# Bringing transparency to commodity markets in India: A real-world mission critical deployment of R

Susan Thomas
http://www.igidr.ac.in/~susant

Shobhana Vyas
http://www.cmie.com

June 14, 2006

## Goals

- Obtaining reference rates for the Indian commodity markets
- The polled price methodology
- Why R?
- The implementation at CMIE
- Looking forward

## Obtaining reference rates for the Indian commodity markets

- Goal: To provide reference rate from commodities spot markets, for settlement of futures contracts.
- Who: National Commodities Derivatives Exchange of India (NCDEX) – the futures exchange
  Centre for Monitoring Indian Economy (CMIE) - statistical system.
- Problem of non-transparency. Commodities are:
  1. Non-standardised.
  2. Priced using quotes from various dealers at many different locations.
- Dangers of price manipulation – non-random outliers.
- Need a framework to consolidate fragmented prices into a robust reference rate.

## The polled price methodology

- **The problem**: N reported prices for a commodity into a single reference price.
  The method has to be robust to outliers – which can arise from simple noise, or an attempt to manipulate prices.
- **What is reported**: Reference price ($\mu_P$) and the "standard deviation" of the quotes for the day ($\sigma_P$).
- **Approach used**: $\mu_P$ is calculated as the "Adaptive Trimmed Mean" (ATM) from the sample.
- **Approach used**: Both the extent of trimming, and $\sigma_P$ is calculated through the bootstrap.

## How much to trim?

- Trimmed means are a popular method of calculating a reference rate in finance (eg. LIBOR).
- How to choose the amount of trim? Let the data decide – the "Adaptive Trimmed Mean" (ATM).
- For each sample, apply trim, $K, = 1, 2, \ldots$.
  Compute $\sigma_{\mu_K}$ at each value using the bootstrap.
  "Best" $K$ is arg min $\sigma_{\mu_K}$.

## Why R?

- A previous hand-coded C implementation was actually faster
- But R is a sound foundation - e.g. bootstrap library developed by top researchers - more reliable than hand-coded C.

  Further, there is a huge repository of source available on the internet, and continuous development of this source.
- R is open source: the price is right and there is full flexibility in deployment.
- R is available on multiple platforms.
- Upside for further development of sophistication – all avenues for future progress will benefit from R language and libraries.

## The implementation at CMIE

- CMIE organises the calculation and dissemination of reference rates,
  - ⋆ thrice a day,
  - ⋆ for 34 commodities over 52 centers,
  - ⋆ making data collected from over 90 locations,
  - ⋆ spread over 3.3 million sq. km.

  $\approx$ a million phone calls a year.
- NCDEX receives the data at 10:45am, 1pm and 4pm.
- From reception at headoffice to when it is disseminated to the exchange, data handling has a window of 15-30 minutes.

## Looking forward

- What happens with more commodities, more locations, same half-hour window of time for analysis?
- "Dealer quality monitoring" - some dealers are more reliable than others.
- Draw upon the field of robust statistics to go beyond the adaptive trimmed mean (ATM). Collaborators welcome!