FORECASTABILITY: INSIGHTS FROM PHYSICS, GRAPHICAL DECOMPOSITION, AND INFORMATION THEORY *by Peter M. Catt*

INTRODUCTION

My aim with this paper is to equip forecasters with some cross-disciplinary theory on forecastability and to provide practical techniques for assessing how forecastable a historical time series is.

A time series is a sequence of values at equally spaced time intervals: days, weeks, months, quarters, or years. The historical time series can be viewed as an outcome (realization) of an underlying* data generating process *(DGP). Assessments of forecastability require an understanding of the DGP and its components.

FORECASTABILITY CONCEPTS FROM PHYSICS

The study of physics offers some fundamental concepts that help establish the theoretical forecastability of a given time series.

Classical physicists assumed that all events were caused by earlier events according to known laws of nature. The French mathematician and astronomer Pierre-Simon Laplace claimed that if the current state of the universe were known with precision, together with all the laws of nature, any event could be precisely predicted for any time in the future. Laplace famously said:

"We ought to regard the present state of the universe as the effect of its anterior state and as the cause of the state that is to follow. Assume an intelligence which could know all the forces by which nature is animated, and the states at an instant of all objects that compose it; …for [this intelligence], nothing could be uncertain; and the future, as the past, would be present to its eyes." (Popper, 1988, Preface XX)

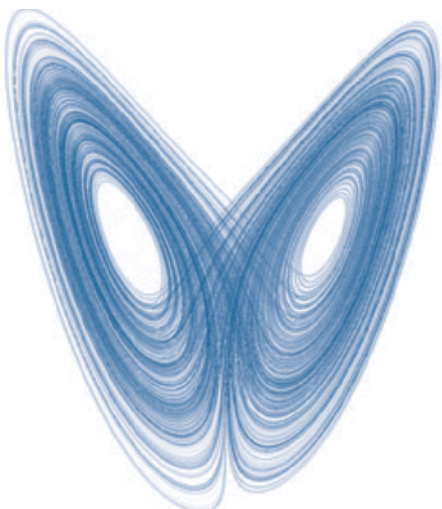
The all-knowing intelligence described above was later coined "Laplace's demon." Although often taken literally, it was intended to demonstrate the idea of scientific determinism – that the future has absolute predictability.

This view wasn't seriously challenged until the development of quantum mechanics in the early 20th century, culminating in Werner Heisenberg's uncertainty principle in 1920. This principle asserts that the more precisely we know the position of a particle, the less precisely we can know its momentum, and vice versa. Rather than having precise values, position and momentum follow a probability distribution, which is to say they are stochastic (i.e., random) in nature. Albert Einstein objected, making his renowned declaration in 1926 that "God does not play dice with the universe." Nevertheless, quantum mechanics continued to demonstrate the presence of such stochastic elements in nature as wave-particle duality, quantum entanglement, and radioactive decay. Today, the prevailing school of thought (the Copenhagen Interpretation) accepts random occurrences at a quantum level.

True randomness implies no underlying deterministic system (i.e., no cause-effect relationship), and is also known as indeterminism (Popper, 1988).

In 1963 the mathematician and meteorologist Edward Lorenz discovered that a simplified weather-forecasting model based on atmospheric convection-rolls never settled down. Instead, it continued to oscillate in an irregular, aperiodic fashion, represented by the now-famous Lorenz Strange Attractor.

Lorenz's Strange Attractor



Lorenz's observations led to chaos theory, which states that some nonlinear systems display a "sensitive dependence on initial conditions," or "error inflation," meaning that two slightly different initial conditions could yield totally different future behaviors, also known as the butterfly effect (Strogatz, 1994). This happens even though chaotic systems are deterministic: their future dynamics are defined by their initial conditions, with no random elements involved. Such systems, while theoretically predictable, pose practical challenges if we cannot accurately measure the initial conditions.

Key Points

- To determine how forecastable a historical time series is, forecasters should consider the underlying data generating process (DGP). Such processes can be broadly classified as deterministic, chaotic, complex, or random. Time series graphs can illustrate the continuum from purely deterministic (completely forecastable) to random (completely unforecastable).
- Coefficients of variation – ratios of the standard deviation to the mean of a time series – have been proposed to assess forecastability. These metrics, however, do not account for structural elements other than trend and seasonality, and so are of limited use in forecastability assessments.
- More valuable are information-theoretic metrics, such as *approximate entropy*, which measure the degree of disorder or irregularity in a series and can detect many patterns, not merely trend and seasonality.
- Analysis of six sample time series shows that an approximate entropy metric gives a truer picture of the relative forecastability of different time series than does the coefficient of variation. A template is provided for calculation of this metric.

Another development relevant to the forecast analyst was the related discipline of complex systems such as animal and human social systems, financial markets and other economic systems. Emergence refers to the way complex systems and patterns arise out of a multiplicity of relatively simple interactions. Emergence can be defined as "the arising of novel and coherent structures, patterns and properties during the process of self-organization in complex systems" (Corning, 2002, p.7). The economy is such a system, reflecting the interactions of millions of constituents. Like chaotic behavior, emergent behavior from complex systems presents real-world forecasting challenges.

This brief review of the physics literature leads to the following classification of the processes underlying the data we observe (the DGPs):

- **Deterministic**: To paraphrase Laplace, we know both (a) the data generating process ("all the forces by which nature is animated") and (b) the parameters (coefficients) on the variables entering the DGP accurately ("the states at an instant of all objects that compose it").
- **Chaotic** (e.g., the Hénon map): We know the data generating process and its parameters, but do not typically know the initial conditions setting the state of the process.
- **Complex**: The data generating process cannot be completely understood.
- **Random**: The data generating process is without pattern.

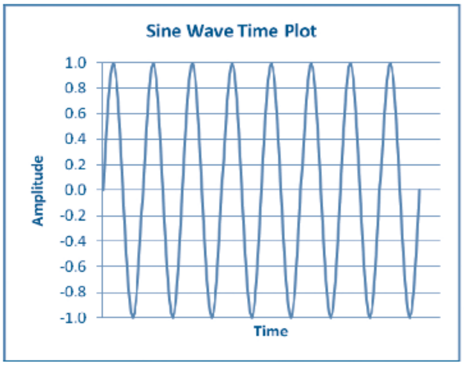
As we move along the continuum from deterministic to random, the process becomes more difficult to identify and hence its future more difficult to forecast. For those interested in a detailed discussion around these issues, I recommend Jean Bricmont's paper "Determinism, Chaos, and Quantum Mechanics" (2003).

Few of the time series that we seek to forecast are purely deterministic or random; they sit somewhere in between. For example, we can forecast the number and positions of airplanes in the sky for every quarter hour tomorrow, simply because we know flight routes (the DGP), departure times, and aircraft speeds (parameters). However, unforeseen circumstances (the random component) lead to delays and cancellations.

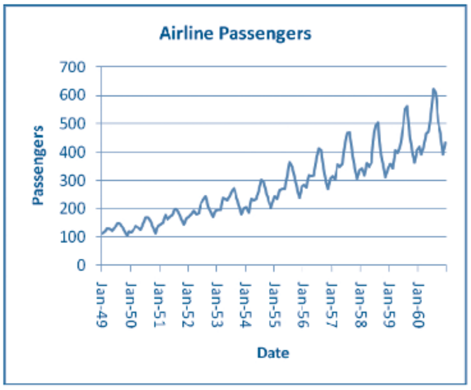
ILLUSTRATIVE TIME SERIES

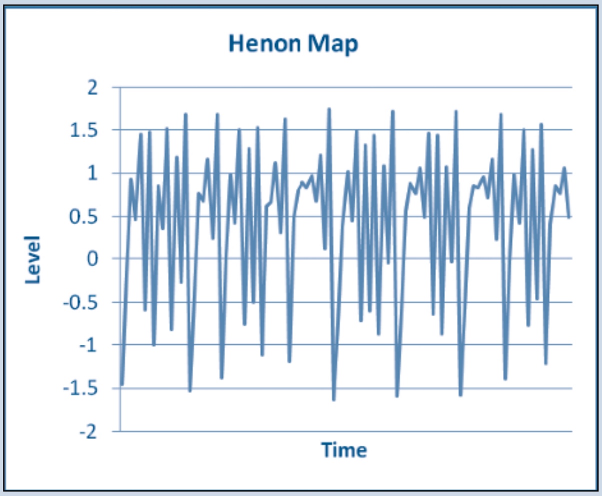
Here are six time series chosen to represent the spectrum of forecastability from deterministic to random.

1. Purely Deterministic: A Sine Wave. 193 data points generated using Microsoft Excel. The DGP is the formula: =SIN(0.2*A), where the argument is in radians. A=incremental row values of 1 to 193. Since there is no random component, knowledge of the DGP (the sine wave) and the parameter value (0.2) enables this perfectly regular, cyclical pattern to be forecast without error.



2. Highly Deterministic: The Airline Series. Monthly boardings at Heathrow Airport, in thousands, Jan. 1949 – Dec. 1960. Source: Box & Jenkins (1976). The DGP includes readily identifiable trend and seasonal patterns.



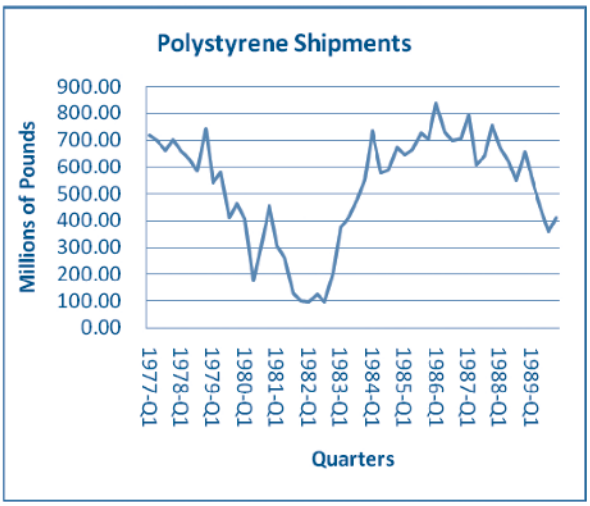


3. Chaotic: Hénon Map. 120 data points. The Hénon Map is a dynamical system that exhibits chaotic behavior. The DGP for the Hénon Map is given by:

$$x_{n+1} = y_n + 1 - ax_n^2$$

$$y_{n+1} = bx_n$$

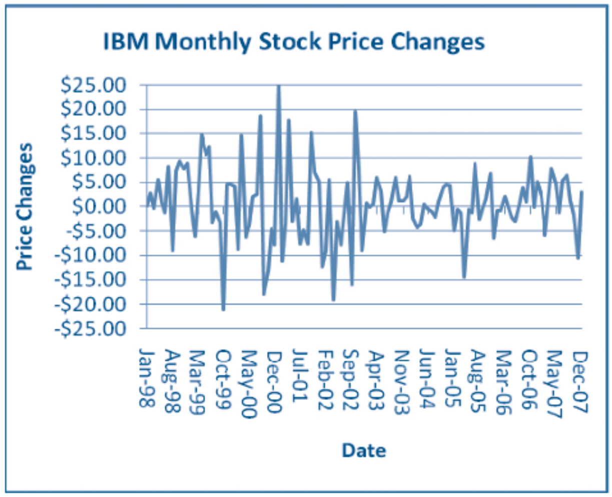
For illustrative purposes, the Hénon map shown has parameter values of: a=1.4 and b=0.3 and initial conditions of y_1=1 and x_1=1.



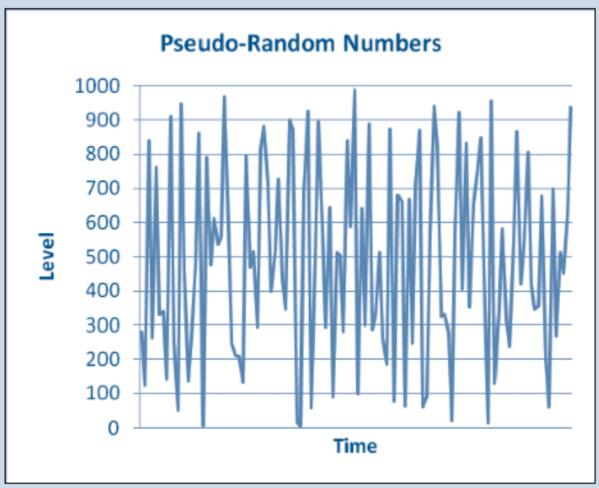
4. Complex Series (Industrial): Polystyrene Shipments. 1977 Q1 to 1989 Q4 (52 data points). Source: U.S. Census of Manufacturing

The polystyrene shipments series results from external demand by automotive, housing, and packaging industries, with price effects, environmental rules, and interactions among all these.

5. Complex Series (Finance): IBM Monthly Stock Price Changes. Jan. 1998 – Dec. 2007 (120 data points). Source: Yahoo Finance (2008). The data generating process is, at best, obscure.



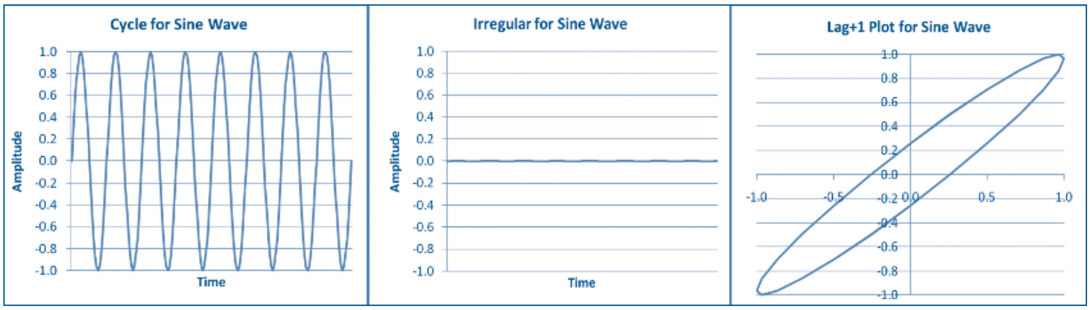
6. Purely Random Series: Pseudo-Random Numbers. 120 data points generated using the Microsoft Excel Rand() function. This simulated series has no discernible deterministic component.



DECOMPOSITION AND FORECASTABILITY

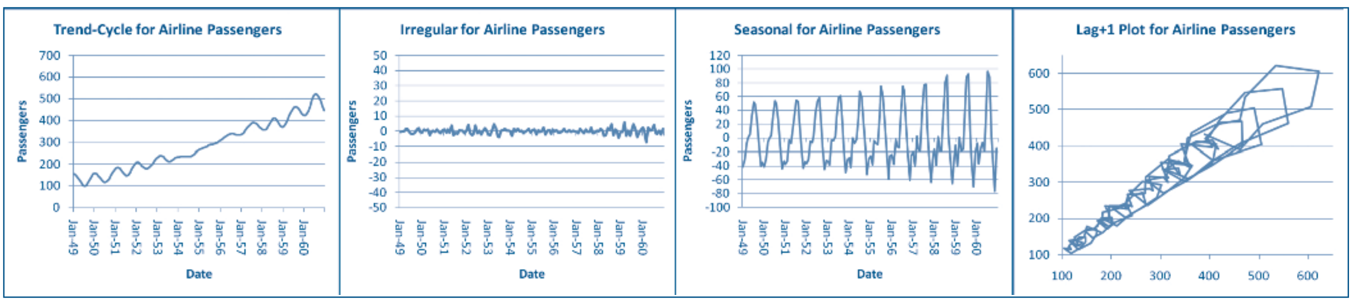
The classical decomposition of a time series into trends, seasonal and other cycles, and irregular movements is intended to reveal the major components of the DGP and the extent to which the process can be forecast based on these components. Decomposition plots show trends, cycles, and movements not attributed to identifiable components ("irregular" movements). A full treatment of decomposition can be found in Makridakis et al. (1998, p. 81). Lag plots can provide further insights into the DGP. A useful introduction to lag plots can be found at ***http://www.itl.nist.gov/div898/handbook/eda/section3/lag-plot.htm***

The Sine Wave



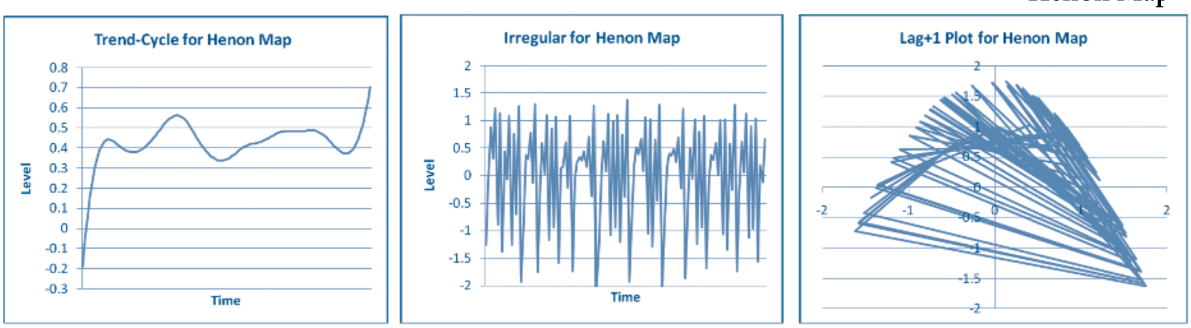
The cycle plot for the sine wave perfectly captures the entire underlying cyclical process, while the plot of the irregular reveals that there is no unaccounted-for behavior to cause forecast error. A lag plot shows the value of a time series against its successive values (e.g. lag +1) . In a lag plot the cycle shows up as a circular pattern with randomness "clouding" that pattern. Here we see a very clear ellipse representing the pure deterministic nature of the series.

Airline Passengers.



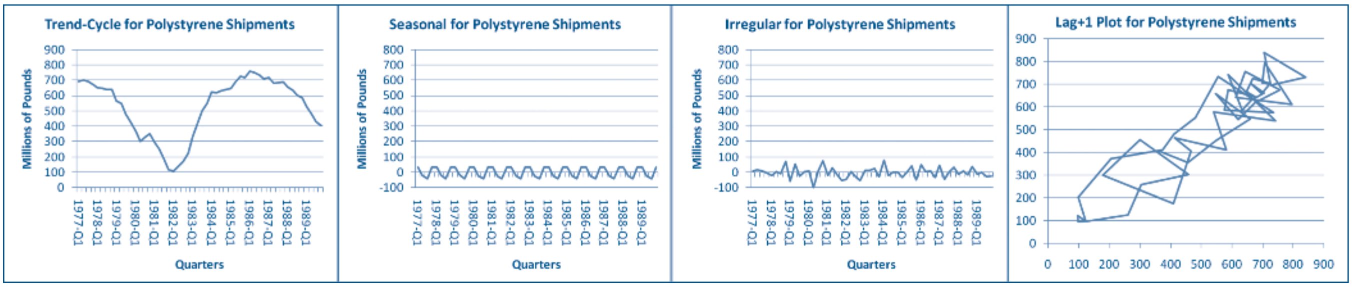
The trend-cycle plot shows a strong upward trend with a lesser cyclic component. The seasonal component for the series shows the strong and growing seasonality, indicating a highly deterministic component of the series. The irregular component not captured by the earlier trend and seasonal plots is fairly slight in comparison, so this series is highly deterministic. The lag plot shows a pentagon-like shape also representing the strong seasonal component.

Hénon Map



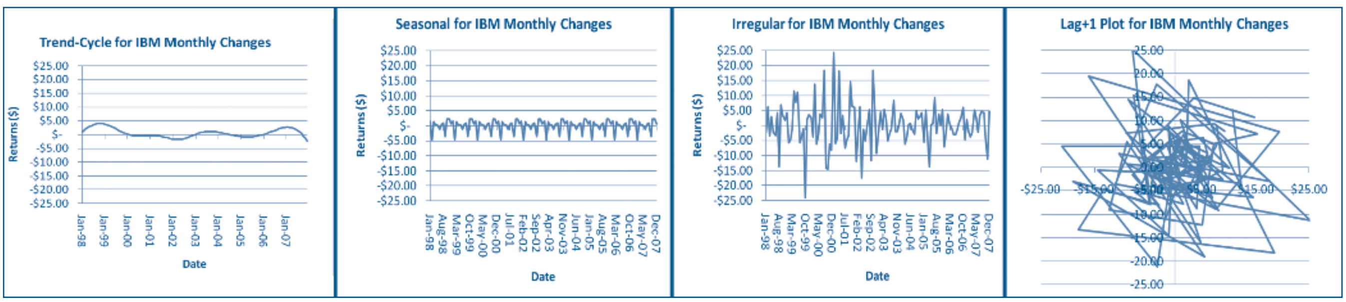
The trend-cycle chart shows a moderate and slow oscillation, suggesting a weak, irregular (aperiodic) cycle. The irregular component is very close to the original series, indicating the series has little in the way of identifiable components using the decomposition method. However, the distinct elliptical pattern in the lag plot reveals the deterministic behavior of the series.

Polystyrene Shipments



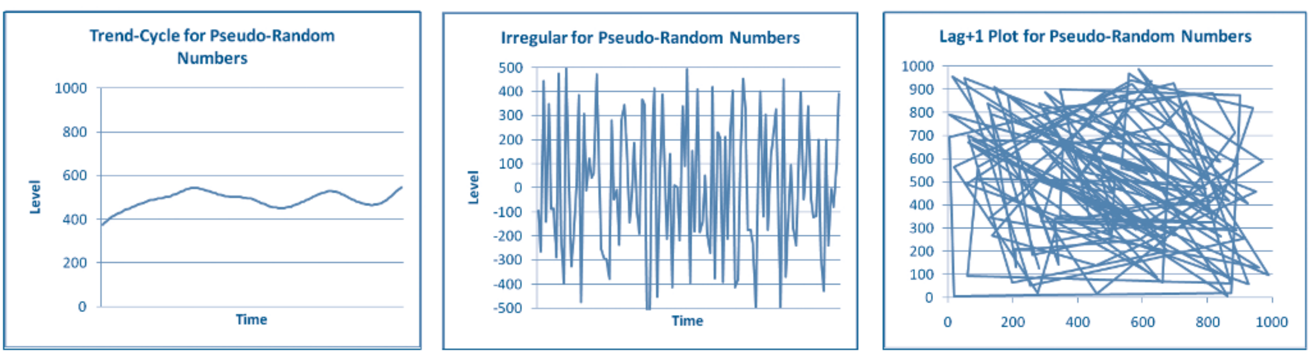
The trend-cycle for polystyrene shipments shows strong yet changing trends, albeit over a considerable time frame. The seasonal component indicates some relatively minor quarterly seasonality. The lag plot shows the cyclic behavior. The irregular component is relatively minor, indicating that this complex series has significant deterministic components.

IBM Monthly Stock Price Changes



The trend-cycle plot, in relation to the range of stock price changes shown in the IBM time plot (-$21 to +$25), shows only minor oscillation around the $0.00 level, indicating little or no meaningful cyclic component. Likewise, the seasonal plot shows minor seasonal fluctuations. The irregular plot is essentially the same as the original series, making this series appear almost purely random. The lag plot does have a concentration of data points around the -$5.00 to +$5.00 region on both axes, and therefore possibly some deterministic component.

Pseudo-Random Numbers



In relation to the range of values shown in the pseudo-random numbers time plot (0-999), the minor oscillation around the 500 level indicates little or no meaningful trend or cyclic component. The irregular component is very close to the original series, indicating the series has little in the way of identifiable underlying components using the decomposition method. The lag plot also shows an obvious lack of pattern, indicating that the series is essentially random.

To summarize: These illustrative time series have moved from completely deterministic (sine wave), to a largely deterministic system (airline passengers) series, to a chaotic process (Hénon map), to complex underlying systems (polystyrene shipments and IBM financial returns), and finishing in a purely random series (pseudo-random numbers).

THE COEFFICIENT OF VARIATION

Ken Kahn (2006) has recommended the use of the “coefficient of variation” as a numerical indictor of the forecastability of a time series. A coefficient of variation (CV) is a measure of dispersion of a distribution around its average and is typically defined as the ratio of the standard deviation to the mean .

$$CV = \sigma/\mu$$

A CV of 0.5 (or 50%) indicates that the series standard deviation is half the size of the series mean. Higher values of a CV indicate greater variability of the data around their mean.

Kahn proposed the following steps for assessing forecastability:

1. Deseasonalize the data, if applicable
2. Detrend the data, if applicable
3. Calculate the mean of the resulting series
4. Calculate the standard deviation of the resulting series
5. Calculate the coefficient of variation (CV)

Steps 1 and 2 are readily performed in spreadsheets and are often available in forecasting software. Deseasonalizing and detrending the data removes the easily forecastable components of seasonality and trend. Without these initial steps, time series with strong deterministic trend and seasonal components, such as the airline passengers series, would yield inflated values of the CV.

I have calculated the CV for the six sample series. Only two of these, the airline passenger series and polystyrene shipments, had to be detrended and deseasonalized.

1. ***Sine Wave***
   Mean: 0.00000030, Std Dev: 0.707,
   CV: 232,104,106%
2. ***International Airline Passengers*** (Deseasonalized/Detrended)
   Mean: 91.80, Std Dev: 18.18, CV: 20%
3. ***Henon Map***
   Mean: 0.426, Std Dev: 0.959, CV: 225%
4. ***Polystyrene Shipments*** (Deseasonalized/Detrended)
   Mean: 457, Std Dev: 200, CV: 44%
5. ***IBM Stock Price Changes***
   Mean: 0.52, Std Dev: 7.67, CV: 1476%
6. ***Pseudo-Random Numbers***
   Mean: 489, Std Dev: 284, CV: 58%

While the graphic analysis suggests that series forecastability deteriorates as we move from the first to the sixth series, the coefficients of variation do not effectively correspond.

- The CV for the deterministic sine wave suggests this series is more difficult to forecast than all the others. A problem here is that when the mean value is near zero, the coefficient of variation becomes highly sensitive and effectively unusable.

- The CV indicates that the pseudo-random number series is more forecastable than the

IBM series and that the Henon Map series is not nearly as forecastable as the polystyrene series.

This lack of correspondence reveals some deficiencies of the CV in assessing forecastability.

First, the Kahn procedure assumes that, once the series is detrended and deseasonalised, there are no further patterns or structure to identify. In other words, any pattern in the series with trend and seasonality removed is considered unforecastable. This is not the case in the chaotic Hénon Map series, or the complex polystyrene series and IBM series, which have other components besides seasonality and trend.

Second, the detrending and deseasonalizing proposed by Kahn is done in a specific way (classical decomposition), which may not be the best way to extract these components. For example, Winters exponential smoothing extract trend and seasonal components using optimized weights rather than the equal weights of the classical decomposition.

Third, the CV is misleading when the series mean value is close to zero, as shown in the sine wave series.

Last of all, there is ambiguity in how the mean of a trending series should be determined. If we have a positively trending series such as Airline, should we use:

- today's level (implying a trend component that moves from a large negative number to about zero)?
- the beginning level (trend component goes from zero to a big positive number)?
- something in the middle?

So the mean of the detrended Airline series could be calculated respectively as about 100, about 500, or about 300, leading to coefficients of variation that can differ up to fivefold.

In conclusion, the relative magnitudes of the CV for different series can give a misleading representation of the relative forecastability of the series. A more reliable metric than CV is proposed in the next section.

INFORMATION-THEORETIC METRICS: COMPRESSIBILITY AND ENTROPY

As far back as the 1940s, information-theoretic concepts were put forward by Claude Shannon (1948) as a means of assessing the complexity of data patterns.

Algorithmic Information Content. One of the simplest ways to represent data complexity is the algorithmic information content (whose initials AIC are not to be confused with Akaike's Information Criterion). This AIC measures the compressibility of a string of numbers that is normally transformed to binary before compressing. For example, the binary sequence "1100110011001100110011001100" can be compressed to read "repeat '1100' 7 times." For this series, the AIC would be minimal, indicating high compressibility.

In contrast, random strings are uncompressible: for example, the minimum description length for "1110101010101101100010111110" is "1110101010101101100010111110" itself, resulting in a maximum AIC.

How does compressibility relate to forecastability? Compressibility metrics give us the ability to measure the deterministic structure versus the randomness of a series. A purely deterministic series can be fully compressed while a purely random series is uncompressible.

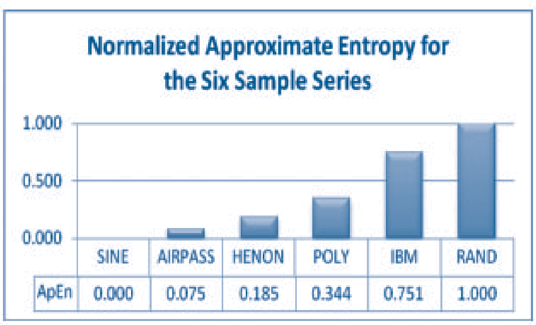
Unfortunately, the AIC has a major flaw: it equates complexity with randomness, thus failing to acknowledge the potential forecastability of complex systems. This failing is similar to that of the CV which equates variability of a detrended/deseasonalized series with randomness.

A more useful foundation than compressibility is *entropy*. In physics, entropy measures the amount of "disorder" of a system. A system with high entropy has a high degree of disorder or irregularity. A system with low entropy is more orderly or regular. Intuitively, the presence of repetitive patterns in a time series renders it more orderly than a time series in which such patterns are absent.

Normalized approximate entropy metric (ApEn). Approximate entropy measures the relative frequency with which blocks of length m that are close together remain close together in the next incremental comparisons, m+1. Small values of ApEn imply strong regularity, or determinism, in a time series; large values of ApEn imply substantial irregularity, or randomness. ApEn has also been found to be a robust measure of regularity when applied to time series that contain noise and outliers (Pincus & Singer, 1996, p. 2083).

The table here shows the values of ApEn for the six sample time series. The values have been normalized to provide a common scale with a range of zero (SINE) to one (RAND). An appendix with the calculation algorithm can be accessed at www.forecasters.org/foresight/documents/catt_issue13.pdf

Approximate Entropy of Sample Time Series (m=2, r=0.5)



The results conform with the forecastability expectations from the graphic analysis. The completely deterministic sine wave has a normalized ApEn of 0.000 (completely deterministic) and our stochastic pseudo-random series (RAND) has a normalized ApEn of 1.000 (random). The airline passenger (AIRPASS) series has a very high degree of regularity with an ApEn of 0.075, very close to zero. The Hénon series is quite regular at 0.185, with the polystyrene shipments (POLY) at 0.344 being a moderately regular series. Based on an ApEn of 0.751, IBM returns are irregular and sit relatively close to our pseudo-random series.

Although a number of variations of ApEn have been developed, the general principle of measuring the regularity of time series features is similar. Unlike the CV, entropy metrics do not assume that only obvious patterns such as seasonality and trend are important, but reveal any form of regularity. Nor are they distorted when the series means is close to zero.

For more insights into ApEn, see Stephan Kolassa's commentary, which follows this article.

ENTROPY AND RELATIVE FORECASTABILITY

In a recent paper on climate prediction, Green and colleagues (2009) define forecastability as the ability to improve upon a naïve benchmark model. The most common of these is the naïve 1 or no-change forecast, which uses the most recent observation available as the future forecast. In principle, in the case of purely random movements, no improvement on the forecast accuracy of a naïve model is achievable.

As Figure 2 has shown, the ApEn metric offers an empirical counterpart to the foregoing definition. For a completely random series (RAND), the ApEn has a value of unity. For a completely deterministic series, the ApEn has a value of zero. Hence the degree to which ApEn falls below unity provides an indicator of the degree of forecastability of the time series. Alternatively stated, the value of 1-ApEn assesses the degree to which we can expect to improve upon a naïve benchmark.

A caveat: analysis remains to be done about the relationship between ApEn and common measures of forecast accuracy such as the MAPE. At this stage, we should not assume a one-to-one relationship in which a particular value of 1-ApEn (such as the 1-0.344 = 0.656 for POLY in Figure 2) translates into the percent reduction in the MAPE that is possible compared to a naïve benchmark.

CONCLUSION

The key questions are: What is the underlying process that we're attempting to forecast and how predictable is it? I have tried to provide a scientific basis to the assessment of forecastability. Time plots, decomposition plots, and lag plots can all be useful techniques in revealing forecastable patterns in time series. Combining the insights from charts with information-theoretic measures such as approximate entropy can help establish where a time series sits on the deterministic-random continuum, and therefore provide an indication of the relative forecastability of the series.

REFERENCES

Box, G. & Jenkins, G. (1976). *Time Series Analysis: Forecasting and Control*, San Francisco: Holden Day.

Bricmont, J. (2003). Determinism, chaos and quantum mechanics. Retrieved February 02, 2008, from ***http://www.fyma.ucl.ac.be/files/Turin.pdf***

Corning, P.A. (2002). The re-emergence of "emergence": A venerable concept in search of a theory, *Complexity*, 7(6), 18-30.

Green, K.C., Armstrong, J.S. & Soon, W. (2009). Validity of climate change forecasting for public policy decision making, *International Journal of Forecasting* (forthcoming). See ***http://kestencgreen.com/naiveclimate.pdf***

Kahn, K.B. (2006). In search of forecastability, *Forecasting Summit* 2006, Orlando, Florida.

Makridakis, S., Wheelwright, S.C. & Hyndman, R.J. (1998). *Forecasting: Methods and Applications* (3rd ed.), Hoboken: New Jersey: John Wiley & Sons.

Pincus, S. & Singer, B. (1996). Randomness and degrees of irregularity, *Proceedings of the National Academy of Sciences*, 93, 2083-2088.

Popper, K.R. (1988). *The Open Universe: An Argument for Indeterminism*, London: Routledge.

Shannon, C.E. (1948). A mathematical theory of communication, *Bell System Technical Journal*, vol. 27, July & October, 379-423 & 623-656.

Strogatz, S.H. (1994). *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*, Reading, MA: Addison-Wesley.

Contact Info:
Peter Catt
Soltius
Peter.Catt@soltius.co.nz

Bio: Peter Catt is a Supply Chain Solution Manager with New Zealand's largest dedicated implementer of Enterprise Resource Planning systems. Prior to becoming involved in ERP systems implementation, Peter held senior operations management positions in the sawmilling, building products, and food industries. He holds a Doctorate in Computing, an MBA, and a postgraduate Diploma in Industrial Production.