Applying Efficiency Measurement Techniques to Central Banks

Loretta J. Mester

Federal Reserve Bank of Philadelphia
and
Finance Department, the Wharton School, University of Pennsylvania

Prepared for the Workshop on Central Bank Efficiency
Sveriges Riksbank
Stockholm, Sweden
May 23-24, 2003

- **Outline of talk**
  - Overview of standard efficiency measurement techniques (to give a flavor of some of the difficulties)
    - Definition of efficiency – efficiency concepts
    - Measurement methods
    - Application of efficiency measurement techniques to central banks

- **Background**
  - Long literature on application of efficiency measurement techniques to financial institutions – Berger & Humphrey (1997) documented 130 studies of financial institution efficiency from 21 countries

  - Explaining differences in efficiency is difficult: Berger & Mester (1997) study of US banks in 1990-95 found 25 variables explained only 7 percent of variance of measured cost efficiency and about 35 percent of variance of measure profit efficiency

  - Small literature applying efficiency measurement to central banks

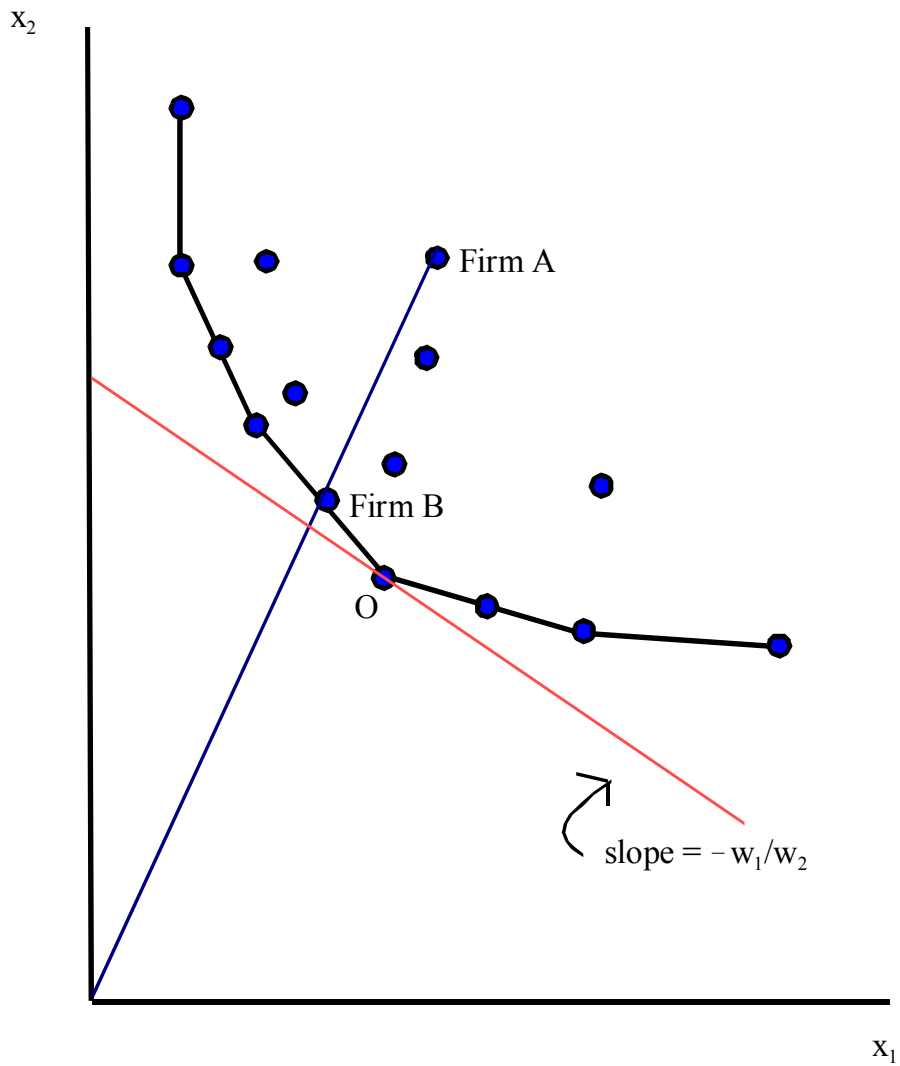
- **Definition of efficiency and efficiency concepts**
  - Efficiency is measured with respect to an objective, e.g., maximization of output, maximization of profits, or minimization of costs
  - Scale economies, scope economies, and X-efficiency are different aspects of performance
    - Scale and scope economies refer to selecting the appropriate outputs, while X-efficiency refers to selecting the appropriate inputs
    - <u>Scale economies</u> refer to how the firm's scale of operations is related to cost
    - <u>Scope economies</u> refer to how the firm's choice of multiple product lines vs. specialization is related to cost
    - <u>X-efficiency</u> refers to how productively the firm is using its inputs
      - Technical inefficiency – wasting inputs
      - Allocative inefficiency – using the wrong combination of inputs to produce outputs

- Sources of X-inefficiency: bad management, lack of managerial control
  - Expense-preference behavior: managerial utility function $U=U(\pi,E)$, where π is profits and E is expenditure on labor (or other input).
    - ♦ Managers derive utility from having large staffs or other perquisites, as well as high profits
- In commercial banking in the U.S., many studies have found large X-inefficiencies, on the order of 20 percent or more of total banking industry costs, and about half of the industry's potential profits.
  - The estimates often vary substantially across studies according to the data source, as well as the efficiency concepts and measurement methods used in the studies.

- Efficiency concepts (= choice of objective function)
  - Output maximization, cost minimization, profit maximization, risk-return frontier
  - <u>Economic</u> optimization in reaction to market prices and competition, rather than based solely on the use of technology.
    - Important to account for allocative inefficiency in misresponding to relative prices in choosing inputs and outputs
    - Firm could be technically and allocatively efficient in producing its chosen level of output, but choosing the wrong level of output in order to maximize profits.

Figure 1

Firm A is inefficient (technical and allocative)
Firm B is inefficient (allocative)



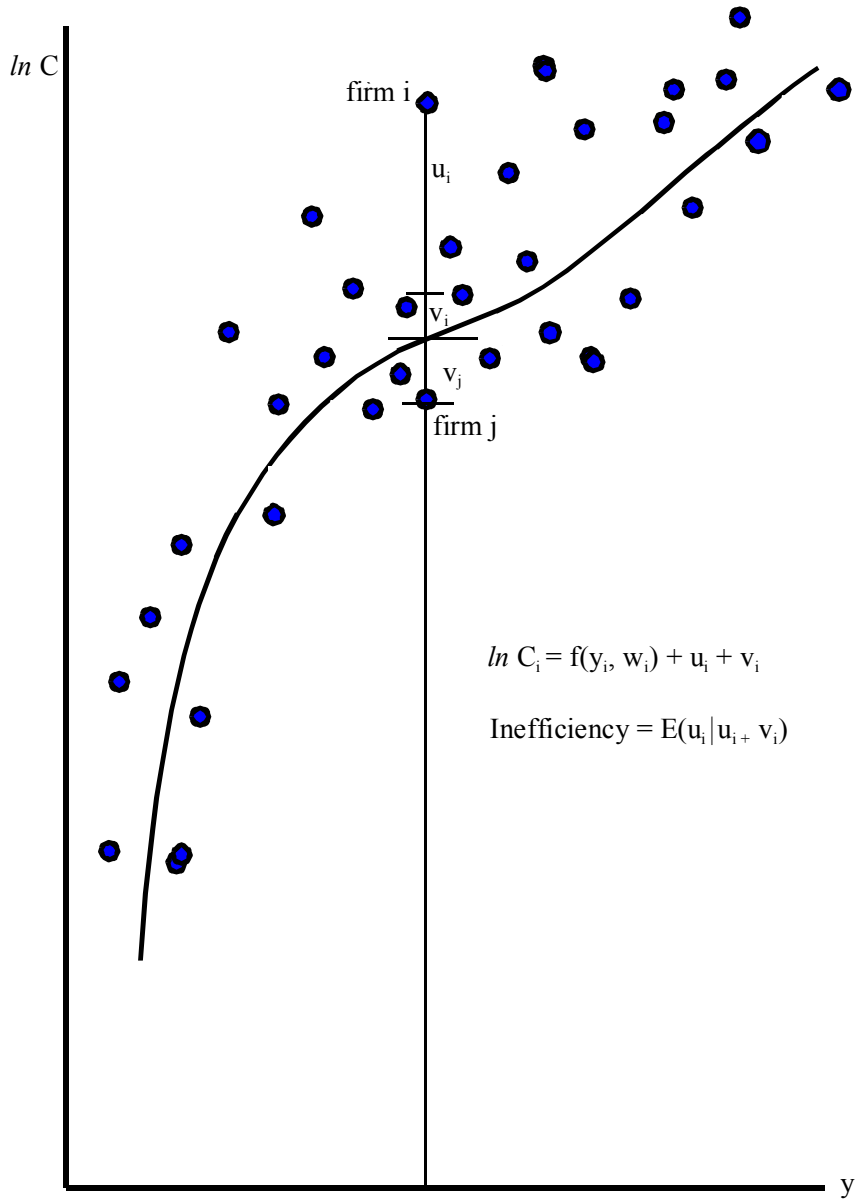
- **Cost Efficiency**

  - Measures how close a firm's cost is to what a best-practice firm's cost would be for producing the same output bundle under the same conditions.

$$ln\ C_i = ln\ f(y_i, w_i, z_i, h_i) + u_i + v_i, \qquad (1)$$

where C measures variable costs,
w is the vector of prices of variable inputs,
y is the vector of quantities of variable outputs,
z indicates the quantities of any fixed netputs
h is a set of environmental or market variables that may affect performance (e.g., regulatory restrictions),
u_i denotes an inefficiency factor that may raise costs above the best-practice level,
v_i denotes the random error that incorporates measurement error and luck that may temporarily give firms high or low costs.

Figure 2


ln C
firm i
u_i
v_i
v_j
firm j
$ln\ C_i = f(y_i, w_i) + u_i + v_i$
Inefficiency = $E(u_i | u_i + v_i)$
y

- **Standard Profit Efficiency**

$$ln\ (\pi+\theta)_i = ln\ g(p_i, w_i, z_i, h_i) - u_{\pi i} + v_{\pi i}, \qquad (2)$$

where π is the variable profits of the firm;
θ is a constant added to every firm's profit so that the natural log is taken of a positive number;
p is the vector of prices of the variable outputs;
$v_{\pi i}$, represents random error;
$u_{\pi i}$ represents inefficiency that reduces profits

- A more comprehensive measure of performance than is cost efficiency, since it accounts for errors on the output side as well as those on the input side.

- Embodies the cost inefficiency deviations from the optimal point, as well as revenue inefficiencies.

- **More Complicated Objectives**

  - Goals of cost minimization and profit maximization may not be general enough, e.g., may want to explicitly recognize tradeoff between return and risk, where risk is a choice variable of the firm

    - Important consideration for financial institutions
    - Prices of bank loans and deposits have an endogenous component since bank chooses its risk level =>When exposure to risk is influenced by production decisions, then cost minimization and profit maximization need not coincide to value maximization.
    - Estimates of efficiency that are derived from cost and profit functions may be mismeasured, since they do not penalize suboptimal choices of risk and quality that then affect prices.

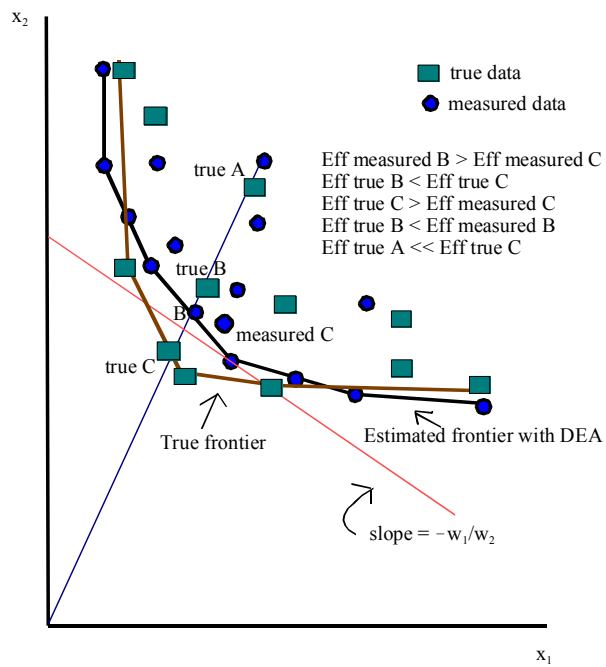
  - General form for efficiency measure:

    $X_i = \alpha_0 + \alpha_1 G_i + \alpha_2 (G_i)^2 + v_i - u_i,$

    where X is firm performance and G defines <u>peer group</u> (no penalty for inefficient choice of G)

- **Efficiency Measurement Methods**
  - Estimation Techniques
    - Nonparametric: DEA and FDH
    - Parametric: Stochastic frontier, distribution-free

  - Measurement error is a problem especially with nonparametric techniques



- Stochastic frontier
  - Distinguish u_i from v_i by making explicit assumptions about distribution
    - v, normally distributed
    - u_i half normally distributed, $u_i > 0$
    - Inefficiency measured by mean of the conditional distribution of u_i given $u_i + v_i$, i.e., $\hat{u}_i \equiv \hat{E}(u_i \mid (u_i + v_i))$

- Distribution-free approach
  - Available for panel data
  - Assume core efficiency for each firm over time; core inefficiency is distinguished from random error by assuming that core inefficiency is persistent over time, while random errors average out over time
  - Inefficiency measured by average (u_i).

- **Functional Form, Variable Selection, and Variable Measurement**
  - Translog and the Fourier-flexible functional form are used
  - Variables to include should be based on the model
    - X-efficiency is the residual => omitted variables (or extraneous variables) can have large effects on measured efficiency; peer groups are defined by the exogenous variables
    - Defining outputs, inputs and measuring them is not trivial: consider banking
      - Asset approach vs. production approach; asset quality; financial capital
    - Can include environmental variables – variables that affect performance but are not a choice of the firm
      - These affect the definition of the peer group
      - The alternative is to leave the variable out of the frontier specification, but then determine whether the efficiency estimates are correlated with the variable.

- **Tests of Expense-Preference Behavior**
  - Expense-preference is a specific type of X-inefficiency
  - Firm managers have utility function $U=U(\pi, E)$, where E represents expenditures on the input the manager likes
  - Tests are based on estimating input demand functions or cost functions
    - Functional forms are derived explicitly from the utility function, which depend on the underlying production function of the firm.
  - Tests cannot give firm-specific measures of inefficiency. They test whether a group of firms is showing expense-preference toward any input
  - Issues of functional form choice, variable choice, and variable measurement discussed above are as relevant in these tests as they are in measuring more general X-inefficiency
    - E.g., Cobb-Douglas production technology vs. more general form

- **Application of Efficiency Measurement Techniques to Central Banks**
  - To implement efficiency measurement you need to choose:
    - Efficiency concept, i.e., firm objective function (this includes specification of the production function of the firm)
    - Estimation technique
    - Functional form
    - Variables and their proxies
  - Can this process be applied to the central bank? Yes and no.
    - For narrow case of whether the central bank is creating its "output" in the most efficient manner in the sense of resource costs, then we can apply the standard efficiency measurement techniques to certain central bank outputs
      - Key is being able to define that output and to specify the central bank's objective function with respect to that output.
    - For other bank activities, this would be difficult because measuring the central bank's output is difficult.
    - A more difficult question is <u>social efficiency</u>, i.e., is the central bank operating to minimize the costs borne by society for a given level of "output"

- In the U.S., central bank activities include monetary policy, bank supervision and regulation, and payment services, to ensure financial system stability and maximum sustainable economic growth.

  - Efficiency measurement techniques can most easily be applied to the question of resource cost efficiency of providing payment services, given that the central bank is a provider of the services.

- **Application to Payment Services**
  - Fed's payment services include check clearing and collection, wire transfers, automated clearinghouse transfers, securities safekeeping, and coin and currency distribution.
  - Fed is mandated to provide some of these services on a competitive basis => assume Fed wants to minimize its costs of producing these services, subject to the constraints under which is must operate (e.g, given the areas it is required to service, particular inputs it is required to use, quality of output it is required to provide) and given the choice of production technology (one that does not subject the payment system to excessive risk).

- So can specify goal — minimize the cost of producing the payment service — and can measure the output of the payment service — e.g., number of checks processed, number of pieces of unfit currency destroyed. So can apply efficiency measurement techniques.
  - Implementation issues come up: whether the central bank wants to minimize costs each year, or over a longer period; data problems, etc.

- Applications
  - Several papers have estimated cost functions for central bank payment services and estimated scale economies based on the estimated cost function.
  - Fewer papers have applied efficiency measurement techniques to central bank payment services (see, e.g., Bauer and Ferrier, 1996; Bauer and Hancock, 1993; Bohn, Hancock, and Bauer, 2001; and Gilbert, Wheelock, and Wilson, 2002).

- Bohn, Hancock, and Bauer (2001) estimate cost functions and cost efficiency for Fed currency operations – example it can be done
  - Outputs are # of fit notes generated by high-speed currency processing operations; # of notes destroyed either on-line by the high-speed machines or off-line at the reconciliation stations; and # of transactions with depository institutions (= # of incoming shipments of currency received + # of outgoing orders for currency filled)
  - Inputs are buildings, labor, equipment, and materials
  - Estimate translog and a hybrid of the translog function cost with the stochastic frontier and distribution-free methods
  - 37 Federal Reserve Banks and Branches do currency operations.
  - Findings
    - Average office operates at more than 80 percent of the efficiency of the best-practice office
    - Cost-efficiency estimates for individual offices vary substantially across models but most and least efficient firms are consistent

- Issues in the application to payment services
  - Assumed choice of technology used in the production of payment services was given; once technology chosen, the central bank would want to provide services in the most efficient manner.

  - In choosing the technology, the central bank may have other goals in addition to efficient production.
    - E.g., to the extent central bank wants to ensure stability of the payment system, it may choose or design a technology or type of payment service that is not least-cost if that means lower risk with regard to stability.

  - Social efficiency of central bank in providing payment services vs. private sector provision
    - Standard efficiency techniques used to investigate resource costs in payments would not capture the social efficiency of technologies and products

- **Application to Banking Supervision and Financial Stability**
  - Don't know of any papers that have applied the standard methods of efficiency measurement to banking supervision
  - Difficult to apply efficiency techniques because difficult to specify the relevant objective function with respect to these activities
  - Goal of financial market stability is a worthy one, but how do we measure it?
    - Can conceptualize the question of whether the central bank efficiently deploys its bank examiners in a way to achieve the highest level of banking industry soundness.
    - Difficult to take question to data because difficult to quantify output
      - Can't just count # banks examined and relate it to cost of bank supervisory staff
      - Could relate the # of bank failures over a certain period with the # or cost of supervisory staff that examined banks in prior period and estimate a frontier across different regulators or over time (holding constant other factors related to failure)
        - Does a high failure rate mean poor supervision? Is a bank failure a failure of the supervisory process or a success?

- **Application to Monetary Policy**
  - Difficult because even a narrow application looking at resource efficiency still requires being able to write down the objective function of the monetary policymaker, measuring the "outputs" and "inputs" of policy
    - Resource allocation does not seem to be the interesting question regarding monetary policy efficiency.

  - Cosier and Longworth (2003) present a good overview of the Bank of Canada's approach to assessing the efficiency of monetary policy

  - Section 2A of the Federal Reserve Act => Fed is to conduct monetary policy to promote price stability, sustainable growth, and financial stability.
    - Translating this into a well-specified objective function can be applied is difficult, partly because the parameters of any loss function are not known and may differ across policymakers

  - Two applications

- **Expense-Preference Behavior of Central Banks**

  - Boyes, Mounts, and Sowell (1988) define the Fed's output as the monetary base (reserves + currency outside the banking system)
  - Assume the Fed has a Cobb-Douglas production function to create <u>Money supply</u>:

    $M_t = AL_{1t}^{b}L_{2t}^{c},K_{1t}^{d}\ K_{2t}^{e}$

    where M_t = monetary base
    L_{1t} = number of employees at the Board of Governors,
    L_{2t} = number of employees at the Federal Reserve Banks,
    K_{1t} = physical capital at the Board of Governors,
    K_{2t} = physical capital at the Reserve Banks.

  - Fed faces Money demand:

    $R_t = j_0 M_t^{j1} Y_{1t}^{j2} Y_{2t}^{j3}$

    where R_t is the appropriate interest rate,
    Y_{1t} is real GDP,
    Y_{2t} is the currency-to-deposit ratio

- If Fed managers act to <u>maximize profits,</u>

$$\pi = RM - W_1L_1 - W_2L_2 - r_1K_1 - r_2K_2,$$

then the implied demand functions for labor are:

$$ln\ L_{1t} = a_0 + a_1\ ln\ W_{1t} + a_2\ ln\ W_{2t} + a_3\ ln\ r_{1t} + a_4\ ln\ r_{2t} + a_5Y_{1t} + a_6Y_{2t}$$
$$ln\ L_{2t} = b_0 + b_1\ ln\ W_{1t} + b_2\ ln\ W_{2t} + b_3\ ln\ r_{1t} + b_4\ ln\ r_{2t} + b_5Y_{1t} + b_6Y_{2t},$$

If Fed exhibits expense preference for labor, maximizing

$$U=U(\pi_t,E_{1t},E_{2t})$$

$$ln\ L_{1t} = \alpha_0 + \alpha_1\ ln\ W_{1t} + \alpha_1\ ln(1-(U_{E1}/U_\pi))_t + \alpha_2\ ln\ W_{2t} + \alpha_2\ ln(1-(U_{E2}/U_\pi))_t + \alpha_3\ ln\ r_{1t} + \alpha_4\ ln\ r_{2t} + \alpha_5Y_{1t} + \alpha_6Y_{2t}$$

$$ln\ L_{2t} = \beta_0 + \beta_1\ ln\ W_{1t} + \beta_1\ ln(1-(U_{E1}/U_\pi))_t + \beta_2\ ln\ W_{2t} + \beta_2\ ln(1-(U_{E2}/U_\pi))_t + \beta_3\ ln\ r_{1t} + \beta_4\ ln\ r_{2t} + \beta_5Y_{1t} + \beta_6Y_{2t}$$

- Proxy expense preference term $ln(1-(U_{E1}/U_{\pi}))_t$ by ln [(Board assessments/number of Board employees)/System profits] $\equiv ln\ p_{1t} - ln\ L_{1t}$.

- Proxy the expense preference term $ln(1-(U_{E2}/U_{\pi}))_t$ by ln [(Federal Reserve Bank expenses/number of Reserve Bank employees)/System profits] $\equiv ln\ p_{2t} - ln\ L_{2t}$.

- Reduced form of the labor demand functions are:

$$ln\ L_{1t} = A_0 + A_1\ ln\ W_{1t} + A_1\ ln\ p_{1t} + A_2\ ln\ W_{2t} + A_2\ ln\ p_{2t} + A_3\ ln\ r_{1t} + A_4\ ln\ r_{2t} + A_5 Y_{1t} + A_6 Y_{2t}$$

$$ln\ L_{2t} = B_0 + B_1\ ln\ W_{1t} + B_1\ ln\ p_{1t} + B_2\ ln\ W_{2t} + B_2\ ln\ p_{2t} + B_3\ ln\ r_{1t} + B_4\ ln\ r_{2t} + B_5 Y_{1t} + B_6 Y_{2t},$$

- Joint test of Cobb-Douglas production technology and expense-preference behavior: estimate system without coefficient restrictions, then test whether coefficient on $ln\ W_{1t}$ = coefficient on $ln\ p_{1t}$, coefficient on $ln\ W_{2t}$ = coefficient on $ln\ p_{2t}$, coefficient on $ln\ W_{1t} < 0$ in the Board demand function, and coefficient on $ln\ W_{2t} < 0$ in the Reserve Bank demand function.

- Mester (1994) rejects the joint hypothesis of Cobb-Douglas technology and expense preference based on these tests.

- Mester (1994) generalizes the tests to allow for a more general production technology than Cobb-Douglas and to allow for the possibility that the Fed shows expense preference toward physical capital instead of or together with labor.

  - Reject expense preference on the part of the Fed.

- Issues in the application of expense-preference tests to central bank
  - These expense-preference tests show that standard efficiency techniques can be applied to monetary policy activities.

  - But tests are based upon a model where an efficient Fed maximizes profits.

  - Not clear that profit maximization is the correct metric for the central bank's money supply activities.

- **Other Objective Functions**

  - Long literature on monetary policy reaction functions, or Taylor-type rules for monetary policy:

$$f_t = r^* + \pi_t + \theta_\pi (\pi_t - \pi^*) + \theta_y y_t,$$

where f_t is the nominal interest rate,
π_t is inflation,
y_t is the output gap (the percentage deviation of output from potential output),
π^* is the policymaker's inflation target, and
r^* is the long-run equilibrium or "natural" real rate of interest

  - Rule can be derived from a model of the economy in which the central bank's goal is to stabilize output and inflation

- Model:

  - Inflation is determined by
    $\pi_t = \pi_{t-1} + \alpha y_t + e_t$, where $\alpha > 0$ and e_t is supply shock

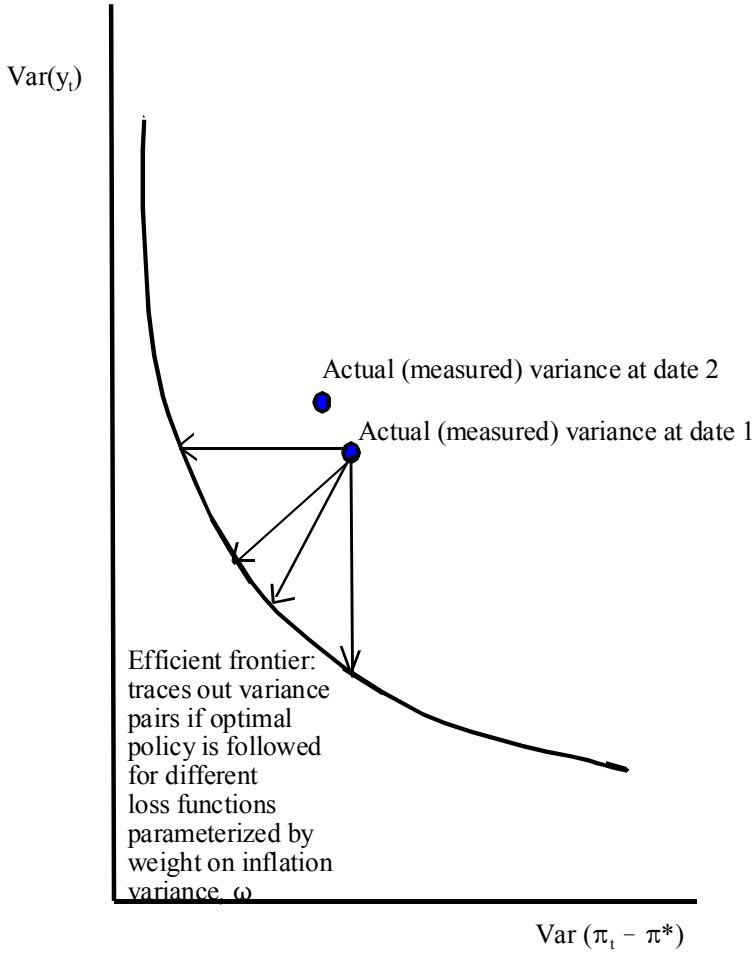
  - Output is determined by
    $y_t = \rho y_{t-1} - \xi(r_{t-1} - r^*) + u_t$, where $\xi > 0$, $0<\rho<1$, and u_t is demand shock

  - Central bank's objective function
    Minimize $L = \omega Var(\pi_t - \pi_t^*) + (1-\omega)Var(y_t)$, $0<\omega<1$

  - Then optimal policy is
    $f_t = r^* + \pi_t + \theta_\pi^N (\pi_t - \pi^*) + \theta_y^N y_t$,

    where $\theta_\pi^N = \dfrac{-\alpha\omega + \sqrt{4(1-\omega)\omega + (\alpha\omega)^2}}{2(1-\omega)\xi}$ and $\theta_y^N = \dfrac{\rho}{\xi}$.

- <u>Efficient frontier</u>: trace out inflation-output gap variance frontier as a function of ω, the weight on inflation in policymaker's loss function.



$$\text{Loss} = \omega \, \text{Var}(\pi_t - \pi^*) + (1-\omega) \, \text{Var}(y_t)$$

- Inefficiency: Compare actual performance to optimal performance
  - If we knew the weight, ω, in the policymaker's loss function, then compare the actual point to the point on the frontier corresponding to ω.

    Inefficiency = ω [Actual − Optimal inflation variability given ω]
    + $(1-\omega)$[Actual − Optimal output variability given ω]

    - Uses the policymaker's loss function to weight the deviations of actual output gap and inflation variance from optimal variances
  - But we will not know the policymaker's loss function parameter, so won't know what point on frontier is the proper reference point as optimal.
    - Horizontal distance ≡ increase in inflation variance for a given level of output gap variance;
    - Vertical distance ≡ increase in output variance for a given level of inflation variance
    - Minimum distance to the frontier
    - Distance to the frontier along a ray through the origin

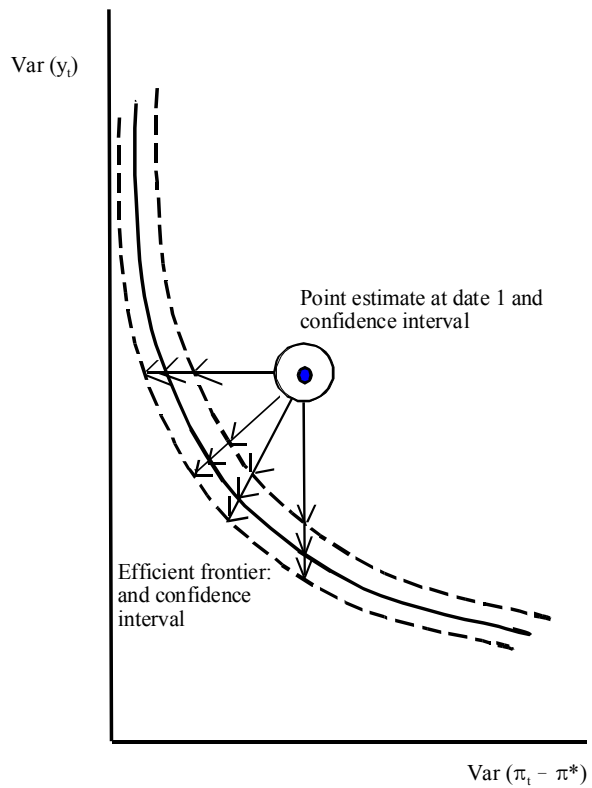
- Cecchetti, Flores-Lagunes, and Krause (2001), and Cecchetti and Krause (2002) studied central bank efficiency across different countries using this type of efficiency frontier.

- Estimated a two-equation model of the economy for 24 countries using data from 1991Q1 to 1998Q4, and trace out an efficient frontier for each country.
  - Assume that the inflation target is 2 percent for all countries.
  - Measure output by industrial production and use trend growth in industrial production as a measure of potential growth.
  - For a given ω, measure loss associated with actual performance of the country using the loss function in equation
  - Part of this loss is due to inefficiency, measured as above.
  - For each country's ω they use the ω that corresponds to the point at the intersection of the frontier and a ray through the origin and the point of actual performance.
- Find high variation in both performance and policy efficiency across countries and that central bank credibility explains most of the cross-country variation in macroeconomic outcomes (independence, transparency, accountability explain little)

- Issues in the application of efficient monetary policy frontier
  - Several embedded assumptions
    - Efficiency measures are based on a particular value of ω, but the policymaker's ω is not known.
    - Policymaker's inflation and potential growth targets are not necessarily explicitly stated by the central bank
    - Measurement issues regarding inflation and output gap
    - Need to believe model of the economy – that economy's dynamics imply there is a tradeoff between output and inflation variability.

  - Even if policy is efficient given the policymaker's ω, it could still be inefficient from society's viewpoint. (E.g., putting no weight on inflation variability would likely lead to very poor economic outcomes.)
    - Society's loss function may differ from policymaker's

- Even if ω were known, there are significant measurement problems with implementing these efficiency estimates.
  - Frontier shifts if the policymaker takes into account measurement error in the data, and this changes efficiency measures.
  - Explicitly recognize that at the time of evaluation, the data on inflation and output are measured with error.



$$\text{Loss} = \omega \text{ Var } (\pi_t - \pi^*) + (1-\omega) \text{ Var } (y_t)$$

- **Conclusions**
  - Application of standard efficiency measurement techniques to central banking is difficult
    - Uniqueness of some of the activities of central banking
    - Difficulty in measuring some of the central banking outputs
    - Complicated and multiple objectives pursued by central banks
  - Certain central bank activities do lend themselves to efficiency measurement, e.g., payment services provision
  - It is much more difficult to apply the techniques to bank supervision and monetary policy