



GRESHAM COLLEGE
Founded 1597

The perverse and the reverse: how bad measures skew markets

Transcript

Date: Monday, 17 October 2005 - 12:00AM

Location: Barnard's Inn Hall



THE PERVERSE AND THE REVERSE: HOW BAD MEASURES SKEW MARKETS

Professor Michael Mainelli

Good evening Ladies and Gentlemen. I'm pleased to see so many of you turn up for a lecture on such a pernicky subject as measurement. However, as attendance is one of our key measures of success, perhaps regardless of the quality of my lecture, we've already succeeded. Or is this something for us to explore further this evening?

Well, as we say in Commerce - "To Business".

Taking the Measure

Tonight's lecture examines measures. We're going to cover a few principles such as Goodhart's Law and Gresham's Measurement Corollary. We're also going to explore what makes measures bad or good. But why do measures matter in the first place? I can only ask you to imagine a world without commercial measures. You are ready to buy or sell something - but you have no idea whether the amount you're told you're being sold is correct. Were the weights and measures tampered with? Is the petrol flow at the station pump reading correctly? Is the currency you're being paid with valid, of value, convertible, at what rate? You have no idea who the counterparty really is, who vouches for them, what are the payment terms, where will disputes be adjudicated? A market without measure is a market without reason. Measures matter.

One of the less disputed areas of the government versus private sector divide is that the majority of people give government a prominent role in running the regimes of measurement. For instance, Section 8, Clause 5 of the US Constitution specifically gives Congress the power "To coin Money, regulate the Value thereof, and of foreign Coin, and fix the Standard of Weights and Measures." People everywhere expect government to set standards and measures, to take censuses or to provide national statistics. But ill-thought-out measures do lead to the perverse and the reverse, and the subjects of this talk, at the levels of consumers, of firms and of governments.

Let's start with the Law of Unintended Consequences. There are many versions of this law, but mine is that 'action to control a system will have unforeseen results'. This is a law we hate to love, but love it we do. The underdog supporter in each of us experiences a small thrill when we see authorities humbled trying to control things we're not sure they should have a say over. Supposedly, when authorities in the state of Vermont tried to control roadside advertising in 1968 by banning billboards, instead of obtaining clear views of the countryside they found that local businesses had acquired such intense artistic interests that an auto dealer commissioned and installed a sculpture of a twelve-foot high, sixteen-ton gorilla clutching a Volkswagen Beetle, and a carpet store erected a nineteen-foot high genie holding aloft a rolled carpet as he emerged from a smoking teapot. On the other hand, Adam Smith's Invisible Hand is a rarer case of the Law of Unintended Consequences having a beneficial effect - huge numbers of selfish people create mutual benefits.

We know a lot about how to specify objectives and targets, but measures are always at the core. There is a lovely acronym for one straightforward approach to setting objectives, the SMART objective. A SMART objective has five characteristics - specific, measurable, achievable, realistic and timely. As an example, think of a call-centre:

- .. specific: a number, percentage or frequency should be used - 'answer the phone within 10 seconds' is much clearer than 'answer the phone promptly';
- .. measurable: the measures must be taken consistently and communicated - 'you don't seem to be answering the phone as quickly as yesterday' is not adequate measurement;
- .. achievable: the objective must be realistic for a reasonable amount of effort - you can't ask people to answer the phone always within one second given they have to breathe or drink;
- .. relevant: the objective must be within people's control - you can't ask call-centre agents to increase market share;
- .. timely: you must clearly set out timescales - e.g. did you mean calls within the next month?

However, this lovely SMART edifice falls down if we set the wrong measure. Let me give you one example. I recently encountered a financial services organisation struggling with its helpline. In order to improve efficiency, the call-centre agents were given time targets, such as 'so many minutes to be spent per customer query'. As bonuses were tied to the targets, the call-centre agents rapidly found they could hit the minutes-per-customer target fairly accurately. It didn't matter that many of the big, profitable transactions, e.g. "I'm thinking of taking out a mortgage", or the ones that matter to customers, e.g. "I need to change a standing payment", took more than the allotted time because the call-centre agents mysteriously found that the call had come to an end. Of

course, the poor customer would retry a few times before concluding that the one transaction that could occur within the narrow time window available was terminating his or her account.

So the financial services organisation noticed two things – first, the popularity of its helpline was rising rapidly and, second, that customers were leaving in droves. Of course action had to be taken. So a special, second helpline was set up to catch customers who made noises threatening to leave. As soon as customers threatened to leave, they were transferred to the special, second helpline. This second helpline was given two special powers, the first was a small amount of money to ‘grease’ the path for the customer back into their high quality service. And the second power? An unlimited amount of time to deal with the query. Of course, with much higher call volumes the regular agents were now stretched handling return phone calls from customers, so their available time per customer needed to be reduced. Strangely, while the special helpline was reasonably successful at retaining customers, the rate of customers threatening to leave increased even more rapidly.

Naturally, the loss of business had to be investigated further, so an ultra-special third helpline was set up to find out why customers were dissatisfied. This third helpline had one role, just to listen. Completely unlimited time to talk to customers. In a final ironic twist, the Chief Executive of this company was in the media talking about the need to “get close to the customer” and talking about special research he was commissioning. It doesn’t take a genius... Nevertheless, while this case was sorted out in a few weeks, we all know of many cases where organisations fail to get on top of these problems and they escalate out of control.

We do love these tales showing that ‘action to control a system will have unforeseen results’. First, we thrill to see plucky call-centre agents triumph over the wicked control system, albeit by ‘throwing a spanner in the works.’ Second, we enjoy the satisfaction of natural justice as disconnected managers get their comeuppance. We tend to ignore the huge amounts of wasted time for customers and call-centre people or the potential loss of jobs as the financial services firm is clearly less competitive.

Is there anything we can learn about avoiding these unintended consequences? Well, one thing we learn is that there is a big difference between measures and targets. Measures are, in theory, isolated facts. Targets are things where some remuneration or incentive is implied. I might have an argument that there is such a thing as an objective, non-human measure, but I certainly don’t have a target that doesn’t involve people and the incentives that motivate people. So how can we start to understand measures and targets. We turn to Gresham’s Law for a starting point.

Weighed and Found Wanting

The most accurate expression of Gresham’s Law is given by Professor Robert Mundell as “cheap money drives out dear, if they exchange for the same price” [1]. For instance, if we force less valuable silver coins to have the same value as gold, but there is an outside market for gold and silver, then the dearer gold coins will be used in the other market leaving us with cheap silver in the one we are debasing. There are numerous examples of control systems trying to force cheap and dear to exchange for the same price, e.g. bimetalism and recoinages where the good money fled to places it was properly valued. Again, the Law of Unintended Consequences takes hold - ‘action to control a system will have unforeseen results.’ Again, we have the plucky yeoman smuggling coinage to a country where he gets full value and we savour the woefully disconnected government getting its comeuppance.

What has this got to do with call-centre measures? Well there is a version of Gresham’s Law operating here; may I label this Gresham’s Measurement Corollary – “a cheap measure drives out a valuable measure, if they exchange for the same price”. In the case of the financial services call-centre we see that the agents are not rewarded for valuable measures such as happy customers or increased customer revenue, so they use the cheap measure of minutes per call. As the call-centre agents are paid for the number of calls, they focus on this and other, better measures suffer. While I recognise that the valuable measures are more difficult to track, it is easy to see how Gresham’s Measurement Corollary leads to the fun of the Law of Unintended Consequences. By taking the easy measure, we lose.

There used to be a lot of Soviet-era jokes about measures taken to the absurd. Once upon a time there was a steel factory that made railroad equipment, especially locomotives. The target was changed to output as much steel as possible, so the managers made the locomotives with as much steel as they could, so heavy the locomotives broke the rails. So central planning converted the locomotive factory to focus on making rails for the railroads. The new target was set as the maximum number of pieces of finished steel, so the managers decided to make exceedingly thin rails, as thin as pins. So central planning decided to convert the factory to make pins, but the railways were too weak to deliver the ore or ship the pins, etc. Yet we post-Communists make similar mistakes. The UK government has been frequently pilloried for its perhaps overly-enthusiastic use of targets. The Economist noted in a leader in 2001 [26 April 2001, “Missing the Point”]:

“At worst, targets create “perverse incentives” - when workers are cleverer than targeters, and find ingenious, and not necessarily desirable, ways to meet their targets. That is why, for example, the government’s commitment to reduce the hospital waiting list is now widely discredited. The target, cutting the number of people waiting for treatment by 100,000, has been met. But the number of people waiting to see a specialist -

waiting to be put on the waiting list, in other words - increased.”

In the relative absence of market forces, the Law of Unintended Consequences seems to recur in the National Health Service with alarming frequency, hand in hand with examples of Gresham's Measurement Corollary. For example, we were told during the recent election that we could always see a General Practitioner within 48 hours of asking for an appointment, the target. The bad information was driving out the good and the government was unable to find out what was happening, because no General Practitioner would make an appointment more than two days hence. I once had the dubious pleasure of watching a UK government minister getting a dressing-down after he smarmily assured a room of business people that there were only a half-dozen targets in the entire National Health Service. The business people easily managed to recite over 20 targets off the top of their heads in under a minute. The final discussion was a Kafkaesque skit on what did we mean by a measure and how could we measure the measures. We could expand on this theme of government's difficulty in setting targets with examples in the UK of fudged A-levels affecting everyone's perceptions of higher education or the fact that the government pays all bills promptly to small businesses as soon as the paperwork is in order, though payments well exceed 30 days because the vast majority of incorrect paperwork originates in the government bodies. We could expand on this theme across Europe as agricultural subsidies distort investment decisions in transportation, energy and the environment, or in the USA with the Jones' Act reducing intra-coastal shipping. However, time doesn't permit me to explore what's worse - a lack of measure, where things such as pension deficits aren't measured, to the point where they exceed 50% of government debt, but aren't recorded at all.

With mundane measures, it always starts with good intentions, “we really want to measure such and such to ensure that we meet our objective”, a golden measure. But often the really valuable measure that might deliver our objective is beyond our ken, difficult to define or difficult to measure. So we find a cheap proxy for the valuable measure - some paper target that will more or less do. While the people we are trying to measure will almost certainly change their behaviour to meet their targets, they almost certainly will not look beyond the measure to the original objective. After all, it is much easier, and cheaper, to deliver the measure that's been set.

Man Is The Measure

Around 450 B.C., Protagoras of Abdera, a Greek Sophist, noted, “Of all things, the measure is Man; of the things that are, how they are; and of things that are not, how they are not.” While this is an intensely anthropo-centric statement, this statement is of fundamental importance to Commerce. Human beings may be regarded as insignificant in the scale of the universe or indistinguishable at a cosmic scale from other life, but in Commerce we believe that the ultimate determinant of value for markets is the value given by people. We have seen how human interests have driven markets for cloves, nutmegs, pepper, tulips, tea, coffee, sugar and precious metals. We lament the decline in literacy, but then race home to read the latest thrillers other people are reading, because our key measure is a book's position on bestseller lists, i.e. what are other people reading. In fact, the rise of Western society during the period from 1250 A.D. to 1600 A.D. may well be due to the need of Western society to quantify in order to simplify the chaos of a cacophony of measures and currencies. Alfred Crosby remarks that, “the epochal shift from qualitative to quantitative perception in Western Europe ... made modern science, technology, business practice and bureaucracy possible.” Markets are anthropo-centric.

Measures that people care about should be the measures that matter for markets. Or, going the other way, measures that matter for markets must be those about which people care. Moreover, markets cannot function without measures. And markets pay a lot for measurement, whether it be to auditors, actuaries, rating agencies, laboratories, standards agencies and inspectors, or lawyers when things go wrong. Markets get a lot from measures. Probably two big things worth pointing out are that measurement reduces information volatility and information asymmetry. If I can trust the moisture content measure in the cargo of wheat you are trying to sell me, we have all saved a lot of time, bother and unnecessary calculations about future volatility in outcomes. If I can trust a third-party who either established a measure or took a measure, then I have reduced somewhat the information asymmetry in our relationship. In fact, the evolution of a market from an innovative product or service, to a widespread product or service, to a commodity, is frequently accompanied by an evolution of measurement. Breaking from a commodity back to an innovative product or service is often accompanied by a revolution in measurement.

As an example, take telecommunications. In the very early days there were a variety of billing methods. Globally, the industry evolved to the point that most 20th century telephone bills were based on measurements of time, distance and national borders. With the long payback on the fixed assets, these charges became increasingly divorced from the underlying costs of delivering telecommunications and a number of people predicted the ‘death of distance’. Nonetheless, telecommunications became a commodity with its own bandwidth exchanges measured by time and distance. For a time, static measurements stifled innovation - new suppliers found that traditional measurements squeezed them between existing supply measures and new propositions to customers, with unfortunate results. Over time, technical innovation led to demand for a flat charge for internet usage, so data bills were based, initially at very high charges, on connectivity. This led to unprecedented innovation to the point that we find the telecommunications industry in turmoil as Voice Over Internet Protocol (VOIP) starts to cannibalise traditional telephony. A cycle from innovation to commodity and

back again, accompanied or instigated by changes in measurement. On the other hand, a market that relentlessly sticks with an outmoded measurement can avoid evolution. The measure skews the market and the market skews the measure.

'You can't manage what you don't measure'. This old management adage is specifically correct, but often fails us in practice, particularly in areas where the thing being measured is complex. Some good examples of complex measurement areas include the quality of customer relationships, the quality of research or happiness. As Robert Pirsig noted in Zen and the Art of Motorcycle Maintenance:

"I think there is such a thing as Quality, but that as soon as you try to define it, something goes haywire. You can't do it." [Pirsig, 1974, page 184].

Managers should abide by the rule of thumb 'that which gets measured gets done' but realise that they may fail to find great measures that can handle the intangible or the serendipitous. Descriptive measurement, e.g. the Pyramids are x height, would seem to be different from success measurement, e.g. the Pyramids were well-suited to their purpose. The unbounded nature of direction, i.e. we may change where we are heading or what we are becoming, complicates measuring success. Unbounded goals are "the ultimate, long-run, open-ended attributes or ends" and "are not achievable" [Hofer and Schendel, 1978]. Before you have achieved your goals, you have often established new ones. Goals may not be directly measurable (e.g. profit *maximisation* – how do you know this is the most profit you could have made?).

'Accuracy' is integral to measurement. Accuracy raises a key question of 'knowing' the results of measurement:

"All you know is that the length [of a piece of steel] is accurate to within such and such a fraction of a millimetre, and that it is nearer the desired length than anything measurably longer or measurably shorter. With the next improvement in machine tools you may be able to get a piece of steel whose accuracy you can be sure of to within an even closer margin. And another with the next improvement after that. But the notion 'exactly six millimetres', or exactly any other measurement, is not something that can ever be met with in experience. It is a metaphysical notion." [Magee, 1973, pp.27-28]

Protagoras understood that measures are not as objective as we like to think. All measurement is political. Measurement seems to take place by (1) assessing against a standard or (2) comparing against like objects or (3) comparing against a prediction or model. These three means of measuring, (1) standard-based, (2) comparative, (3) predictive, can be fused, e.g. average the comparisons and one has a standard, or consider a prediction to be just another comparison. The three means of measuring are worth considering independently as (1) is about an absolute measure, (2) is about conditional measure and (3) focuses on expected outcomes. Measurement immediately begs the questions, by whom and for what end.

So what do people do with measures? I would contend that most measurement is in aid of one or more of four general purposes:

- .. set direction – we intend to achieve X;
 - .. gain commitment – we agree with our audience that they should expect us to achieve X;
 - .. keep control – we know what our key measure is and how 'on target' we are;
 - .. resolve uncertainty – we believe in X as a key measure, regardless of some of the volatility around us.
- Be confident.

Here's an interesting statement from Warren E. Buffett that embodies all four purposes:

"Our long-term economic goal (subject to some qualifications mentioned later) is to maximize Berkshire's average annual rate of gain in intrinsic business value on a per-share basis. We do not measure the economic significance or performance of Berkshire by its size; we measure by per-share progress. We are certain that the rate of per-share progress will diminish in the future - a greatly enlarged capital base will see to that. But we will be disappointed if our rate does not exceed that of the average large American corporation."

[Warren E. Buffett, "An Owner's Manual" for Berkshire's Class A and Class B shareholders, June 1996.]

Note that Buffett has incorporated all four purposes:

- .. he has 'set direction' as this "average annual rate of gain in intrinsic business value" is a long-term economic goal;
- .. he has 'gained commitment' from shareholders by focusing on a per-share basis;
- .. he has 'kept control' by showing that he realizes increasing scale will make the measure more difficult to achieve;

.. he has 'resolved uncertainty' by acknowledging that Berkshire Hathaway's "per-share progress will diminish", but also by comparing Berkshire Hathaway's performance to the average large American corporation. If Berkshire Hathaway fails to beat the average, then things are uncertain, otherwise, they're on target.

An obvious, but crucial, point is that people really mess up measures. People mess up measures in two big ways, they love single numbers and they are obsessed with meeting targets.

Wanna BET%?

Selecting measures is fraught with opportunity and danger. A good example of seizing opportunity is the example of a soft drink distributor whose country managers insisted that they had largely reached market saturation. In a savvy move, the soft drink distributor moved country managers' goals from the market share of processed drinks to the market share of the throat - what percentage of each country's liquids consumption had been through their bottling plants before going down consumers' throats. Measurement redirection of this sort will work at least once.

On the other hand, selecting measures can be dangerous. Sticking with insurance, let me present you with a hypothetical test. You are the Chief Executive Officer of a large insurance company affected by an unusually large hurricane. You are being hounded by the press to 'come clean', to declare the effect of the hurricane on your company. Your press office gives you two draft press releases from which you can select your company's official position:

.. Press Release A - our losses will be £Q, a specific amount;

.. Press Release B - at this point in time our losses are estimated to be between £X and £Y, probably around £Z, at a probability of P%.

Which one do you choose, and why? [HANDS UP] Let's look at what really happened after some recent disasters.

After the Twin Towers disaster of 11 September 2001, a Lloyd's press release a fortnight later on 26 September 2001, said:

"The estimate at this time shows that Lloyd's net exposure arising from the attacks is £1.3 billion (US\$1.9 billion). This exposure is equivalent to 12 per cent of the market's 2001 capacity."

Berkshire Hathaway's press release from the day following the tragedy, 12 September 2001, said:

"Berkshire Hathaway cannot now make an intelligent estimate of industry-wide losses. Historically, Berkshire's share of catastrophe payments ... has been in the range of 2-3% of the industry's total losses. Berkshire's present guess is that it will incur 3-5% of the industry's loss in this catastrophe due to a different mix of policy coverages than that which prevails in natural catastrophes."

So Lloyd's announces a specific number of \$1.9 billion two weeks after Berkshire Hathaway says things will be in line with the industry. Who's the canner? Well, putting my neck out a bit, I suggest that Berkshire Hathaway is clearly the canner. First, they haven't boxed themselves in because they used a wide range from 3% to 5%; 5% is 167% of 3%. Second, they have made their measure comparative with the industry, not specific to themselves or an absolute number. Further, the range is based on an overall figure, "the industry's loss", that is also somewhat conveniently indefinite. Berkshire Hathaway have a lot of room for management and manoeuvre. In contrast, Lloyd's is pinned to a specific number. Sure, Lloyd's have said it's provisional, "at this time". Equally, Lloyd's could have said they're over 99% likely to be wrong. The odds of the market's net loss being exactly \$1.9 billion are close to zero. It's almost impossible to provide a measure of losses from an event such as a terrorist incident using just a single number only two weeks afterwards. In the event, Lloyd's spent a lot of time adjusting and revising the figure as more information emerged, while Berkshire Hathaway's measure turned out to be 'fit for purpose'. By January 2002 Lloyd's admitted to paying out \$454 million, by May \$1.1 billion and by December 2002 estimated its ultimate loss at \$3.26 billion (£2.02 billion) and some 150% of the estimate.

Just to underscore that the Lloyd's approach is consistently specific, Hurricane Katrina hit New Orleans early on the morning of 29 August 2005. A Lloyd's insurance market press release two weeks later on 14 September 2005, said:

"Lloyd's provisional estimate is that the market's net loss as a result of Hurricane Katrina is £1.4bn (\$2.55bn)."

On 19 September 2005, a Berkshire Hathaway Inc. press release announced:

"Due to the extraordinary devastation created by Hurricane Katrina, it is particularly difficult to estimate an industry loss for this event and we don't intend to at this time. However, Berkshire Hathaway has

previously stated that it expects its share of industry losses from catastrophes such as Hurricane Katrina to be 3-5%. Berkshire continues to believe this to be true and thus expects it will incur 3-5% of the industry losses associated with Hurricane Katrina.”

Again, I forecast that Lloyd’s is likely to come under pressure as it is forced to update this figure and Berkshire Hathaway is unlikely to be held to account. Insurance companies will always be pressured to estimate losses long before they can make intelligent estimates. In practice though, Berkshire Hathaway’s approach has been more realistic and canner than Lloyd’s. In fact, Berkshire Hathaway is more truthful, any loss estimate is a range. While Lloyd’s comes under pressure during any revision - if upwards, clearly there is even worse news to come; if downwards, they are not on top of things, tried to manipulate opinion and are probably overly optimistic - Berkshire Hathaway just has to be within the range. The Economist proffers an old rule, ‘predict a figure or a date, but never both’ [25 April 1992, page 91].

If nothing else, I hope that the next time someone pressures you for a measure after this lecture, you do not respond with a single number. I’ll leave you with another acronym, BET%. Think about the risks and the rewards of a single number versus a range. The acronym stands for Bottom, Expected, Top and the % likelihood of anything happening. If your boss asks you what sales will be this year, instead of saying £5million, far better and safer to say bottom: £4million, expected: £5million, top: £6million, at a 90% likelihood, particularly if this is your target. Are you off target if instead of £5million of sales you hit £4,999,999?

Goodhart’s Law

Let’s turn to the second big way in which people mess up measures - an obsession with meeting targets. I contend that a measure associated with an incentive becomes a target. Professor Charles Goodhart was Chief Adviser to the Bank of England when he formulated an observation on regulation, Goodhart’s Law. The original form of Goodhart’s Law was:

“As soon as the government attempts to regulate any particular set of financial assets, these become unreliable as indicators of economic trends.”

Goodhart’s initial observations emerged from monetary policy and regulation. As he noted, “financial institutions can... easily devise new types of financial assets.”

Goodhart’s original expression evolved to his preferred formulation:

“Any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes.”

Professor Marilyn Strathern restates Goodhart’s Law more broadly:

“When a measure becomes a target, it ceases to be a good measure.”

An example of Goodhart’s law in action has been the focus on benchmarks for investment managers. Initially, the statistical measurement of investment managers made a lot of sense. Investment management benchmarks emerged, from Morningstar to comparisons with the S&P 500. No problem. However, investors frequently have conflicting objectives. For instance a typical investor might wish to simultaneously (a) beat the annual performance of a benchmark index (b) minimize the variability of annual returns and (c) avoid significant losses. No one will have the guts to point out that (a) “beat the annual performance of a benchmark index” conflicts with (b) and (c). While luck might permit all three to occur in any particular year, over time, all three will conflict. If you want to beat a benchmark you have to accept either higher variability of returns, significant losses, or both. How can you ostensibly provide all three objectives without contradiction?

So beating a benchmark moved from a statistical measurement to a target because investors added the incentive of attracting their money to the measure of the benchmark. Investment managers needed to beat the target in order to attract investors. Investment managers realized that investors, e.g. pension fund managers, wanted excess return over the benchmark, ‘without risk’. While this excess return over an index (often known as ‘alpha’) was desirable, it was also technically impossible. But honesty can be a poor sales policy. In order to attract investors, investment managers found themselves stuck either having to breach some technical portfolio issues or duration or credit quality. The easiest thing to breach has been credit quality. For instance investment managers may add to their benchmark portfolio a one-year corporate security that pays 100 basis points more than an equivalent one-year Treasury bond. The security and the bond have the same duration, but the credit quality is completely different. As long as the corporate security pays out, the investment manager is able to claim he or she beat the benchmark by delivering more than ‘alpha’. Though corporate securities rarely fail, if this happens, then the investment manager folds up his or her tent and walks away. A small change in credit quality allows the benchmark to be beaten. Goodhart’s law is alive and well. Because the benchmark became a target, it ceased to be a good measure. In fact, we find large numbers of investment managers that beat particular benchmarks, but we are unable to evaluate them without taking account of survivor bias.

Goodhart’s Law bears gentle comparison with Heisenberg’s Uncertainty Principle. Heisenberg noted that the

more precisely the position of a particle is determined, the less precisely the momentum is known. Measuring a physical system disturbs it. In biology, measuring a cell may well kill it. While I am loath to apply lessons from physical systems directly to social systems, it is true that measuring social systems disturbs them and alters their behaviour. If measurement did not alter behaviour, we would hardly see such an emphasis among politicians, managers and others on measurement. In some ways, the more precisely we try to measure things, the worse things become. One of the things I have noticed over the years is that an expert is frequently someone who is very good at understanding what are the appropriate measures in a particular situation. It can be a delight to watch a professional at work as they pursue a line of questioning – how do you define utilisation? how do you incorporate overtime in utilisation? does utilisation include all working days, or just days that people turn up for work? are bonuses based on utilisation or on value-added? etc.

So why do firms measure what they measure? In part they measure what they can. Sometimes this is fine, sometimes it is a big mistake. A lot of current thinking tries to get managers to think as scientists, how can we measure impact and what measures relate cause to effect. The measurement of impact may relate to long-term goals and the structure may be fairly stable, but the measures of cause and effect should change over time. As we learn what causes success, it should become normal; as we learn what causes failure, it should be made rare. In turn, new measures should evolve.

Quis Metit Ipsos Metatores?

Who measures the measurers? Poor thinking about measurement is at the heart of a number of looming problems in markets today. To take just one example, we often hear about the hidden reserves of German companies. These reserves arise from a number of timing differences between actual profit and taxable profit. German companies do, as do many companies the world over, keep different accounting statements for different purposes, estimating profit, managing cashflow or supporting tax calculations. German tax is rather generous in the near term with allowances that German companies love to take and consequently firms build up substantial reserves that they don't show on their statutory accounts. German banks realise this and lend accordingly. At the same time, the global banking industry continues to become more open and it turns out that the ostensibly weak balance sheets of German companies now require German banks to either price their loans to German companies higher, or retain higher reserves of their own. In turn, this leads to less lending to German companies because they have ostensibly weak balance sheets, or to a stronger position for foreign banks with less exposure to Germany over German banks with local knowledge. All of this is exacerbated by German companies adopting the new measures in the International Accounting Standards, and German banks adopting new Basel II capital measures, at the same time. So a bad measure of profit skews the accounts and in turn skews the measure of bank capital and in turn skews the banking market.

So what can we do?

First, if we apply BET% to a number of commercial measures, we see the importance of ranges. Take one fundamental area of commercial measurement, financial accounts. Accounting measures are presented as specific numbers, not ranges. We laugh at the financial accounts of large firms being accurate to the last penny. How absurd, we remark, when we all know that no company could ever be so accurate to even a rounded profit of £46,205,182. Yet we don't follow through on the obvious implication, a specific number is the wrong measure. Too many things in profit, as in all accounting statements, are ranges, from the estimate of gains in freehold land value to the likely profit on individual contracts to the value of insurances, etc. To ensure total clarity, we litter the financial accounts with explanatory footnotes to the point that only highly sophisticated financial analysts can understand them. When the accounts are presented, these financial analysts tear them apart in order to try and re-build estimates based on ranges. Intriguingly, the auditors get off very, very lightly, practically skipping away. How do you hold an auditor to account? Is being off by £1 enough to claim the accounts are invalid? Certainly not. £2? Well, when? In fact auditors have cleverly avoided giving us anything substantive to go on, such as "we are 95% certain that profits were between £X and £Y". Let's think about forcing auditors to lay these ranges out clearly. In fact, let's pin down all commercial measurers to their estimates using BET%. Perhaps I can venture to say, in honour of that famous statistical cleric, that their 'Bayes are numbered'.

Secondly, we need to be more flexible. Some areas are perhaps more amenable to softer measures. A good example comes from scientific research where crude measures such as published papers can be 'gamed' more easily than competitive peer review. Yet competitive peer review is hardly ideal for measuring research quality on its own. Some areas require us to pursue a form of 'pendulum management', moving back and forth across a range of measures over time, and then perhaps repeating them. We might start with measures of customer satisfaction, then move to cost control measures and then move back again to customer satisfaction measures.

Thirdly, we need to be more critical. Garrison Keillor welcomes you to Lake Wobegon, "where all the women are strong, all the men are good-looking, and all the children are above average." One of my favourite, misleading measurement statements from purportedly competitive companies is "our key strength is that we have the best people." One would assume over time that there would be regression to the mean. What are the odds that a large global company with, say, 100,000 employees is able to prove any statistically different metric on the

quality of their people from 100,000 other randomly selected business people? In the very rare case where the quality of people might even be measured, it is a strange thing to brag about. One would assume that the best people cost more. In fact, I'd be more impressed if someone said "our key strength is that we take mediocre people at lower cost and through our superior systems deliver excellent results." During this 200th anniversary of Trafalgar, may I submit that that statement sounds like the mission statement for the very successful 18th century Royal Navy and its press gangs.

And yet, before I encourage you to be too critical of other people's measures I leave you with a cautionary quote from the Bible that urges tolerance. A number of scholars believe that this quote from Matthew Chapter 7 Verse 2 is the source of Shakespeare's inspiration for the title of "Measure for Measure":

"For in the same way you judge others, you will be judged, and with the measure you use, it will be measured to you"

Well, it wouldn't be a Commerce lecture without a commercial. The next Commerce lecture will follow the theme of better choice and explore "Perceptions Rather Than Rules: The (Mis)Behaviour of Markets" here on Monday, 14 November 2005 at 18:00.

Thank you.

Notes

[1] Because Gresham was not a 16th century fan of the soundbite, Gresham's Law is usually inaccurately expressed as "bad money drives out good". In fact, history repeatedly shows that the best Gresham's Law soundbite would be "good money drives out bad". What Thomas Gresham was pointing out was that the more solid currency will, over time, be valued more and reduce the circulation of the cheaper currency. Do see Further Reading, especially Professor Robert Mundell, the Nobel laureate, for more background.

Further Discussion

1. Where could slightly different measures make the biggest differences to society?
2. Where does BET% not apply?
3. Can you find a contradiction of Goodhart's Law where a measure is used for control but this enhances both the measure and the control?

Further Reading

Alfred W Crosby, The Measure of Reality, Cambridge University Press (1997).

Charles W Hofer and Dan Schendel, Strategy Formulation: Analytical Concepts, West Publishing Company, 1978 (1986 ed).

Bryan Magee, Popper, Fontana Press, 1973 (1985 ed).

Robert M. Pirsig, Zen and the Art of Motorcycle Maintenance, Bantam Books, 1974.

Thanks

My thanks to Mark Duff for pointing out the difference between the Lloyd's and Berkshire Hathaway reactions to events and to William Joseph, Ian Harris, Mary O'Callaghan, Robert Pay and Andrew Smith for sparking some ideas. Particular thanks are due to my brother Kel.

