Systems are reflexive in the sense that beliefs about what will happen often influence what does happen.  

4. In the recent announcement of the discovery of the ‘God Particle’ – the Higgs Boson – the European Lab, CERN said that its level of certainty was 4.9 sigma, implying that physicists came tantalisingly close to the ‘Gold Standard (5σ)’ – or one in 2 million chance of the result being a random fluctuation or noise.  

5. We cannot even dream of such precision in statistical estimates of economic variables where errors can arise from multiple sources such as errors in reporting, errors in aggregation, errors in measurement of relationships among the variables and so on. In other words, the chance of being fooled by randomness is much higher in the assessment of economic events. That makes deciphering the right signals from economic data and separating the grain from the chaff, that is distinguishing the trends and cycles from spurious noise, much more complex. The lesson from all this is that in collecting, analysing and interpreting data for economic analysis, we should be mindful of its inherent and unavoidable imperfections.

Data Gaps and Central Banks  

6. The theme for this year’s conference is ‘Data Gaps and Central Banks’. This is a subject that is always valid and one that should always be on our check-list. Nevertheless, I believe the motivation for the theme derives also from the global financial crisis that we have been going through for the past five years. In the context of the crisis, there are three main critiques hurled at economists and policy makers: that they helped brew the crisis, they failed to see it coming and that they have no idea how to fix it.  

7. Let us put these questions in the context of data and statistical analysis. Were data gaps and shortcomings in statistical measurements behind the crisis? Were the analysis and interpretation of data so flawed and
egregious that signals were missed? Are data gaps and
data analysis hindering crisis management? The short
answer to all three questions is that while data gaps
may not have caused the crisis, the crisis did reveal
several data gaps. Equally, shortcomings in the
measurement, analysis and interpretation of data
inhibited appropriate and timely response. As Claudio
Borio of the BIS said, ‘The main reason why crises occur
is not lack of statistics, but failure to interpret them
correctly and to take remedial action.’

**Have We Learnt from Past Crises?**

8. Since the global financial crisis is so much a part
of our consciousness, a relevant question is, have
statisticians learnt from past crises? As one of President
Obama’s advisers said, ‘This crisis is too good to be
wasted’. That is a dictum that economic statisticians
have always been mindful of.

9. The Great Depression of the 1930s, led to the
development of a new statistical framework for
economic analysis. As observed by the UN Statistical
Commission¹, national accounts were born out of the
Great Depression and were developed as a consistent
and comprehensive measure of economic activity for
policymakers. The lessons of the Great Depression also
led to the development of statistically estimated
macroeconomic time series models by the likes of
Tinbergen and Haavelmo, although the approach was
criticised by Keynes, who had reservations about the
model and the methods (Sims, Nobel Lecture, December
2011).

10. In the more recent period, following the Latin
American financial crisis, the IMF introduced the Data
Standards Initiatives to promote transparency of
economic and financial statistics comprising the Special
Data Dissemination Standard (SDDS) in 1996 and the
General Data Dissemination System (GDDS) in 1997.
These were aimed at enhancing the availability of
timely and comprehensive statistics that would
contribute to the pursuit of sound macroeconomic
policies. Earlier this year, the IMF put out SDDS plus
as an upper tier of its data standards initiative especially
targeted at addressing the data gaps identified during
the global crisis.

11. The global crisis that began in 2008, and which is
still with us, has triggered a collective initiative under
the G-20 umbrella to address data gap issues. The G-20
Data Gaps Initiative aims to bridge data gaps on the
build up of risk in the financial sector, cross-border
linkages, vulnerabilities of domestic economies to
shocks and cross-sectoral interconnectedness and to
improve the communication and dissemination of
official statistics. Several international bodies including
the United Nations Inter-Agency Group, IMF, FSB,
OECD, BIS and ECB are engaged in this exercise.

12. Data gaps have been broadly classified into two
categories: the first category is gaps for which a
conceptual and statistical framework and a system of
collection of data already exists. and the second is
where such frameworks need to be further developed.
The data gaps include strengthening financial stability
indicators, data on non-bank financial institutions,
information on exposure of cross-border
interconnectedness of financial institutions, sectoral
balance sheets, international investments and property
price data. These targets are best achieved through
creating a wide range of economic and financial
statistics that are mutually consistent thereby
eliminating contradictory signals due to measurement
issues.

13. The Reserve Bank has been actively engaged in
these initiatives pertaining to the financial sector and
has taken important steps towards addressing data
gaps, including contributing to India achieving SDDS
compliant status in 2001. For example, the Reserve
Bank compiles and reports financial soundness
indicators, international banking statistics, the
international investment position as well as data on
consolidated direct and portfolio investment. India’s
Annual Observance Report of the SDDS for 2011, posted
on the IMF website, shows that there are no major
deviations from the SDDS undertakings.

**Statistics at RBI: Some Introspection**

14. As a public policy institution, the Reserve Bank is
a large producer and consumer of statistics. Consistent

¹ UN Statistical Commission. February 2011: Draft Guidelines on Integrated
Economic Statistics
with our mandate, we collect and analyse data both for internal policy formulation and for external dissemination. Notwithstanding our best efforts, we must admit that we do not measure and analyse everything that we need to measure and analyse. Let me draw attention to a few important areas where the Reserve Bank needs to make a greater effort to fill gaps.

Data on Non-Bank Financial Companies

15. As the central bank, the Reserve Bank has responsibility for compiling monetary and liquidity aggregates. While our coverage of the banking sector is quite robust, our coverage of non-bank financial companies (NBFCs) is much less so. This is a gap we must fill as the role of NBFCs in the financial intermediation process has been growing rapidly. The number of NBFCs has increased from around 8,500 in 2000 to about 12,500 in 2011. As on March 2011, NBFCs had assets of `1.16 trillion and public deposits of `120 billion. NBFCs source their funds largely from borrowings, with nearly half of the borrowings from banks and other financial institutions. For the Reserve Bank to have a better understanding of the influence of NBFCs on liquidity and the credit creation process, we need more comprehensive and higher frequency data on NBFCs.

Data on Co-operative Banks

16. The co-operative banking sector plays an important role by providing financial intermediation services to agriculture and allied activities, small scale industries and self-employed workers. As at end March 2011, there were as many as 97,410 co-operative banks, of which more than 98 per cent were rural co-operatives. Yet our data on co-operative banks is weak – suffering both from coverage and reporting lags. Considering their systemic importance, geographical reach and role in financial inclusion, the Reserve Bank, in coordination with NABARD, needs to improve the statistical data reporting system of co-operative banks.

Data on Financial Inclusion

17. Financial inclusion has been among the Reserve Bank’s policy priorities. Yet, understanding how financial inclusion works at the grassroot level remains a challenge for us. There is this book, Portfolios of the Poor: How the World’s Poor Live on $2 a Day (Collins and others, 2009), a book written with great empathy that provides a sobering account on how the poor manage their everyday finances. The authors used an innovative method of survey data collection by having a panel of impoverished slum households in Bangladesh, India, and South Africa maintain financial diaries on an every day basis of their incomes and expenditures over a period of one year. They then analysed the transactions, one by one, penny by penny, to understand how poor households manage their money. The findings show that most poor households do not live hand to mouth. They fight poverty by employing surprisingly sophisticated financial management linked to informal networks and family ties. In the authors’ words, ‘They push money into savings for reserves, squeeze money out of creditors whenever possible, run sophisticated savings clubs, and use micro-financing wherever available.’

18. In order to pursue meaningful financial inclusion, the Reserve Bank needs many studies from across the country like ‘Portfolios of the Poor’. We need a better understanding of the levels and patterns of income of the poor, their patterns of expenditure, their marginal propensities to consume and save, the risks they face and how they cope with them and the transaction costs they encounter in dealing with formal financial institutions.

Basel III and Identifying Inflexions in Economic Cycles

19. A critical component of the Basel III package is a countercyclical capital buffer which mandates banks to build up higher levels of capital in good times that could be run down in times of economic contraction consistent with safety and soundness considerations. This is conceptually neat, but operationalising it poses many challenges, as indeed evidenced by Spain’s recent experience. The foremost challenge is identifying the inflexion point in an economic cycle which would trigger the release of the buffer. This needs to be based on objective and observable criteria. It also needs long series data on economic cycles. So, what we need is both a better database and more refined statistical skills in interpreting economic cycles.
Data on Corporate Saving

20. In the compilation of national accounts which is the domain of the CSO, the Reserve Bank is involved in the estimation of saving, compilation of flow of funds and external accounts. The standard practice for estimating saving is to estimate household saving in financial assets and corporate saving, and to determine households’ saving in physical assets as the residual. The residual method is a valid statistical approach (for that matter, GDP is also estimated as residual of total output after adjusting for intermediate consumption). Nevertheless, the method implies that any error in the measurement of corporate saving can create an error in the estimate of household physical saving, and thereby distort macroeconomic assessment. The Reserve Bank’s data set on the corporate sector is extensive and rich, but its accuracy has been challenged by analysts. We need to examine how we can improve this.

Data Consistency

21. The Reserve Bank has a wide mandate. The multiple policy objectives have resulted in multiple data systems. Each of these data systems is tailored for its specific objective, and has evolved over time to meet that specific objective. The problem, however, has been that each data system has evolved independently. We now need to focus on integrating the data systems to improve consistency across the aggregate database. This is perhaps most important in the data relating to the priority sector, segments of which have undergone definitional, conceptual and coverage changes over the years.

Constraints in Statistical Analysis at RBI

22. Having indicated areas where the Reserve Bank needs to focus on filling data gaps, let me now move on to some constraints we face in data measurement and interpretation.

(i) Measurement of Inflation

23. By far the most important statistic for the Reserve Bank is inflation which has traditionally suffered from measurement problems. I must admit that even at a personal level, I do not know how to interpret inflation. Twenty years ago, when I had a thick mop of hair, I used to pay `25 for a haircut. Ten years ago, after my hair started thinning, I was paying `50 for a haircut. And now, when I have virtually no hair left, I am paying `150 for a hair cut. I struggle to determine how much of that is inflation and how much is the premium I pay the barber for the privilege of cutting the Governor’s non-existent hair.

24. In India, we have several measures of inflation – one wholesale price index (WPI) and three legacy consumer price indices (CPIs). The different CPIs capture the heterogeneity of the economic structure and the large differences in the consumption basket across different population segments. In addition, the Government introduced, with effect from January 2011 (with base 2010=100), a new CPI series which has CPI (Rural), CPI (Urban) and CPI (Combined), representing the entire country and a weighting diagram based on the 2004-05 consumer expenditure survey.

25. One of the problems we have to contend with in assessing inflation trends is the divergence between WPI and the CPIs which is due to differences in coverage and weights. Food has a weight of only 24 per cent in the WPI as against weights in the range of 37-70 per cent in the CPIs. Metals and a few other bulk commodities, whose prices have been volatile in the recent period, have a weight of 10.7 per cent in WPI, but are not directly included in the CPIs. Services, whose prices have been on the rise, have weights in the range of 12-25 per cent in the CPIs but are not reflected in the WPI. The difference in weights and coverage and the divergence in price movements not only create a wedge between the different inflation measures but also sometimes move them in opposite directions.

26. Another problem is the measurement and interpretation of core inflation. Core inflation is usually estimated by excluding food and energy prices from the basket of goods and services that represents a household’s typical spending. The rationale for exclusion is that the prices of food and energy tend to fluctuate sharply and such volatility from the supply side, if passed on into the general price index, makes it difficult to interpret the overall trend. The surmise
is that core inflation, being less volatile, gives a better sense of future price trends.

27. Even in advanced economies where food and fuel constitute a relatively small proportion of the consumption basket, the focus on core inflation for policy formulation is being challenged. Lorenzo Bini Smaghi, a member of the executive board of the European Central Bank argues that core inflation is not a good predictor of headline inflation and adds that ‘it may have taken a while to realize this but, as the Romans used to say, to err is human, to persist in erring is diabolical.’ There are, of course, counter arguments: economists like Paul Krugman argue that the efficacy of core inflation in forecasting headline could be period specific. If one takes a longer series of over three years, there is some evidence that core inflation does have statistically significant predictive power.

28. Theoretically, the rationale underlying core inflation is valid if variations in the prices of food and energy products are temporary and do not, on average, differ from other prices. If that is not the case, core inflation is neither a good estimate of the underlying inflation nor a reliable predictor of future inflation.

29. Even as the use and interpretation of core inflation in advanced economies remains contentious, in an economy like India, we have to contend with additional questions. In our economy, where food constitutes nearly 50 per cent of the consumption basket and fuel has a weight of 15 per cent, can a measure of inflation that excludes them be called core? Inflation in fuel and certain protein food items has been persistent over the last three years. Can a persistent component be excluded from the core measure?

30. Also, structural changes in our economy over the past decade have created an unprecedented demand for commodities. In the absence of a supply response, this has resulted in a lasting change in the price level. Therefore, our headline measure of inflation will necessarily have a larger momentum than core inflation. In the Reserve Bank, we need to be mindful of these factors in interpreting the movements in headline and core inflation.

31. Against the backdrop of these conceptual and measurement problems, let me now respond to some criticism against the Reserve Bank with reference to inflation indicators.

32. One criticism has been that analysts are confused about what measure of inflation the Reserve Bank uses for its policy. I can only reiterate what we said several times in the past, which is that we study all measures of inflation, both headline and disaggregated, and in interpreting their levels as well as rates of change, we try to be mindful of their conceptual and measurement shortcomings.

33. Another criticism has been that the Reserve Bank should be guided by the CPI rather than the WPI. That we now have the new CPI data is a welcome development. Theoretically, CPI, which measures changes over time of the general level of prices of goods and services that households acquire for the purpose of consumption, is considered a better measure of inflation than WPI. But, the new comprehensive CPI does not have adequate history to support data analysis and to be used as a sole headline measure of inflation. At the same time, the Reserve Bank cannot ignore a price index which arguably reflects the most updated economic structure. So, in our assessment of the inflation situation, we use the new CPI as also the legacy CPIs, but not them alone.

34. This is not to argue that the WPI is flawless. In its present structure, the WPI does not capture the price movement of services. Also, it is a hybrid of consumer and producer price quotes. For example, the index captures the price of important commodities like milk from the retail markets; not at the producer level. In contrast to CPI, Producer Price Index (PPI) measures price changes from the perspective of the seller. Sellers’ and purchasers’ prices differ due to government subsidies, sales and excise taxes, and distribution costs. For these reasons, it is, therefore, desirable that we move towards developing a PPI that measures the average change over time in the sale prices of domestic goods and services.
(ii) Issues Relating to Expenditure Side GDP Estimates

For the purpose of monetary policy, it is important to assess aggregate demand conditions by studying the GDP data. Despite their well-known limitations, expenditure-side GDP data are being used as proxies for components of aggregate demand. However, in recent quarters, the difference between identified components of GDP data measured by the output method and the expenditure method has increased. For example, in the expenditure side data for Q4 of 2011-12, if we add up the year-on-year growth in private final consumption expenditure, government final consumption expenditure, gross capital formation and net exports, we end up with a 9 per cent year-on-year GDP growth. But the growth rate derived from the supply side data for Q4 of 2011-12, if we add up the year-on-year growth in private final consumption expenditure, government final consumption expenditure, gross capital formation and net exports, we end up with a 9 per cent year-on-year GDP growth. But the growth rate derived from the supply side, by adding indirect taxes and subtracting subsidies yields a number of 5.6 per cent. The difference between the two numbers as high as 3.4 per cent of GDP – is the error component. Such a large error component not only makes the data non-transparent but complicates the assessment of aggregate demand.

36. This is possibly an appropriate context for me to cite a few concerns about some official statistics. Economic measures such as income, debt, deficit and balance of payments are primary off-shoots of good book-keeping cum accounting systems. Many of the variables in these data have to be estimated using statistical methods and simplifying assumptions. This is in part because it is costly and time consuming to keep or process every record. This is also because some of these variables cannot be measured directly: think about the huge unorganised informal segment of the Indian economy.

37. How do we cope with this? A part of the data is supplemented by regular surveys. But surveys suffer from sampling and non-sampling errors. Also economic cycles, particularly economic slow-down, pose significant challenges to the accuracy of economic indicators, including those that are used as source data for the early GDP estimates. During this period, some firms go out of business, some may complete surveys only partially, and some may choose not to respond at all to voluntary surveys. Therefore, devising reliable and accurate methodologies for what goes into which account, even with today’s technology, is no trivial task, especially for a modern, complex economy.

(iii) Estimation of Potential Output

An accurate estimate of potential output is critically important for central banks to assess demand conditions and the output gap. Reliability and timeliness are critical. For example, an output gap measurement mistake led the Federal Reserve to pursue policies that eventually led to the Great Inflation of the 1970s.  

39. Potential GDP cannot be directly observed, but has to be estimated using statistical techniques. Such an estimate is vulnerable to errors for three main reasons. First and most obviously, real GDP data is often revised – sometimes substantially – after the initial estimates are published. Second, there is considerable uncertainty in the level of productivity growth at any point in time. Finally, there do not exist widely agreed assumptions or a unique methodology for estimating potential GDP. These reasons get accentuated in the case of India because of lack of comprehensive and consistent data on employment.

40. The uncertainty surrounding economic activity has heightened in the post-crisis period. India is no exception. Assessing India’s potential growth rate, consistent with our objective of low and stable inflation, remains a challenge. In its annual report for 2009-10, the Reserve Bank had reported that the potential output of the Indian economy may have dropped from 8.5 per cent pre-crisis to 8.0 per cent post-crisis. Latest assessment following the standard filtering technique suggests that potential output growth may have further fallen to around 7.5 per cent.

(iv) Market Reference Rates

41. The recent LIBOR controversy has drawn the world’s attention to how a few large global financial institutions allegedly manipulated one of the most commonly used market rates.

42. Can we have problem similar to LIBOR here in India? The counterpart of LIBOR in India is the less used MIBOR which is set by FIMMDA-NSE. The Reserve Bank, on its part, compiles and publishes every day the reference exchange rates for spot USD-INR and spot EUR-INR following a transparent statistical method. Earlier on, we used to set the rate by averaging the mean of the bid/offer rates polled from a few select banks as at 12 noon on every working day. A few months ago we changed the method. Now we poll a select list of contributing banks at a randomly chosen five minute window between 11.45 am and 12.15 pm and put out the reference rate at 12.30 pm. The contributing banks are selected on the basis of their market-share in the domestic foreign exchange market and representative character. The Reserve Bank periodically reviews the procedure for selecting the banks and the methodology of polling so as to ensure that the reference rate remains a true reflection of market activity.

Data Interpretation: Need to Look Behind the Data

43. Information Technology has improved data availability. Paradoxically, this has also increased the likelihood of misinterpretation of data. In his recent book, *Thinking, Fast and Slow*, Daniel Kahneman explains where we can and where we cannot trust our intuition and how we can tap into the benefits of slow thinking.

44. Consider an example: in a recent financial daily, an analysis based on last quarter operating results of Indian companies claims that interest costs were 11.1 per cent of total revenues, and infers on that basis that a cut in Reserve Bank’s rate can make a significant difference to the prospects of profits of firms. Arithmetically, this is correct and will make the reader believe that high interest costs have adversely impacted profitability. The reader would also intuitively attribute the deceleration in real GDP growth to 6.5 per cent in 2011-12 to the increase in the policy rate of the Reserve Bank by 375 bps during March 2010 – October 2011. But a closer look at the same sample suggests that the data used for this analysis includes financial and insurance firms which typically have interest to sales ratio of as high as 65 per cent. In contrast, non-financial firms, which account for more than 85 per cent of business, have interest cost of only 2.7 per cent of sales. Therefore, it is necessary to look behind the data and explore what lies underneath.

45. Consider another important variable – the real interest rate. Research in the Reserve Bank shows that real interest rate and investment activity are inversely related. A crucial question is, whether high real interest rate is the only factor behind the investment slowdown?

46. Recently, the Reserve Bank published detailed time series data of weighted average effective lending rates (WALR) of commercial banks, which have been compiled on the basis of contractual interest details of individual loan account data since 1992-93. The trend of WALR clearly indicates that the real lending interest rate today is lower than it was in the high growth period of 2003-08, a period when investment boomed. This is the case irrespective of whether we use WPI or the GDP deflator.

47. Some analysts have contested the Reserve Bank’s assessment by using the Benchmark Prime Lending Rate (BPLR) to derive the real lending rate. Use of the BPLR for this analysis is flawed. Flawed because the BPLR of banks failed to be a floor for lending rates: as much as 60-70 per cent of lending by commercial banks used to be done at rates below the BPLR. This was the reason why the Reserve Bank replaced the BPLR by the Base Rate. The BPLR has practically become defunct as a meaningful reference rate and as such real lending rates derived from BPLR would obviously tend to be artificially higher.

48. The Reserve Bank maintains that interest cost is only one of the several factors that have dampened growth, and the increase in policy rate by the Reserve Bank alone cannot explain the investment slowdown. I have asked our economic research department to do a detailed study on the time-series relationship between

---

* Business Standard, July 5 2012, (p.11).

real interest rate and investment activity. We expect to put out that report in the public domain in the next couple of months.

**Summing up**

49. Let me now conclude. I have spoken about statistical analysis in the context of the crisis, and about how every major economic and financial crisis has led to improvements in data standards and in the tool kits for statistical analysis. I referred to the international response to the data gaps identified in the ongoing global financial crisis. I then looked inwards at some of the more important data gaps within the Reserve Bank and highlighted why it is important to rectify them. Following that, I moved on to some constraints in data measurement and interpretation, and where necessary and possible, indicated how we address those constraints.

50. In conclusion, I want to reiterate that the Reserve Bank is deeply conscious of positioning itself as a knowledge institution. We place great store by the timeliness and accuracy of data and in the skill sets for analysing and interpreting that data. Oftentimes, the complex problems of the real world test the adequacy and the integrity of our data sets as also our ability to analyse and interpret them. We make best efforts to improve our data collection and analysis and to be mindful of the conceptual and measurement errors. We are not the best practice, but will spare no effort to reach the best practice, if not set the best practice itself.

51. Finally, a hearty welcome once again to our external experts, and my best wishes for the success of this Annual Statistics Conference.

**References**


Lorenzo Bini Smaghi (2011): Ignoring the Core can keep inflation at bay. Financial Times, June 1.


"Your statistics course was the one I hated most at University and was more than glad when it was over. But in my working career it has been the most valuable of any of the courses I took!" So, despite its challenges and controversies, taking time to get to grips with statistical concepts and techniques is well worth the effort. Its aim is to provide an explanation of the key techniques and formulas of statistical analysis, often using examples from widely available software packages. It stops well short, however, of attempting a systematic evaluation of competing software products. A substantial range of application examples is provided, but any specific selection inevitably illustrates only a small subset of the huge range of facilities available. Statistical analysis is fundamental to all experiments that use statistics as a research methodology. Most experiments in social sciences and many important experiments in natural science and engineering need statistical analysis. Statistical analysis is fundamental to all experiments that use statistics as a research methodology. Most experiments in social sciences and many important experiments in natural science and engineering need statistical analysis. This article is a part of the guide: Select from one of the other courses available.