+1}. \tag{3.67}$$

For ARMA models in general, the prediction equations will not be as simple as the pure AR case. In addition, for n large, the use of (3.64) is prohibitive because it requires the inversion of a large matrix. There are, however, iterative solutions that do not require any matrix inversion. In particular, we mention the recursive solution due to Levinson (1947) and Durbin (1960).

Property 3.4 The Durbin–Levinson Algorithm

Equations (3.64) and (3.66) can be solved iteratively as follows:

$$\phi_{00} = 0, \quad P_1^0 = \gamma(0). \tag{3.68}$$

For $n \geq 1$,

$$\phi_{nn} = \frac{\rho(n) - \sum_{k=1}^{n-1} \phi_{n-1,k}\ \rho(n-k)}{1 - \sum_{k=1}^{n-1} \phi_{n-1,k}\ \rho(k)}, \quad P_{n+1}^n = P_n^{n-1}(1 - \phi_{nn}^2), \tag{3.69}$$

where, for $n \geq 2$,

$$\phi_{nk} = \phi_{n-1,k} - \phi_{nn}\phi_{n-1,n-k}, \quad k = 1, 2, \ldots, n-1. \tag{3.70}$$

The proof of Property 3.4 is left as an exercise; see Problem 3.13.

Example 3.19 Using the Durbin–Levinson Algorithm

To use the algorithm, start with $\phi_{00} = 0$, $P_1^0 = \gamma(0)$. Then, for $n = 1$,

$$\phi_{11} = \rho(1), \quad P_2^1 = \gamma(0)[1 - \phi_{11}^2].$$

For $n = 2$,

$$\phi_{22} = \frac{\rho(2) - \phi_{11}\ \rho(1)}{1 - \phi_{11}\ \rho(1)},\ \phi_{21} = \phi_{11} - \phi_{22}\phi_{11},$$
$$P_3^2 = P_2^1[1 - \phi_{22}^2] = \gamma(0)[1 - \phi_{11}^2][1 - \phi_{22}^2].$$

For $n = 3$,

$$\phi_{33} = \frac{\rho(3) - \phi_{21}\ \rho(2) - \phi_{22}\ \rho(1)}{1 - \phi_{21}\ \rho(1) - \phi_{22}\ \rho(2)},$$
$$\phi_{32} = \phi_{22} - \phi_{33}\phi_{21},\ \phi_{31} = \phi_{21} - \phi_{33}\phi_{22},$$
$$P_4^3 = P_3^2[1 - \phi_{33}^2] = \gamma(0)[1 - \phi_{11}^2][1 - \phi_{22}^2][1 - \phi_{33}^2],$$

and so on. Note that, in general, the standard error of the one-step-ahead forecast is the square root of

$$P_{n+1}^n = \gamma(0) \prod_{j=1}^n [1 - \phi_{jj}^2]. \tag{3.71}$$

An important consequence of the Durbin–Levinson algorithm is (see Problem 3.13) as follows.

Property 3.5 Iterative Solution for the PACF

The PACF of a stationary process x_t, *can be obtained iteratively via (3.69) as* ϕ_{nn}, *for* $n = 1, 2, \ldots$.

Using Property 3.5 and putting $n = p$ in (3.61) and (3.67), it follows that for an AR(p) model,

$$\begin{aligned} x_{p+1}^{p} &= \phi_{p1}\, x_p + \phi_{p2}\, x_{p-1} + \cdots + \phi_{pp}\, x_1 \\ &= \phi_1\, x_p + \phi_2\, x_{p-1} + \cdots + \phi_p\, x_1. \end{aligned} \tag{3.72}$$

Result (3.72) shows that for an AR(p) model, the partial autocorrelation coefficient at lag p, ϕ_{pp}, is also the last coefficient in the model, ϕ_p, as was claimed in Example 3.15.

Example 3.20 The PACF of an AR(2)

We will use the results of Example 3.19 and Property 3.5 to calculate the first three values, ϕ_{11}, ϕ_{22}, ϕ_{33}, of the PACF. Recall from Example 3.9 that $\rho(h) - \phi_1\rho(h-1) - \phi_2\rho(h-2) = 0$ for $h \geq 1$. When $h = 1, 2, 3$, we have $\rho(1) = \phi_1/(1-\phi_2)$, $\rho(2) = \phi_1\rho(1) + \phi_2$, $\rho(3) - \phi_1\rho(2) - \phi_2\rho(1) = 0$. Thus,

$$\begin{aligned} \phi_{11} &= \rho(1) = \frac{\phi_1}{1-\phi_2} \\ \phi_{22} &= \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2} = \frac{\left[\phi_1\left(\frac{\phi_1}{1-\phi_2}\right) + \phi_2\right] - \left(\frac{\phi_1}{1-\phi_2}\right)^2}{1 - \left(\frac{\phi_1}{1-\phi_2}\right)^2} = \phi_2 \\ \phi_{21} &= \rho(1)[1-\phi_2] = \phi_1 \\ \phi_{33} &= \frac{\rho(3) - \phi_1\rho(2) - \phi_2\rho(1)}{1 - \phi_1\rho(1) - \phi_2\rho(2)} = 0. \end{aligned}$$

Notice that, as shown in (3.72), $\phi_{22} = \phi_2$ for an AR(2) model.

So far, we have concentrated on one-step-ahead prediction, but Property 3.3 allows us to calculate the BLP of x_{n+m} for any $m \geq 1$. Given data, $\{x_1, \ldots, x_n\}$, the m-step-ahead predictor is

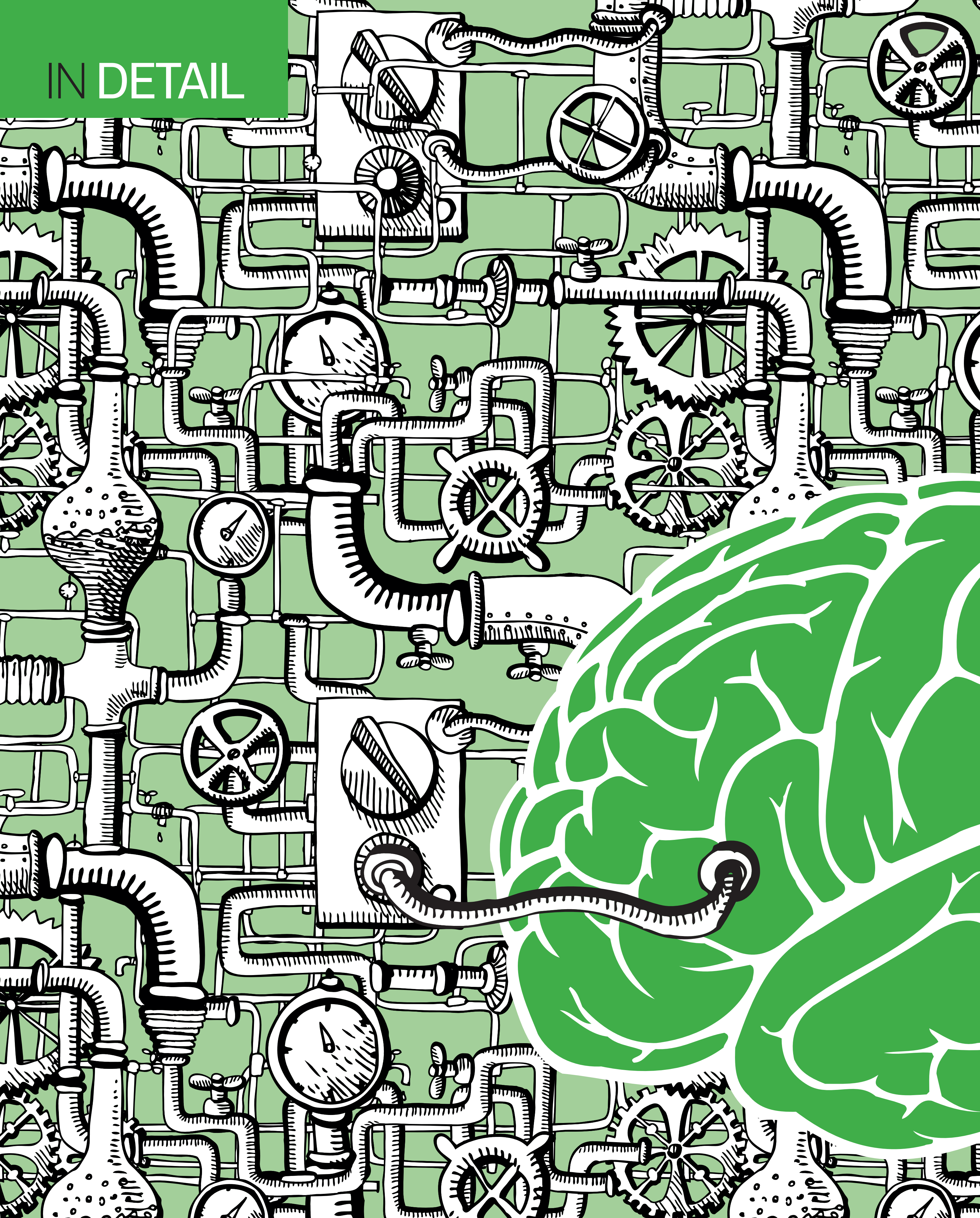
$$x_{n+m}^{n} = \phi_{n1}^{(m)} x_n + \phi_{n2}^{(m)} x_{n-1} + \cdots + \phi_{nn}^{(m)} x_1, \tag{3.73}$$

where $\{\phi_{n1}^{(m)}, \phi_{n2}^{(m)}, \ldots, \phi_{nn}^{(m)}\}$ satisfy the prediction equations,

$$\sum_{j=1}^{n} \phi_{nj}^{(m)} E(x_{n+1-j}x_{n+1-k}) = E(x_{n+m}x_{n+1-k}), \quad k = 1, \ldots, n,$$

or

IN DETAIL



Great expectations: The past, present and future of prediction

From ancient oracles to statistical models, **Graham Southorn** gets to grips with forecasting methods, how they have evolved, and what the future might hold

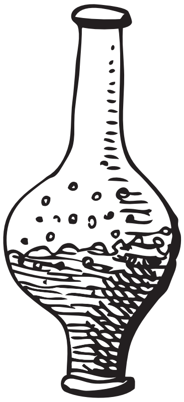
Graham Southorn is a freelance science writer, and former editor of *BBC Focus Magazine*

We are all fascinated by the future. Whether it is the rise and fall in interest rates, the outcome of elections, or winners at the Oscars, there is sure to be something you want to know ahead of time. There is certainly no shortage of pundits with ready opinions about what the future might hold – but their predictions might not be entirely reliable. A 20-year study, published in 2006, showed that the average expert did little better than guessing on many of the political and economic questions asked of them.[1]

But expert predictions are only part of the forecasting story (see "Prediction versus forecasting", page 19). A raft of methods – from mathematical models to betting markets – are promising new ways of seeing into the future. And it is not only academics and professionals who can do it – online services allow anyone to have a go.

Forecasting could probably stake a claim to being one of the world's oldest professions. Beginning in the eighth century BC, a priestess known as the Pythia would answer questions about the future at the Temple of Apollo on Greece's Mount Parnassus.[2] It is said that she, the Oracle of Delphi, dispensed her wisdom in a trance – caused, some believe, by the hallucinogenic gases that would seep up through natural vents in the rock.

By the second century BC, the ancient Greeks had moved on to more sophisticated methods of prediction, such as the Antikythera mechanism, whose intricate bronze gears seemed capable of predicting a host of astronomical events such as eclipses.



Over time, our astronomical predictions became more refined, and in 1687 Isaac Newton published his laws of motion and gravitation. Newton's friend Edmond Halley predicted in 1705 the return of the comet that now bears his name. But forecasters also started to concern themselves with more mundane, earthly matters. By the nineteenth century, the new technology of long-distance telegrams meant that, for the first time, data from a network of weather stations could be transmitted in advance of changing conditions. This did not only spur developments in meteorology. People began to believe that similarly scientific measurements might be useful in other areas, such as business.

Forerunners

Early economic forecasters, such as Roger Babson, have been profiled by Walter A. Friedman in his book *Fortune Tellers*.[3] Their work was not only inspired by weather forecasters, says Friedman, in an interview with *Significance*. "It developed because of the sharp ups and downs of prices in the late nineteenth and early twentieth century, coupled with the desire of businesspeople to make future plans. Babson developed his service, for instance, after the 1907 panic. The way he shaped his forecasting method was reliant on traditions of barometers, ideas about the business cycle, and the availability of data."

Whereas Babson's method looked at trends in data over time, his contemporary, Irving Fisher, built a machine to model how the flow of supply and demand of one commodity affected that of others. It was a hydraulic computer. When Fisher adjusted a lever, water flowed through a series of tubes to restore equilibrium between the prices of goods. A similar device was built by William Phillips at the London School of Economics in 1949. It used channels of coloured water to replicate taxes, exports and spending in the British economy.

Babson was alone among his contemporaries in predicting the 1929 stock market crash. However, says Friedman, he and other early forecasters made assumptions that were very simplistic and often misguided.

Wrong though they may have been, the methods employed by Babson and Fisher were forerunners of the range of statistical tools available today. Babson worked with time series, plotting the aggregate of variables like crop production, commodity prices and business failures on a single chart that forecast how the economy would fare. Fisher's hydrostatic machine, on the other hand, did not include a time element. In Friedman's book, the machine is described as the grandparent

Cover and feature design: Kevin Reed/CPL

© 2016 The Royal Statistical Society

of the economic forecasting models developed after World War II and run on computers.

The science of forecasting

Techniques that we know today were refined in a variety of different fields, says Rob Hyndman, professor of statistics at Monash University in Australia. "In the last 50 years or so, people started doing more time series models that try to relate the past to the future. They became extremely useful in sales forecasting and in predicting demand for items. In other fields such as engineering, they started trying to build models to predict things like river flow based on rainfall. Models for electricity demand were developed so that they could plan generation capacity."

The science of forecasting got going properly in the 1980s, says Hyndman. "People realised that if you took all of the ideas that people had developed in different fields, and you thought of it as a collection of techniques and overlaid that with analytical and scientific thinking, then forecasting itself could be considered a scientific discipline."

But how do we know whether or not quantitative forecasting will work in any given area? Hyndman, who is editor-in-chief of the *International Journal of Forecasting*, believes the predictability of an event boils down to three factors. "The first is whether you have an idea of what's driving it – the causal factors. So you might just have data on the particular thing you're interested in, but you don't understand why the fluctuations occur or why the patterns exist in the data. You can still forecast it, but not so well. If you haven't understood the way in which the thing you're interested in reacts to the driving factors, you're going to lose that coupling over time," he says.

The two other factors involved are the availability of data and whether a forecast will itself affect what it is that you are trying to predict. If an exchange rate rise is predicted, for instance, it will affect prices in the real markets, which will end up influencing the rate itself.

These factors aside, a successful forecaster also needs a toolbox of statistical methods and the know-how to pick the right method for a particular situation, says Hyndman (see "Forecasting techniques explained", page 17). Statistical models, though, only work in the short term. "They're not very good for long-term forecasting because the big assumption – that the future looks similar to the past – slowly breaks down the further you get into the future," he says.

"Then there are problems where there's just not really enough data to be able to build good models, or situations that are not reflected at all historically, such as technological changes," says Hyndman. "There's no data available that will tell you what's going to happen."

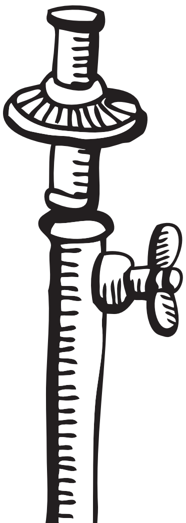
Superforecasters

Unforeseen developments – whether technological, political or social – pose an interesting dilemma for those whose job is to anticipate such things: national security agencies, for instance. What if such developments are predictable, not from a single data set or time series, perhaps, but from the aggregated opinions of groups of individuals? In 2011, the Aggregative Contingent Estimation (ACE) programme set out to answer that question, with funding from the Intelligence Advanced Research Projects Activity (IARPA). ACE announced a forecasting tournament that would run from September 2011 to June 2015, in which five teams would compete by answering 500 questions on world affairs.

One of the teams was the Good Judgment Project (GJP), which was created by Barbara Mellers and political scientist Philip Tetlock, the man behind the 2006 research on expert predictions. The GJP attracted over 20 000 wannabe forecasters in its first year. In his 2015 book, *Superforecasting*, Tetlock recalls how the forecasters were asked to predict "if protests in Russia would spread, the price of gold would plummet, the Nikkei would close above 9500, or war would erupt on the Korean peninsula".[4]

The GJP won the tournament hands down. According to Tetlock, it beat a control group by 60% in its first year and 78% in its second year. After that, IARPA decided to drop the others, leaving the GJP as the last team standing.

As described in *Superforecasting*, the GJP continually assigned its participants an ever-changing rating called a "Brier score", which measures the accuracy of predictions on a scale from 0 to 2; the lower the number, the more accurate the prediction. By doing so, they were able to identify the best among them, whom they called superforecasters. Some superforecasters were plucked from the crowd and placed in 12-person "superteams" that could share information with each other. Would the superteams perform any better?



We need to talk about Nate

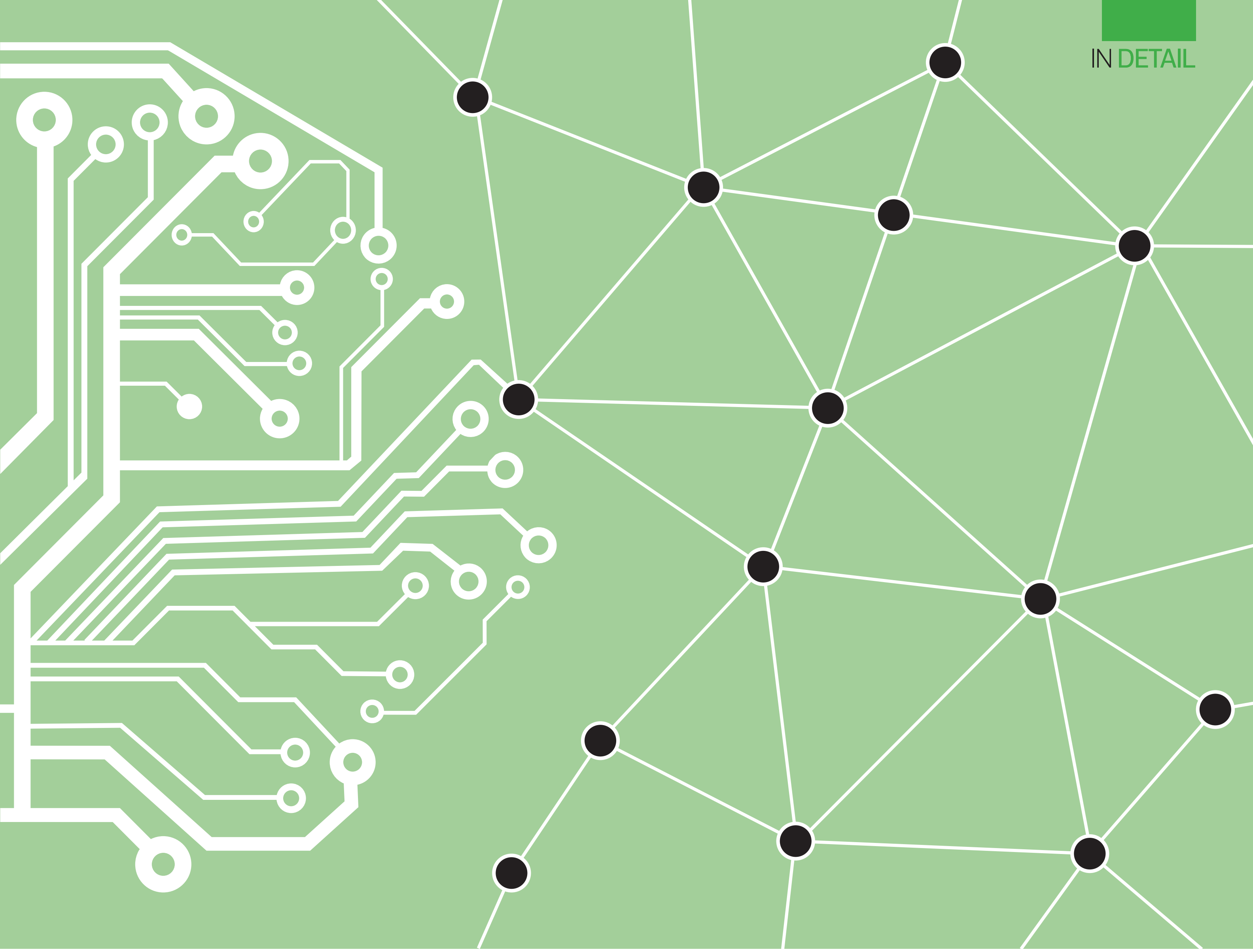
One name crops up alongside virtually every mention of forecasting. Nate Silver is the high-profile American statistician best known for developing a mathematical model that correctly called the results in 49 of 50 states in the 2008 presidential election. The clever part was the way it incorporated polling data, explains Professor Rob Hyndman. "At least for the last two presidential elections, Nate Silver developed some very good Bayesian techniques for combining all of the prediction polls to get good forecasts of what would happen on election day."

Bayesian techniques are rooted in Bayes' theorem, which provides a means of updating levels of belief in the light of new evidence. Thus as new polls are taken, Bayesian methods allows their findings to be combined with current polling evidence to produce an updated level of belief about, say, the various outcomes of an election.

But Silver's foresight may not be as startling as it first appears. In his book *Superforecasting*, Philip Tetlock points out that a "no change" forecast, in which the political party that won a state in the previous election merely holds onto it, would have correctly predicted 48 out of 50 results.

The same idea in meteorology is known as persistence forecasting. That is, the weather in future is forecast to be the same as it is now. The technique is only useful for very short-range forecasts or slowly evolving weather patterns.

Finally, it is worth noting that a forecast is only as good as the data it is based on. In the 2015 UK General Election, Silver forecast that the Conservative Party would win more seats than Labour – but not as many as the Conservatives actually gained: the model's prediction interval fell short of the final tally of 331 seats. In this case, the opinion polls used by Silver, and many others, failed to accurately reflect voting intentions, which had a knock-on effect on prediction accuracy.



Forecasting techniques explained

Rob Hyndman, professor of statistics at Monash University, outlines the different kinds of statistical forecasting methods

On time-series forecasts

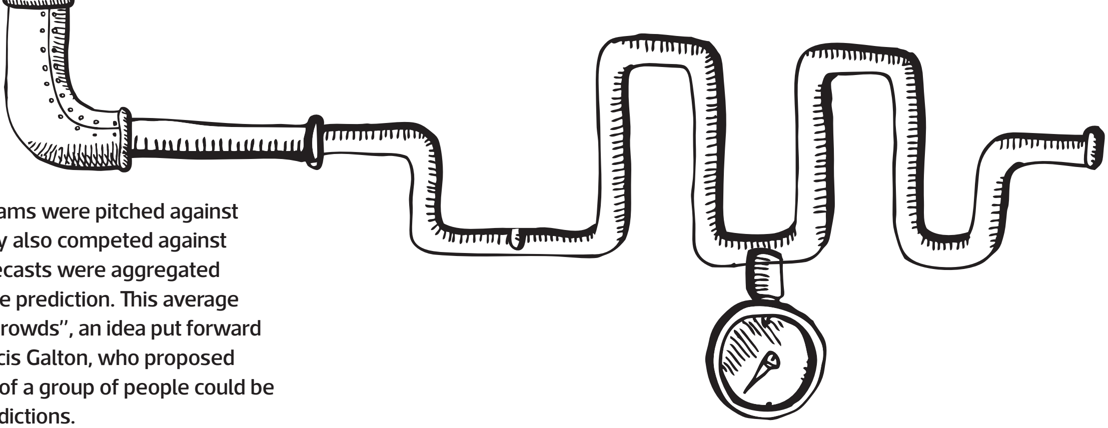
A purely time-series approach just looks at the history of the variable you are interested in and builds a model that describes how that has changed over time. You might look at the history of monthly sales for a company. You look at the trends and seasonality and extrapolate it forward, but you do not take any other information into account.

On explanatory models

This is where you relate the thing you are trying to forecast with other things that might affect it. So if you are forecasting sales you might also take into account population and the state of the economy. All you actually need is for the drivers to be good predictors of the outcome, whether or not there is a direct causal relationship or something more complicated. You can also have models that combine the two. So you have some time series and some other information, and you build a model that puts it all together.

On probabilistic forecasts

For a long time people have been producing predictions with prediction intervals – giving a statement of uncertainty using probabilities. The new development is not just giving an interval but giving the entire probability distribution as your forecast. So you will say, "Here's a number: the chance of being below this is 1%. Here's another number: the chance of being below this is 2%," and so on over 100 percentiles. But if you are giving an entire probability distribution, you cannot measure the accuracy of your forecast in a simple way. New techniques for measuring forecast accuracy that take this into account have become popular in the last five years or so: these are called probability scoring methods.



It turns out they could. Superteams were pitched against teams of regular forecasters. They also competed against individual forecasters, whose forecasts were aggregated to produce an unweighted average prediction. This average represented the "wisdom of the crowds", an idea put forward in 1907 by the statistician Sir Francis Galton, who proposed that the accumulated knowledge of a group of people could be more accurate than individual predictions.

The superteams faced one final group of competitors: forecasters who had been assigned to work as traders in prediction markets, a popular form of forecasting in which people place bets on the outcome they think is most likely to happen. Writing in *Superforecasting*, Tetlock says: "The results were clear-cut … . Teams of ordinary forecasters beat the wisdom of the crowd by about 10%. Prediction markets beat ordinary teams by about 20%. And superteams beat prediction markets by 15% to 30%."[4]

Want to bet?

The fact that Tetlock's superteams were able to beat the markets was something of a surprise. Futures exchanges, such as the Iowa Electronic Markets (IEM) and PredictIt, have an enviable track record in predicting outcomes, especially political outcomes. In November 2015, Professor Leighton Vaughan Williams, director of the Betting Research Unit at Nottingham Business School, and his co-author Dr James Reade, published "Forecasting Elections", a study that compared prediction markets to more traditional methods.[5] "We got huge amounts of data from InTrade and Betfair, plus statistical modelling, expert opinion and every opinion poll. We compared them over many years and literally hundreds of different elections using state-of-the-art econometrics," says Vaughan Williams. "We found that prediction markets significantly outperformed the other methodologies included in our study, and even more so as you get closer to the event."

Similarly, a study conducted by Joyce E. Berg and colleagues, published in 2008, compared IEM predictions to the results of 964 polls over five US presidential elections since 1988. They found that "the market is closer to the eventual outcome 74% of the time" and that "the market significantly outperforms the polls in every election when forecasting more than 100 days in advance".[6]

So why were prediction markets beaten by superforecasters in Tetlock's research? One reason could be, as Tetlock himself writes, that the prediction markets in his contest lacked "liquidity" – in other words, they did not feature substantial amounts of money or activity. Vaughan Williams believes liquid markets would normally win for particular kinds of event. "If you ask [superforecasters] to beat a market on who's going to win the Oscars or the Super Bowl or Florida in the 2016 US election they'd find it tough – because millions of pounds will be traded," he says. "But if you ask them: 'Will David Cameron have a meeting with Jean-Claude Juncker by Thursday night?', in that situation there's no real market for them to beat."

Markets are not very good at predicting things that are inherently unpredictable, however. "A prediction market can't

> Mathematical models are simplifications of reality – and life is sometimes too complex to model accurately

aggregate information on something that people can't work out, like the Lottery or earthquakes," says Vaughan Williams. "You know where earthquakes are more likely to happen, but you can't predict on what date they'll occur."

Foreseeing terrorist attacks is something else that markets are just not built for, he adds. "Terrorists are hardly going to be putting their money in and tipping their hand – if anything, they would be doing the opposite."

But where they can be used, prediction markets have another advantage, says Vaughan Williams. "Often what's just as important as knowing what the future will be is knowing it before somebody else. I've just accepted a paper for the *Journal of Prediction Markets* showing that the IEM's influenza market is effective for predicting outbreaks. It's because the market aggregates information from everyone on the system in real time, second by second. If you see everyone around you sneezing with what looks like flu, you could go to your computer and start trading."

Managing complexity

Predicting epidemics is an area where statistical forecasting methods struggle, says Rob Hyndman. "We just don't have the data on which viruses are brewing and which mosquito populations are breeding, so it's extremely difficult – if not impossible," he says. Google infamously sought a way round this problem by analysing search activity to predict the spread of the influenza virus. Early estimates were reliable and accurate, but in time the model produced overestimates – in part because it failed to fully account for the fact that flu-related searches might be made by healthy individuals.[7]

One area which has seen steady progress, however, is weather and climate prediction. A review published in *Nature* in September 2015 revealed the "quiet revolution" in numerical weather prediction that has made today's six-day forecasts as accurate as the five-day forecasts of 10 years ago.[8] Forecasters are able to give one to two weeks'

Prediction versus forecasting

The terms "prediction" and "forecasting" are often used interchangeably – as is the case in this article. But as far as anyone has managed to pin down a definition, one school of thought holds that forecasting is about the future – tomorrow's temperature, for example. Prediction, in contrast, involves finding out about the unobserved present. If you want to determine how much your house will sell for, you could make a prediction based on the prices of houses in your neighbourhood.

In a blog post on this issue (bit.ly/21vlCGI), Galit Shmueli, Distinguished Professor of Business Analytics at Taiwan's National Tsing Hua University, wrote: "The term 'forecasting' is used when it is a time series and we are predicting the series into the future. Hence 'business forecasts' and 'weather forecasts'. In contrast, 'prediction' is the act of predicting in a cross-sectional setting, where the data are a snapshot in time (say, a one-time sample from a customer database). Here you use information on a sample of records to predict the value of other records (which can be a value that will be observed in the future)."

lead time for extreme events like Russia's recent heat wave and the US cold snap, while fluctuations in sea surface temperatures following El Niño can be predicted three to four months beforehand.

The improvements have been driven by number-crunching power from supercomputers, backed by a hierarchy of models of varying complexity, and global data from satellites.

However, there are inherent limitations to modelling complex systems like the climate, according to David Orrell, who runs the scientific consultancy Systems Forecasting. "If you look at the formation of a cloud, you have an interaction between minute particles of something that forms a seed for a droplet. The droplet grows, and that process is incredibly non-linear and very sensitive to small changes. It involves things over all scales, from the microscopic scale to the scale of a cloud."

Orrell, whose work involves forecasting the effects of cancer drugs, says organic systems like the climate and the human body are fundamentally different from the kinds of mechanistic systems we are good at modelling. "The dream is that if we just add more and more levels of detail we'd be able to capture this [behaviour], but there's a fundamental limitation to what you can do with mechanistic models."

Whatever next?

So how will forecasting evolve in the future? In terms of opinion-based predictions, look no further than Almanis, a cross between a prediction market and the Good Judgment Project. Describing itself as a "crowd wisdom platform", Almanis incentivises forecasters with points, not pounds, although it awards real money prizes to the most accurate users. It is a commercial entity, making money by charging companies or governments to post questions.

Services like Almanis will be commonly used within the next 10 years, believes Leighton Vaughan Williams. "As academic research further improves their efficiency I think they'll become a key part of corporate forecasting and information aggregation. Say I want to reduce the waiting list in an eye clinic and I've got a budget of £100 000. Should I hire a doctor, or two nurses? A prediction market can give you that sort of information."

In terms of mathematical forecasting, Rob Hyndman says that methods are being developed to cope with massive data. "One trend that's happening at the moment is that a lot of the techniques that computer scientists have developed in machine learning in other fields are coming into forecasting. It's very interesting to see what's going on."

Statistical models are also being used to combine other types of predictions into meta-forecasts. One example is PredictWise, an academic project started by David Rothschild, an economist at Microsoft Research. The tool combines information from prediction markets, opinion polls and bookmakers' odds to come up with probabilities for everything from the next James Bond to the likelihood of the UK leaving the European Union.

Probabilities are not the same as certainties, however. "In 2008, Hillary Clinton had a 20% chance of winning the New Hampshire primary and she won it," recalls Vaughan Williams. "People said the prediction markets had got it all wrong. But as any statistician would know, what they're saying is that one time in five it's going to happen."

As for mathematical models, Rob Hyndman makes the point that they are always just simplifications of reality – and life is sometimes too complex to model, whether accurately or approximately. The future, or parts of it, therefore, will remain unforeseen. But it is safe to predict that forecasters will keep trying to catch a glimpse of what lies ahead. ■

References

1. Tetlock, P. E. (2006) *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton, NJ: Princeton University Press.

2. Orrell, D. (2007) *The Future of Everything: The Science of Prediction*. New York: Basic Books.

3. Friedman, W. A. (2014) *Fortune Tellers: The Story of America's First Economic Forecasters*. Princeton, NJ: Princeton University Press.

4. Tetlock, P. E. and Gardner, D. (2015) *Superforecasting: The Art and Science of Prediction*. New York: Crown.

5. Vaughan Williams, L. and Reade, J. (2015) Forecasting elections. *Journal of Forecasting*. doi: 10.1002/for.2377

6. Berg, J. E., Nelson, F. D. and Rietz, J. A. (2008) Prediction market accuracy in the long run. *International Journal of Forecasting*, **24**(2), 285–300.

7. Lazer, D. and Kennedy, R. (2015) What we can learn from the epic failure of Google Flu Trends. *Wired*, 1 October. bit.ly/1L8xCUK

8. Bauer, P., Thorpe, A. and Brunet, G. (2015) The quiet revolution of numerical weather prediction. *Nature*, **525**, 47–55.