INFORMATION SYSTEMS TO PROVIDE LEADING INDICATORS OF ENERGY SUFFICIENCY:

A Report to the Federal Energy Administration

MIT-Energy Laboratory Working Paper #MIT-EL-75-004WP

by

MIT Energy Laboratory

June 13, 1975

Final Working Paper

ENERGY INDICATORS

Submitted to

Office of Data Policy

Federal Energy Administration

In Connection with A Study of

Information Systems to Provide Leading Indicators of Energy Sufficiency (FEA Contract No. 14-01-001-2040)

by

M.I.T. Energy Laboratory

June 13, 1975

OUTLINE

APPENDICES

A. Possible Areas for Construction of "Snapshot" Indicators

B. Data Documentation

C. Use of Virtual Machines in Information Systems

D. Code for a General Purpose System Identified and Evaluator (GPSIE)

CONTRIBUTORS

The following M.I.T. faculty, staff, and students participated in the preparation of this study.

Daniel J. Deren

John J. Donovan

Marvin J. Essrig

John M. Fox

Louis M. Gutentag

Henry D. Jacoby

Ellen F. Nangle

David W. Peterson

Norman L. Rasmussen

Bruce E. Stangle

Jonathan B. Tepper

Sandra M. Viglione

1. INTRODUCTION AND SUMMARY

This is the sixth and final working paper of a project to study "Information Systems to Provide Leading Indicators of Energy Sufficiency." The purpose of this effort has been five-fold:

(1) To study and evaluate various types of energy indicators and their potential use in the area of energy sufficiency, and to develop a framework within which such a set of indicators might be developed.

(2) To study problems of data availability, data gathering, and data transfer that may be encountered in constructing such a set of indicators.

(3) To design the information management systems and modeling facilities that are needed to support a continuing program of development and maintenance of energy indicators.

(4) To demonstrate a sample implementation of selected indicators.

(5) To make recommendations for further development in this area by the Federal Energy Administration (FEA).

The period over which this activity has been carried out has seen significant changes in the focus and organization of those federal agencies concerned with energy, and in the stated goals of national energy policy. These changing circumstances have been reflected in the relative emphasis given to these five objectives and the specific direction of the work effort.

This final working paper presents several aspects of the work carried out over the project period:

Section 2 discusses the conceptual framework for energy indicators and distinguishes between different types of indicators.

Section 3 presents a sample implementation of a particular type of indicator (i.e., static or "snapshot" indicators) that was given high priority in the latter half of the project.

Section 4 gives a brief discussion of more complex forms of indicators (i.e., those of a dynamic-stochastic type), which were given a low priority in this particular project, particularly in its latter stages.

Section 5 summarizes the work on management information systems for the design and implementation of energy indicators of all types and addresses broader issues of the types of analytical and data management systems the FEA will need.

Other topics covered in earlier working papers are not necessarily repeated here, and therefore it is useful to summarize the history of the project and the material covered in previous submissions.

1.1 Project History

The Energy Indicator project began on June 15, 1974, and though the bulk of the activity has been carried out in Cambridge, the project has been marked by frequent contact between project members and FEA personnel.[1] Five working papers have been submitted over the project period, each dealing with a different aspect of the indicator issue, and a workshop presented to FEA personnel.

---

[1] During the summer of 1974, two project members worked in the FEA Office of Energy Data for a total of 104 man-days. During this period of residency in Washington, D.C., the time of these individuals was split between tasks associated directly with this project and other tasks assigned by officials of the Office of Energy Data Policy. Over the course of the study, an additional 23 man-days have been devoted to meetings in the Office of Energy Data and elsewhere in FEA, for a total of 227 man-days or 6 man-months in Washington itself. Project communication has been furthered by three visits to M.I.T. by officials of the Office of Energy Data Policy.

Working Paper No. 1: An Analysis of Basic Data Series [4]. The petroleum data collected by FEA are described and then organized in terms of a descriptive economic framework in order to investigate the ways various data series may be used to produce information about alternative stages of the fuel cycle and related economic processes. In addition, available data series (as of August 1974) are cataloged in terms of reporter frequency.

This paper was part of a longer-term effort within the FEA to catalog and document the data series available internally. Clearly, where data collection policies and procedures are changing rapidly, this is a difficult task to maintain. But it is an important one nonetheless, particularly, for any effort to take advantage of data series developed by FEA for purposes of indicator construction.

Working Paper No. 2: Problems of Data Transfer and Management [5]. Addressing the current FEA information systems (as of August 1974) this paper analyzes the key issues involved in the construction of an indicator information system and formulates a set of thirteen questions about the characteristics of each of the information systems in use within the FEA. More than 30 separate systems were identified at the time this paper was written. The answers to these questions would provide a means for discussing FEA information systems in a consistent fashion and thereby facilitate analysis of the integration of systems. These data on existing information systems also would have facilitated the transfer of data to a data base specifically designed for preparation of energy indicators.

Unfortunately, much of the specific data on these internal systems was not readily available to the M.I.T. team, and therefore this analysis could not be completed. It was decided that the FEA would carry on this task as an internal matter, and it was agreed that M.I.T. would not attempt to complete the work laid out in this first working paper.

To date, such a description of FEA internal data systems has not been completed, perhaps due to the fact that the systems themselves are continuing to evolve. At any rate, such an evaluation and common documentation would be very useful to any group continuing with work on energy indicators, and below we include the completion of this task among our recommendations.

Working Paper No. 3: Conceptual Framework for Energy Indicators [6]. Basic areas in which a set of energy indicators would be useful are outlined: monitoring the energy sector, assessing vulnerability, and evaluating policies and programs. A hierarchy of indicators is introduced which stresses that indicators develop logically out of data and models. The hierarchical relations are illustrated in terms of increasing complexity, with examples given for "snapshot," dynamic and stochastic indicators. Section 2 below summarizes many of the conceptual points made in this working paper.

Working Paper No. 4: Preliminary Results on Selected Sufficiency Indicators [7]. Building upon concepts used in Working Paper No. 3, this report explores issues of indicator design and interpretation. A model for interrelating the basic components of energy sufficiency is introduced, and the prospect for developing indicators with a clear "leading" character is addressed and shown to be problematic. Then, starting at the simple end of the hierarchy, a series of sample formulations are presented, and preliminary results submitted for review and comment.

Working Paper No. 5: "Snapshot" Indicators of Energy Sufficiency [8]. During meetings in October and November 1974, the Office of Energy Data Policy indicated that strong emphasis was to be put on simple, intuitive, static of "snapshot" indicators, and that work on the more complex dynamic and stochastic indicators[2], which require more complex analytical models, was to be de-emphasized. Working Paper No. 5 presents a first collection of indicators of this type and discusses others that might be constructed. Since the emphasis on simple, static indicators has continued through the end of the project, the bulk of the indicators shown in Section 3 are extensions and elaborations of ideas suggested in Working Paper No. 5.

In addition, Working Paper No. 5 also contains a discussion of issues in the presentation of data and some suggestions of possible improvements in the manner in which certain data series are displayed in the FEA's Monthly Energy Review. None of that earlier discussion is repeated in this final working paper.

Workshop on Data Management and Modeling Systems. In January 1975 the M.I.T. group gave a presentation at FEA in Washington of the work on the design of an information management and modeling facility to support a continuing program of development and maintenance of energy indicators. The implications of this discussion are broader than this specific task of indicators and have a relevance to a wide range of activities in the FEA.

The mechanisms needed to publish reports and perform data analyses for leading indicators are essentially the same as those needed to maintain and manipulate data for a wide range of policy analyses and studies within FEA. The requirements to meet FEA needs for data base and modeling capability are particularly trying due to the inherent problems associated with energy, energy data, and the broad and diversified implications of energy. For example: System uses are not well-defined; the uses and requirements change;

---

[2] The distinction between "snapshot," dynamic and stochastic indicators is elaborated in Section 2.

different data series become available; other data series become unavailable; protection requirements are complex; different groups of users within FEA have preferences regarding the modeling facilities they wish to use (e.g., TROLL, TSP, etc.); each facility runs under a different operating system. None of the modeling facilities have good data base capabilities; there are models that were and are being developed outside FEA that are running under different operating systems that FEA personnel would like to use on their system.

At this workshop the M.I.T. group presented a scheme and demonstrated that it would allow many modeling facility to run on a single 370 computer. The scheme advocated the use of VM/370 (Appendix D discusses this scheme). The M.I.T. group also demonstrated an advanced data base system, TRANSACT.

1.2 Summary of Contributions

The accomplishments of this project include the following items:

- Problems of developing indicators have been explored;
- A framework of indicators has been developed;
- Some potentially good "snapshot" indicators have been developed;
- FEA computational needs have been analyzed;
- The use of VM technology for FEA needs has been explored;
- Advanced data base systems and concepts have been elaborated for possible FEA use;
- Sample indicators have been constructed using advanced information management systems.

In this section we summarize two of the above accomplishments, sample "snapshot" indicators, exposure of VM technology for FEA needs.

1.2.1 Sample "Snapshot" Indicators

In Section 3 below a prototype set of indicators of the "snapshot" type is presented. The sample of over thirty indicators presented there is not meant to be a complete display of all the indicators that could be presented, or even of the most interesting ones discussed in the course of this project; indeed, as shown in Appendix B, there are data series in the data bank prepared for this project that would allow construction of many more indicators of various aspects of recent and current developments in the energy sector.

Of the set of indicators developed in Section 3, several seem particularly good at conveying an impression of the some aspect of the energy economy and are strong candidates for further refinement and eventual publication by the FEA. First, there are two indicators of the condition of the domestic fuel sector which deserve special attention. Figure 1.1 displays the oil and gas reserve-production ratio for the period 1960 to 1974. It shows the decline in the stock of reserves that stand behind current consumption levels in the United States;

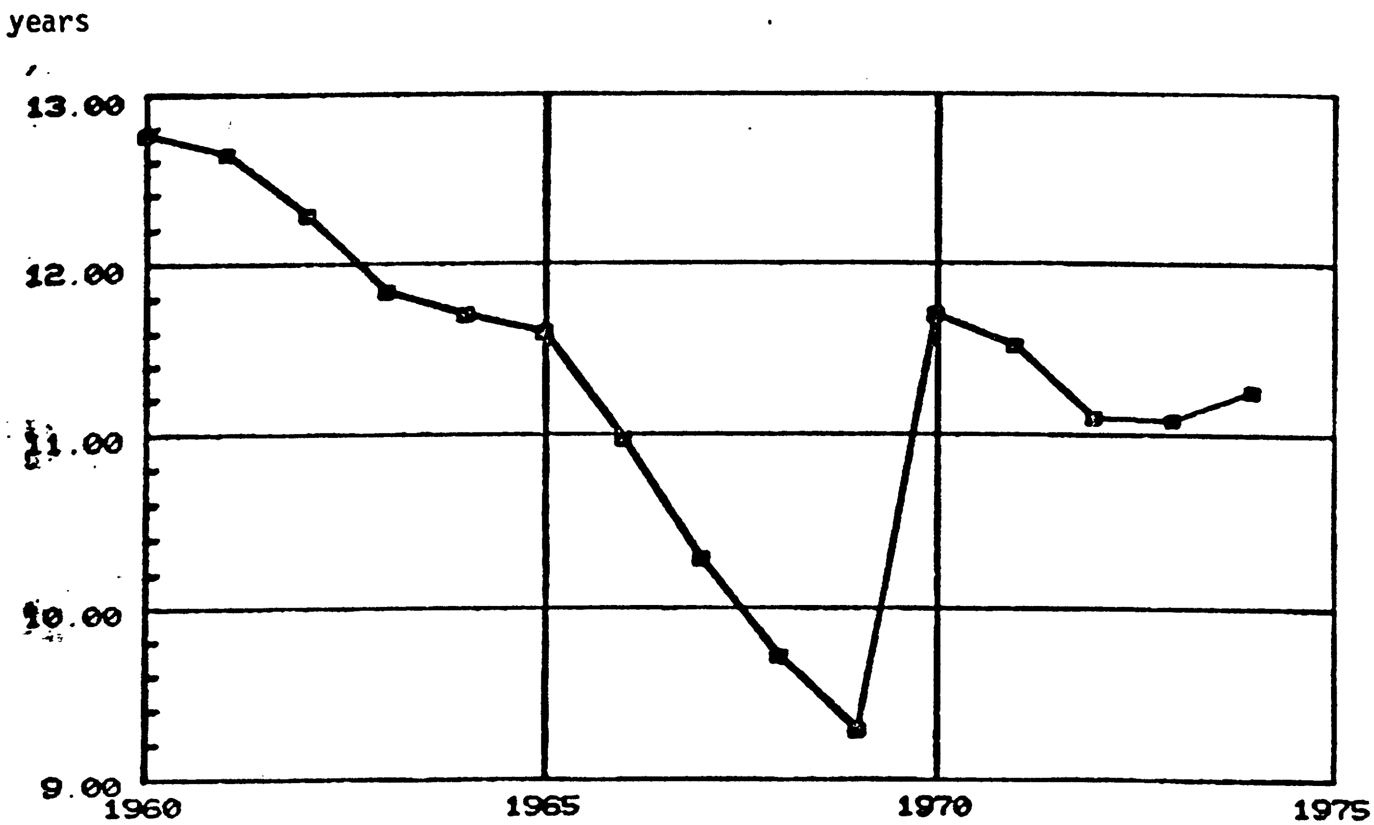


Figure 1.1

Crude Oil Reserve-Production Ratio

NAME: Crude Oil Reserve-Production Ratio

DEFINITION: The ratio of proved crude oil reserves at end of year to annual crude oil production.

INTERPRETATION: If one assumes that no new crude oil reserves are found, that crude production continues at a constant rate, and all other factors (prices, technology, etc.) remain unchanged, then this indicator shows how much time remains before proved reserves are fully depleted.

UNITS: Years

FREQUENCY: Annual

INPUTS: PET.RSVS--Proved reserves of crude oil (estimated as of December 31 of any given year) are the estimated quantities of all liquids statistically defined as crude oil, which geological and engineering data demonstrate with reasonable certainty to be recoverable in future years from known reservoirs under existing economic and operating conditions. Source: API, AGA, CPA, Bluebook.

PET. PROD--Crude oil production is the volume of liquids statistically defined as crude oil, which is produced from oil reservoirs during a year. The amount of such production is generally established by measurement of volumes delivered from leased storage tanks (i.e., the point of custody transfer) to pipelines, trucks, or other media for transport to refineries or terminals. Source: API, AGA, CPA, Bluebook.

FORMULA:

$$\frac{\text{Proved Crude Oil Reserves}}{\text{Crude Oil Production}} = \frac{\text{PET.RSVS}}{\text{PET.PROD}} \quad \frac{(\text{bbl})}{(\text{bbl/year})}$$

OUTPUT: Vertical axis on graph is in years and horizontal in time. Table of data is also given.

CONTACT: John Curtis, FEA, Office of Energy Data Policy, (202) 961-7986.

Figure 1.1a

Documentation for Indicator PET.RP

COLCOMB(PET.RSVS,PET.PROD,PET.RP)

	PET.RSVS	PET.PROD	PET.RP
1960	31613.	2473.	12.7833
1961	31758.	2508.	12.6627
1962	31389.	2552.	12.2998
1963	30969.	2613.	11.8519
1964	30990.	2645.	11.7164
1965	31352.	2699.	11.6162
1966	31452.	2864.	10.9818
1967	31376.	3047.	10.2973
1968	30707.	3161.	9.71433
1969	29631.	3189.	9.29163
1970	39001.	3328.	11.7191
1971	38062.	3299.	11.5374
1972	36339.	3274.	11.0993
1973	35299.	3185.	11.0829
1974	34250.	3043.	11.2553
	10^6 bbl	10^6 bbl/yr.	

Figure 1.1b

Numerical Data for Indicator PET.RP

from approximately a sixteen-year stock there has been a rapid decline in this index to the present, with the decline halted only temporarily by the large Alaskan discovery which enters in 1970. At present, the domestic petroleum system and natural gas system is operating at about an 11 to 1 reserve-production ratio, and given normal operating procedures and the availability of imports, one can expect production to continue at about this relationship in the future. So long as the ratio remains at this low level and imports continue to be significant, one cannot expect that self-sufficiency is near at hand.

Figures 1.1a and 1.1b show the types of documentation that is included with each of these indicators in Section 3.

This circumstance is shown in another way by Figure 1.2 which compares the proved petroleum reserves actually available in the United States year by year with reserves that would be needed to sustain a specified level of "self-sufficiency." In this particular plot, self-sufficiency is defined as meaning that a 10 to 1 reserve-production ratio is maintained in the petroleum sector while imports are held to no more than 15 per cent of total domestic consumption of petroleum products. In that event, the reserves required would be as shown by the jagged line in the figure. The actual reserves are represented by the smooth line (which is simply the annual data for reserves smoothed into a monthly representation). As the figure shows, the United States came through the latter part of the 1960's in a condition of approximate self-sufficiency. But beginning around the turn of the decade, reserves have fallen behind the level needed to sustain self-sufficiency in this fuel. The gap between the two lines indicates the level of effort that would be required--either in dampening demand or in increasing the reserve finding rate--in order to close this gap in the available crude reserves and sustain the defined level of petroleum consumption.

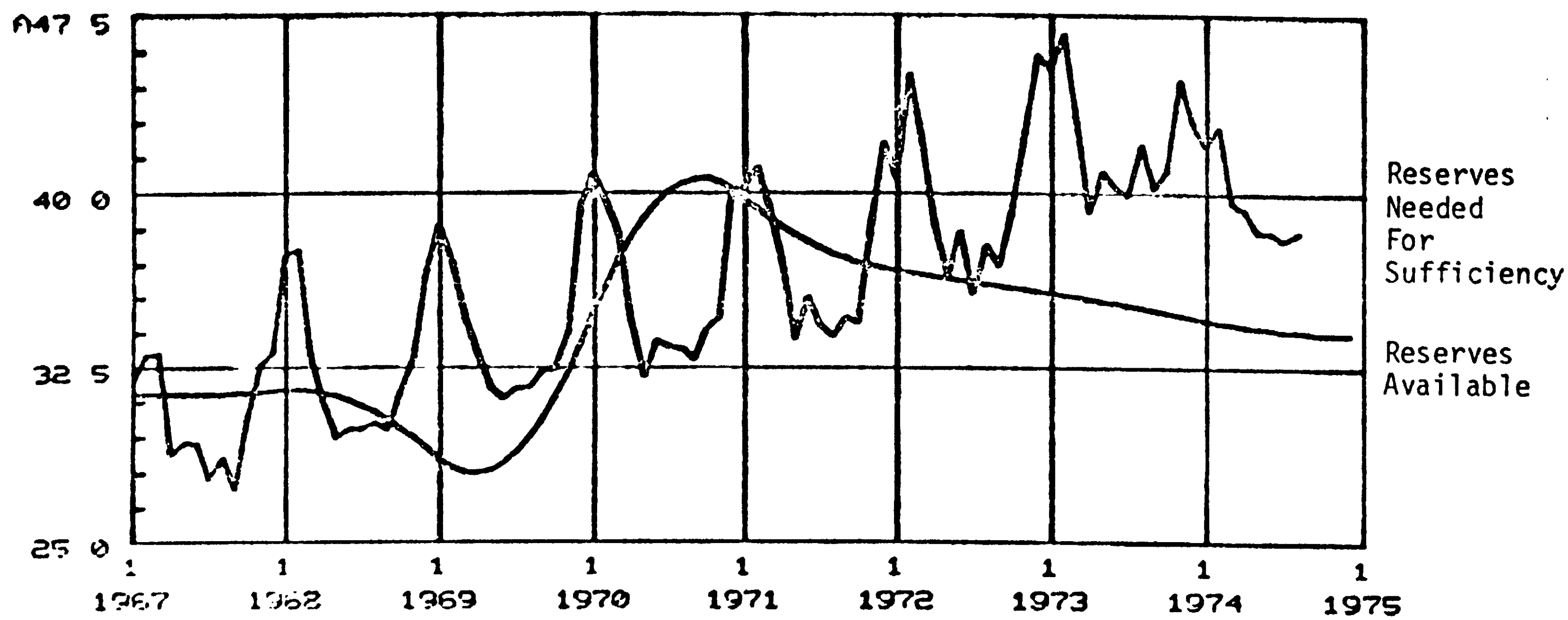


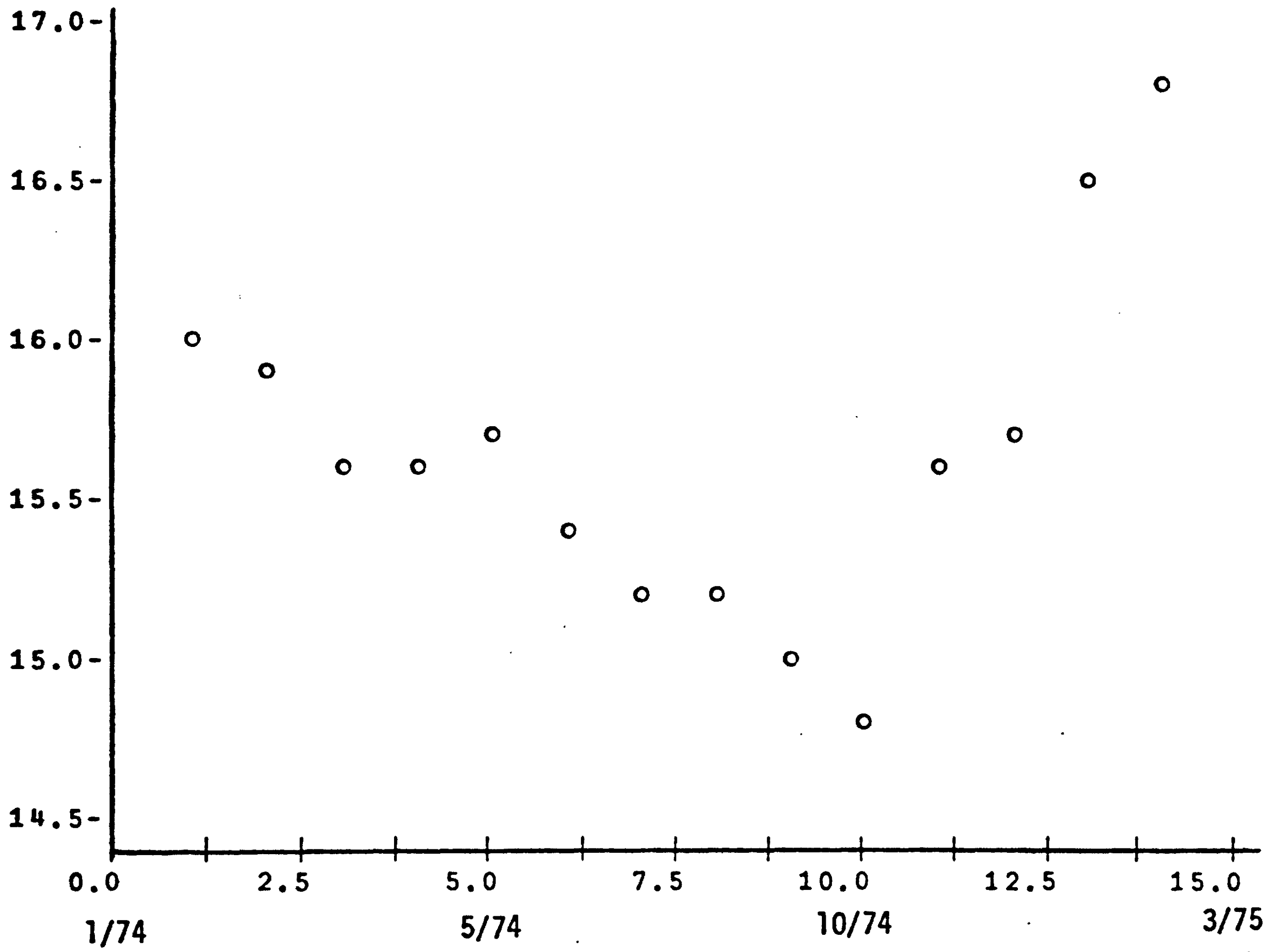
Figure 1.2

Petroleum Reserve Adequacy

Figure 1.3 presents a sample of the types of indicators that might be accumulated to indicate developments in long-run consumption patterns. It shows the weighted-average fuel consumption of new vehicles sold in the United States since January 1974. As can be seen in the figure, there has been a response to rising fuel prices and a continuing concern with energy problems in the United States. As a result of these phenomena, two things are taking place: drivers are becoming more conscious about fuel economy of vehicles they buy and are shifting their purchase patterns toward lower gas-consuming models; and in response to this shift in demand patterns, the aggregate fuel consumption of the mix of vehicles being placed on the market is declining. Figure 1.3 is an example of the use of an indicator to monitor the progress toward a specific policy goal, for one of our stated national objectives is to reduce the consumption of new cars by 40% over the next few years.

In indicating the short-run domestic supply condition of the country, and our likely exposure to disruption by an interruption in world oil supply, the conventional stock-flow ratio proves very informative. Figure 1.4 shows the ratio of the quantities of crude oil and petroleum products held in primary stocks, expressed as a ratio of the current level of imports of crude oil and petroleum products. The resulting index then represents the number of days of imports that are held in primary stocks. As the figure shows, in the late 1960's the United States held as much as 200 - 250 days of imports in primary stocks, but in the early 1970's, due to rising import levels, the amount of oil held in these stocks fell to between 100 and 150 days of imports. To the extent this index rises or falls, it indicates an increase or decrease in the degree of insurance that the country has against disruption of import flows.

26 60 6 *PLOT 'MPG'*



ABSCISSA = TIME STARTING FROM 1974 1

o = *MPG*

Figure 1.3

Average Gasoline Mileage of New Cars

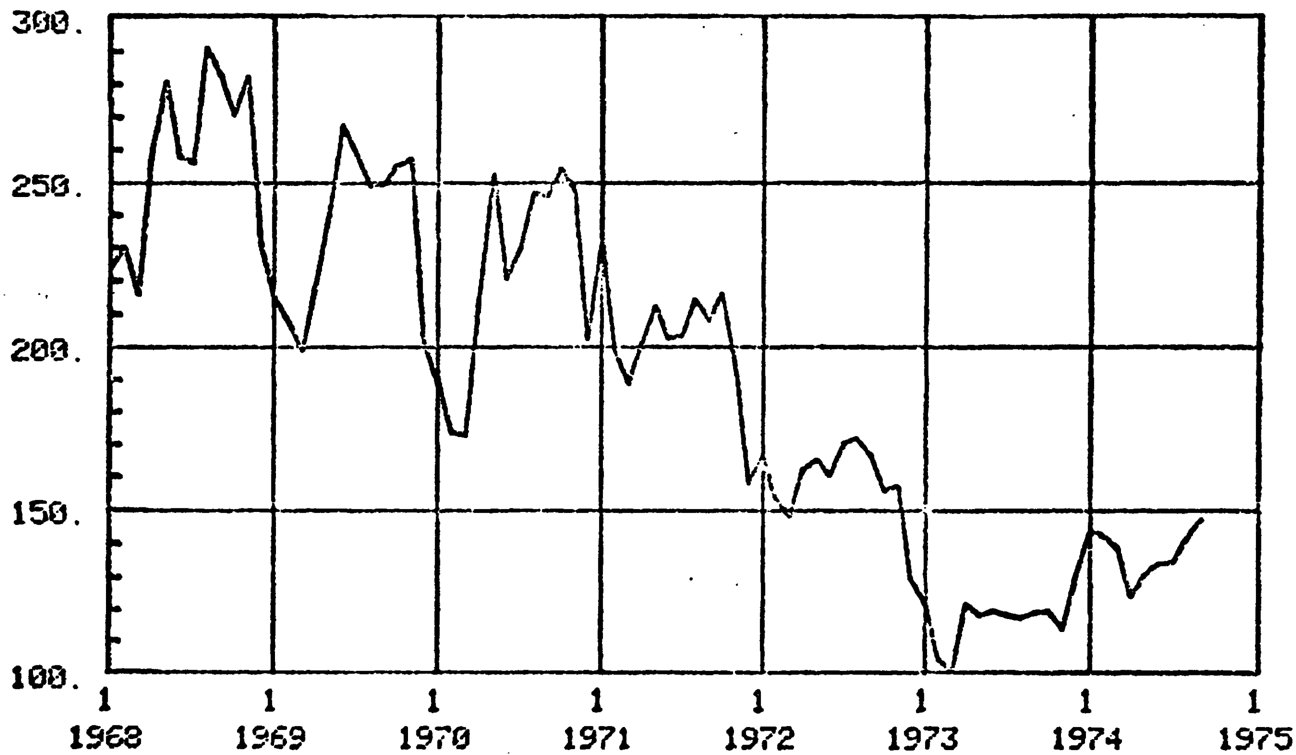


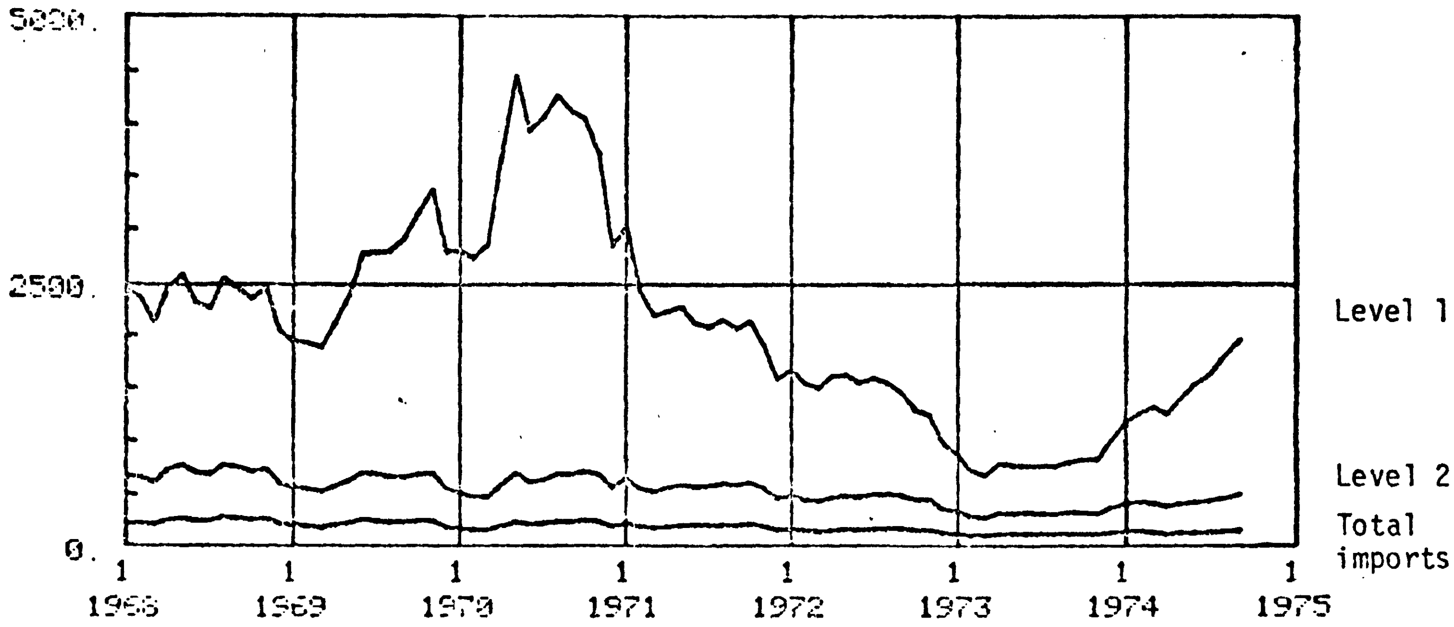
Figure 1-4

Days Supply of Petroleum Imports
(volume weighted)

Another way of showing this same phenomenon is shown in Figures 1.5 and 1.6. Figure 1.5 indicates the number of days of imports held in primary stocks as does Figure 1.4, but it also shows the number of days of insecure imports that these stocks represent. Here for illustrative purposes, insecurity is defined in two ways. First, stocks of crude oil and refined products are expressed in their ratio to crude oil imports from the Persian Gulf Arab nations. As the figure shows, U.S. primary stocks range between a thousand and three thousand days of Arab imports. Of course, many of the imports from the Persian Gulf directly to the United States are in the form of crude oil, but this is not the only oil coming to the United States from this point of origin; a good deal of the oil imported in the form of refined products from Carribean and European refiners is in fact oil from Arab Persian Gulf sources. The middle line in Figure 1.5 indicates the number of days of imports that are represented by the sum of Arab crude imports and product imports that may be from Arab sources.

Needless to say, the data in Figure 1.5 are very rough indicators of vulnerability; a much more detailed level of analysis and more explicit modeling procedure would be needed to take account of the types of changes that actually take place in an embargo, and to gain a more accurate representation of the true level of vulnerability that imports from insecure sources represent. But these numbers do give a rough indication of the level of security we now have, and the rising or falling of these indicators shows whether or not we are becoming better or worse off as time goes by in this regard.

Figure 1.6 shows the same thing only in a different way. Instead of reflecting imports in relation to stocks, it simply shows total import levels in relation to total consumption of refined products in the United States, and indicates the

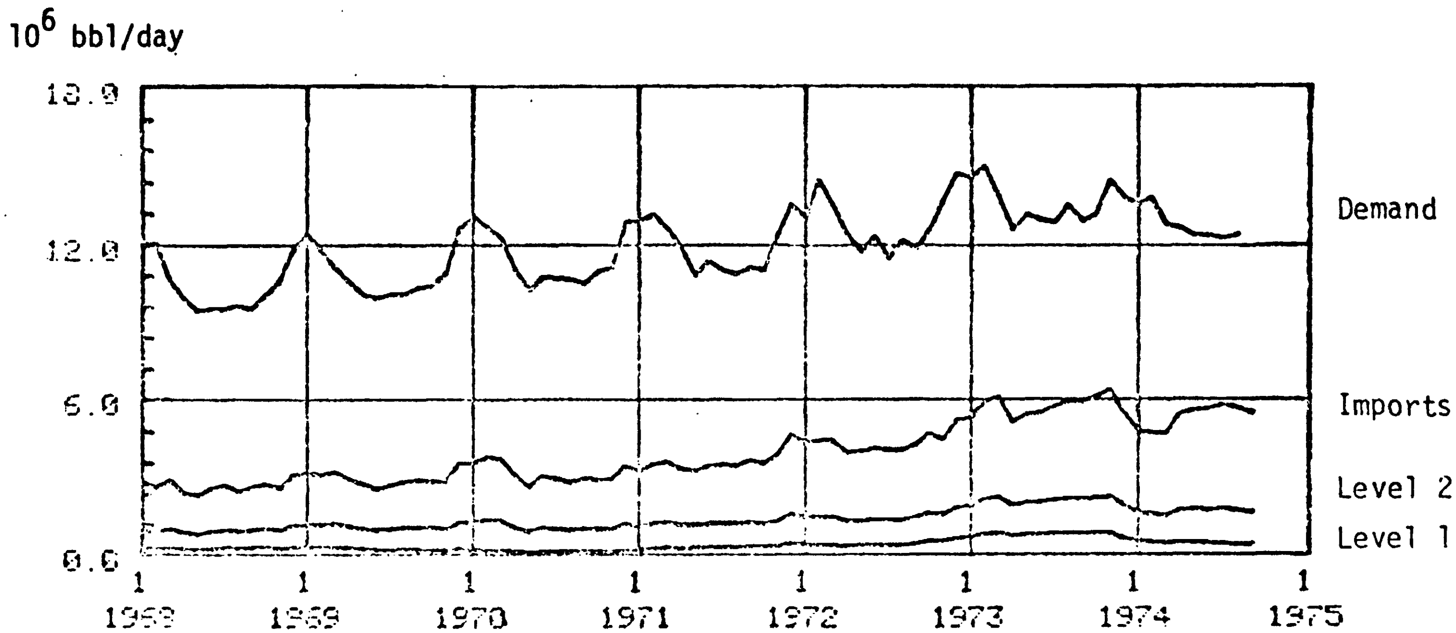


Level 1 indicates stocks in relation to direct Arab Persian Gulf imports.

Level 2 indicates stocks in relation to imports directly from Arab Persian Gulf sources, and from European and Caribbean refining centers which may originate from Arab Persian Gulf sources.

Figure 1.5

Days Supply of Petroleum Imports



Level 1 indicates imports directly from Arab Persian Gulf sources.

Level 2 indicates imports directly from Arab Persian Gulf sources, and from European and Caribbean refining centers which may originate from Arab Persian Gulf sources.

Figure 1.6

Import Dependence
Petroleum Demand, Total Imports, and Imports from Arab Persian Gulf Sources

fractions of total imports that are represented by sources of varying levels of insecurity. Once again in Figure 1.6 one can clearly see the rising dependence of the United States on imported petroleum and petroleum products, and, to a lesser extent, a rising role of imports from sources that might be considered insecure under the definition developed above for illustrative purposes.

Tables 1.7 and 1.8 illustrate another type of indicator that can be constructed by manipulation of data available from the price series prepared by the U.S. Department of Commerce. These are indicators of the relative price of certain energy products in the U.S. economy. Figure 1.7 shows the ratio of the electricity consumer price index to the index for all consumer prices; and as can be seen, electricity prices rise dramatically at around the first of 1974 and have remained at a higher level since. Figure 1.8, on the other hand, shows the experience with gasoline. Gasoline prices rose dramatically starting in the third and fourth quarters of 1973, but they peaked out in mid-1974 and have fallen significantly in recent months <u>relative</u> to the prices of all consumer products. Many other indicators of this type could be constructed, and various weighted indicators summing over several energy products should also prove informative.

Finally, Figure 1.9 shows a summary indicator of developments in the world oil market. The bottom line shows production from the OPEC countries. The top line shows an estimate of net OPEC capacity. Note that, in the early months of 1974, the excess capacity was in the range of 10 percent, which seems reasonable given the long planning horizons for petroleum capacity and the need to have some excess for purposes of adapting to short term fluctuations. Beginning in the second quarter of 1974, an ever-widening gap has opened between the productive capacity of the oil cartel and its ability to sell oil in the world market, and

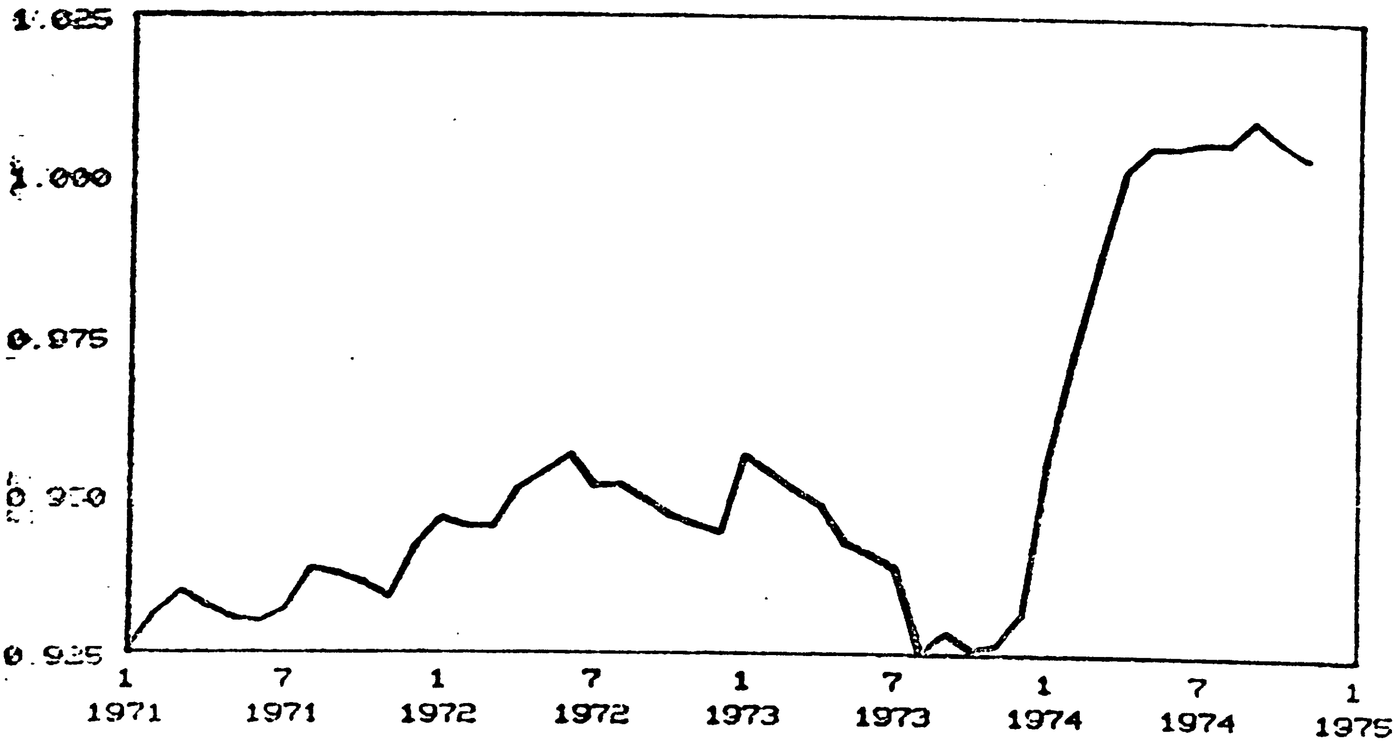


Figure 1.7

Relative Price of Electricity

(value in 1967 = 1.0)

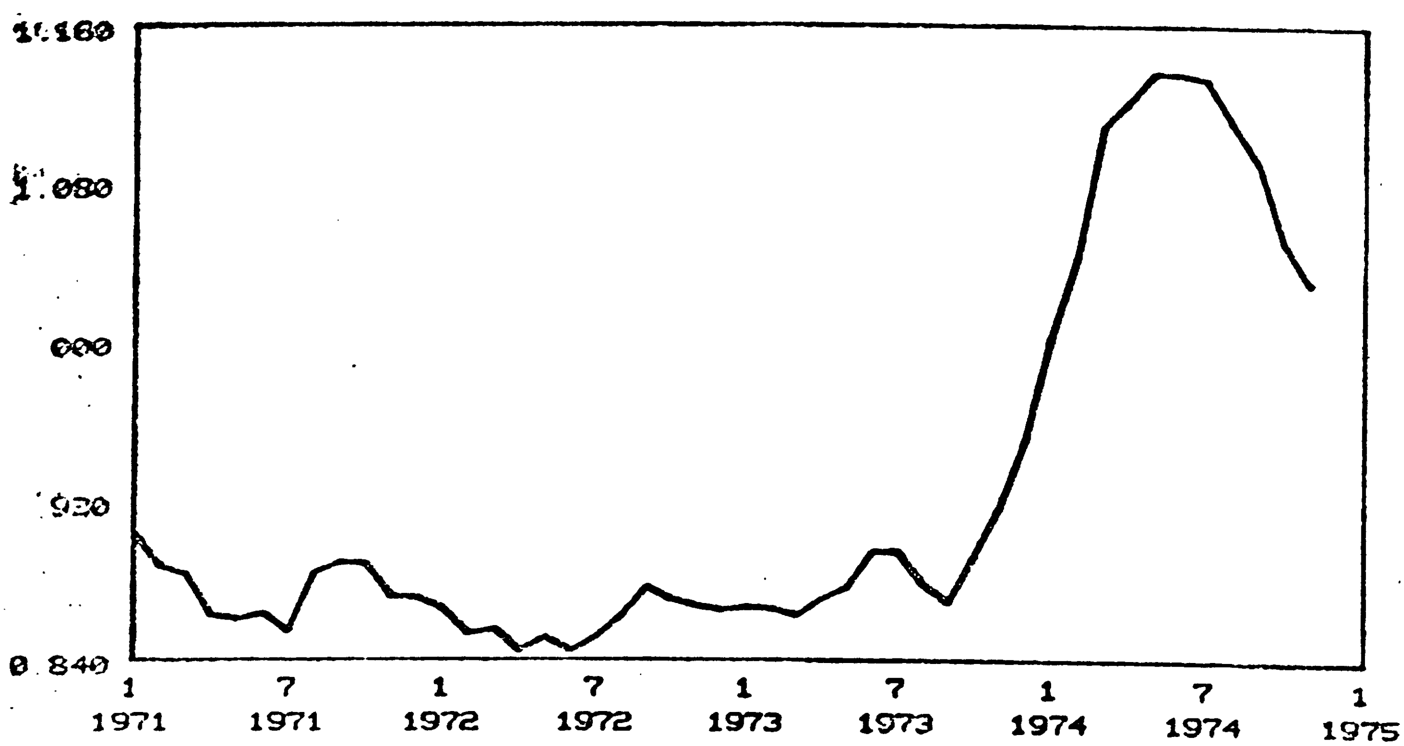


Figure 1.8

<u>Relative Price of Gasoline</u>

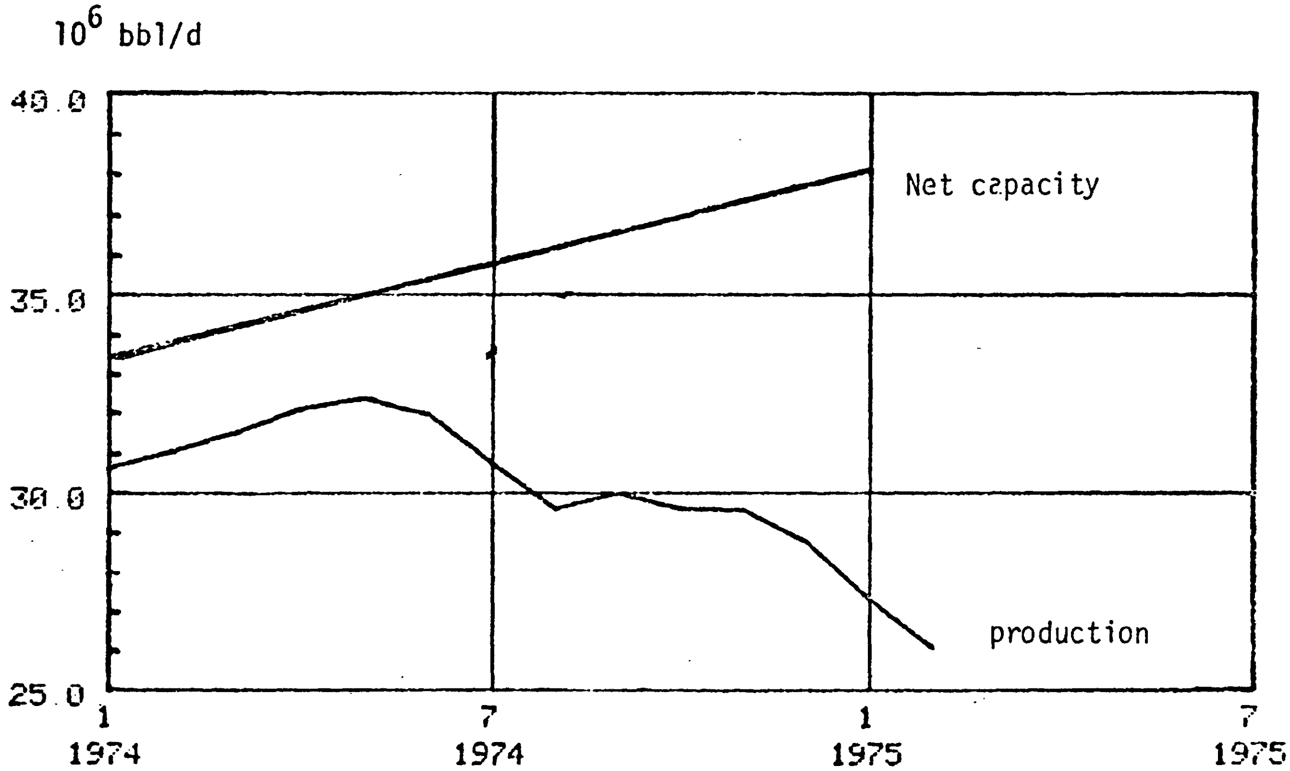


Figure 1.9

OPEC Excess Capacity

to some degree the magnitude of this gap is a rough indicator of the type of pressure this cartel is under.

These nine indicators are only a sample of the types of indicators that could be prepared. They are a first summary set which we would recommend that the FEA consider developing further and ultimately publishing in some form. Eventually, some subset of 10 to 15 summary indicators such as this should become a regular feature of the FEA's publication program.

Lying behind the indicators shown here is a larger group of indicators displayed in Section 3.

1.2.2 Data and Modeling Facility

Our work in information systems (Section 5) has involved three key aspects:

(1) Attempting to show the importance of operating systems presently available on the hardward the Agency is now purchasing, and showing the flexibility of these operating systems (namely, how VM can be used to allow many incompatible modeling and data base systems to work together).

(2) Indicating a direction that management information systems are likely to take in the near future, so that these developments can be taken into account in the planning of the Office of Data.

(3) Demonstrating a prototype system for producing energy indicators using these new concepts,with the possibility that the FEA may want to consider adopting this system even in its current experimental stage.

Exposure of VM Technology for FEA Uses. During the course of this project, it has become clear that good analytical and data base capabilities are needed if the FEA is to continue to make effective contributions to the analysis of complex energy problems. Without good tools (modeling facilities, data base facilities) the FEA analytical effort could be seriously hampered.

Examples of the problems to come can be seen in the current special needs within the Administration:

- Different groups both within and outside FEA support and have strong preferences for different modeling facilities. Each of these modeling facilities run under a different operating system.

- The needs of the FEA computational facility are not well-defined and are, in fact, changing. For example, the country has gone from concerns about petroleum allocation, to possible problems with coal, to economic impact problems, etc.
- The FEA needs quick and inexpensive ways of introducing new data series and performing analyses on them.
- Energy data has complex data validation requirements.
- FEA personnel have needs for facilities to build complex models.
- There is a need to inexpensively and quickly transfer existing energy models (that have been and are being developed outside FEA) onto the FEA's machine.

To accomplish these multiple tasks, we have concluded that an operating system VM/370 is the best scheme for the Agency. Figure 1.10 depicts a configuration of virtual machines as they might be utilized by FEA. Across the top of the figure we depict three or more virtual machines, each of which is running a modeling system under its own operating system, e.g., TSP, running under the operating system, MVT. All these modeling systems have very poor data handling capabilities. Thus we advocate separating the functions of data handling into a data base system. Among the currently commercially available systems, we do not advocate any one particular system. We do suggest that several of the modeling facilities be connected to one data management system, as this connection offers a way of having multiple users access the same data base.

Several different and perhaps incompatible data base systems could be running on their own virtual machines as in Figure 1.10. The architecture of VM could allow any modeling machine to access data stored in any data base system. For FEA to have such a facility operational would require VM/370

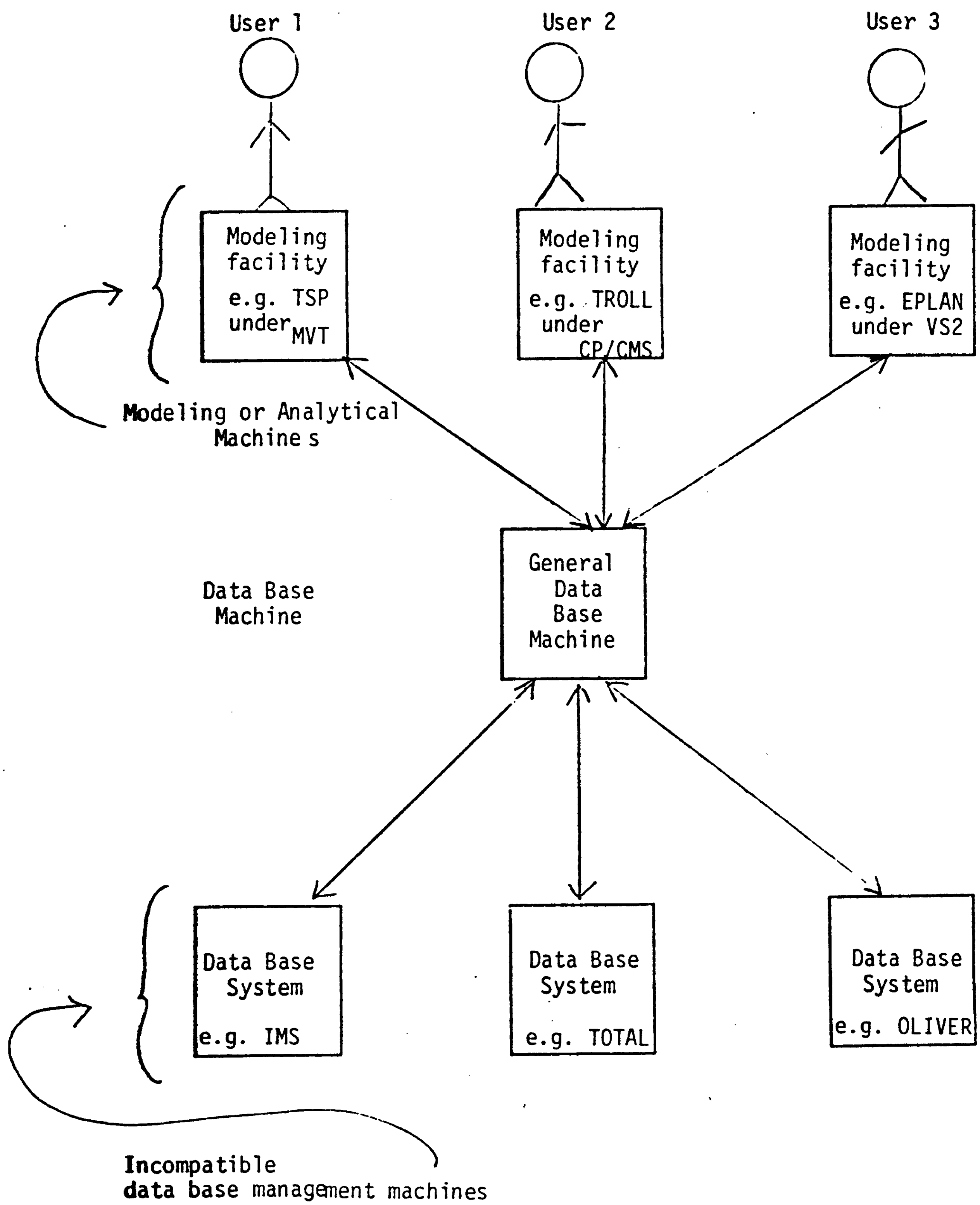


Figure 1.10

Configuration of Virtual Machines

and some very simple interface programs as outlined in Appendix C.

Not only would the VM solve the needs previously mentioned; it would also have the following cost advantages:

- no conversion cost in bringing up existing models as long as they run on an IBM machine (independent of language or operating system).
- no retraining cost involved as programmers may use whatever system they are familiar with.
- little cost involved in implementing the simple interfaces.

The possible disadvantage is performance, which is reflected in additional overhead costs. It is our intuitive feeling that the degradation costs are more than compensated for by the increased effectiveness of FEA staff using such a system. Below we recommend empirical study of this issue.

Exposure of Advanced Data Base Systems. We have developed and implemented a very flexible data base system. This system is representative of systems of the future, and FEA may want to keep abreast of this technology. The major points of the system that should be noted are:

- hierarchical implementation
- hierarchical user view
- security mechanisms
- relational base model of data
- flexibility and case of introducing new data series
- ease of implementing interfaces to modeling facilities
- mathematical soundness

1.3 Recommendations for Further Work

The FEA has on-going within it a rich series of studies involving data collection and processing and modeling activities of various kinds, and out of these various modeling projects there comes a wealth of data--indicators if you will--of the likely developments in the energy economy in the months of years ahead. These modeling results, in terms of our taxonomy of indicators, are classified as producers of dynamic and/or stochastic indicators. and should be considered part of any publishing program for information of this type. "Snapshot" indicators, such as the ones produced as part of this study, are an important part of any such program. Based on our experiences in constructing the indicators shown here, and in Section 3, we would make the following recommendations with regard to further development in this area:

(1) A selected subset of indicators such as the ones presented above (a total of no more than 10-15) should be developed and published in a special static indicator section of the Monthly Energy Review. These summary indicators would pull together information which is now scattered throughout the many series that are plotted and printed in that document as it stands, and would help to give an overall impression of the current status and trends in the energy sector.

(2) The inclusion in such a set of summary energy indicators of the results of the various energy forecasting activities underway within the FEA was outside the focus of our study, as noted above. However, we would recommend that special attention be given to a survey of the possibility of including model results, and the dynamic indicators that they represent, in any such new publication.

(3) Regarding the organization for preparing indicators for the Monthly Energy Review, we would recommend that this activity be carried on by a team of specialists within the staff of the FEA

Office of Data Policy. This would involve the transfer of the data base developed by this project to the FEA computer, the maintenance of this data base, and the addition of new series to it. Given this facility--which could be constructed with either of the management information systems used in this project (see Section 5)--continuing analysis and refining of existing indicators would take place, along with the design of new ones. High priority items for further development are noted in Section 3.

Such an activity would require two FEA staff to carry out: There is a need for one person experienced with energy data and with the use of these indicators, who has the analytical background and computer experience to construct simple models using a computer-based information and modeling system. And there would be a need for one data technician to maintain the data base itself.

(4) A good deal of the data used in this study had to be copied from the Monthly Energy Review, and either punched on cards or directly entered into a computer via a terminal. It was not possible to transfer data directly from FEA data banks to the data bank prepared for this project, or to address FEA computer data series directly. To some extent this was because information about those data facilities was not available, and to some extent it was because the systems were not compatible with each other. We recommend that a high priority be given to the preparation of clear documentation of the data management facilities now extant in the FEA and to the coordination of these various data banks through some common computer facility. Otherwise, all attempts to construct indicators either from raw data or from model results would be

hampered by the problem of gaining access to the information the FEA already has.

On the issue of the type of computational facility to use to create and maintain indicators, and on the even broader issue of the computational facility that FEA should develope to make their personnel effective, we recommend the following:

(5) Virtual Machine Consideration. The FEA is in the process of choosing an operating and data management system. In Appendix C we present a scheme using VM for analytical and Data Management uses in a way that heretofore has not been exploited. The scheme to using VM in this way and the advantages thereby derived should be weighed in FEA's present considerations of their applications. Our recommendation is that FEA adopt VM. We see it as the preferable system to meet FEA's needs. To choose one operating system and limit FEA to only those applications that run under that system seems unnecessarily restrictive, especially in view of the alternative that FEA could always run one operating system under VM (at an overhead cost) and hence the VM choice is only an improvement of the facility.

(6) Performance Study. We recommend that FEA initiate a study of the performance and cost sequences of adopting VM for their uses.

(7) Use of the GMIS System. It is possible for FEA to use the GMIS system in an experimental and limited fashion. We recommend that some FEA personnel become users of GMIS to gain firsthand knowledge of such facilities.

2. A FRAMEWORK FOR ENERGY INDICATORS

The use of indicators to record the course of the U.S. economy, and to look ahead to future developments, began early in this century. Wesley C. Mitchell first organized the method into a consistent framework [9], and over the years much has been added to his work by the National Bureau of Economic Research (NBER) [10], and by various federal agencies, most notably the Office of Business Economics, the Census Bureau, and the Bureau of Labor Statistics. In the energy sector, several federal agencies have collected and published data series, the most important work over the past years being that of the Bureau of Mines (BOM) [14], and with major responsibility now being taken up by the FEA. Several industrial groups also have collected energy data, the most detailed being that of the American Petroleum Institute (API) [1]. Though the energy data collected by BOM, and API and others were not usually finalized into "indicators" of the NBER type, many of the key aggregates have served this function in the past.

Clearly, of course, not all "useful data" is productively thought of as an "indicator", else the concept would be so broad as to be meaningless. Therefore, in this report we speak of an "indicator" as involving both data and some model or conceptual framework which is imposed on the data, either explicitly or implicitly, in order to derive useful information from the data. Thus, the discussion to follow is based on the following definition:

> An Indicator Set is a group of data series logically related to one another and to the energy economy by a model (or conceptual framework). "Indicator" may refer to either an indicator set or a single data series of an indicator set.

The conceptual framework or model associated with an indicator set may take many forms, as will be seen in succeeding sections. In the simplest case, the indicator may consist of a single series of raw data, interpreted by a "model" which is unstated or only implicitly understood. In the most advanced case a large number of data streams might be explicitly inter-realted by a formal simulation model, with stated assumptions about the confidence bounds of both model and data. The "output" of such an indicator would be not only the model and raw data, but also projections from simulation runs and various statistical analyses made possible by the explicit mathematical model.

Most existing sets of indicators--in particular, those published by the NBER [10]--fall somewhere in between the two extremes just mentioned. For example, the NBER indicator set consists of several series of economic aggregates, interrelated by a model that consists of both a formal part and an informal part. The formal part of the model consists of a list of attributes of the data series: whether each series is leading or lagging, the "batting average" of each series, etc. The informal part of the model consists of an implicit understanding of the dynamics of the U.S. economy and how the data series are related to it.

Naturally, the confidence or accuracy of an indicator is a function of the confidences associated with both the corresponding data and models. For example, if a model is vague or poorly formulated, no degree of accuracy in the data can guarantee accurate assessments or predictions. Similarly, even the most sophisticated model will produce unreliable output if it is coupled with poor data.

2.1 Types of Indicators

Under the definition above, indicators naturally fall into a hierarchy depending upon the degree of formalism and complexity associated with the model that lies behind each indicator set. In general, the more sophisticated and explicit the model associated with an indicator set, the more the data of the indicator set can be transformed into projections of the future, and the more accurately the confidence of those projections can be assessed.

Table 2.1 outlines the hierarchy, with the simplest indicator sets at the top of the table. The indicator sets are grouped into three broad categories:

1) Snapshot indicators are based on relatively unprocessed data. The model interrelating the different data series with each other and with the future is usually unstated. Snapshot indicators, if they assume anything about the future, assume constant flows. They may be presented as single numbers representing current values, or as series covering the past and present, but usually without explicit projection into the future. They are static in nature.

2) Dynamic indicators assume a causal structure of interactions among stocks and flows that project changes in both stocks and flows into the future. The model of changing flows may be simple, as in straight-line extrapolation of trends; or the model may be complex, as in a computer-simulation model. The presentation of data series associated with dynamic indicators may include graphs of expected future trends.

Indicator Class	Form of Model	Form of Indicators
Snapshot	no model : list of attributes : projection of flows as constants (e.g., stock/flow ratio	historical Data Series
Dynamic	: projection of flows as trends : integration of arbitrary future flows : simulation models	future projections added
Stochastic	: explicit modeling of error and uncertainty	confidence bounds added

Table 2.1

Hierarchy of Indicators

3) Stochastic indicators contain explicit descriptions of the error and uncertainty associated with both the data and the model underlying the indicator set. A simplifying assumption is less apt to cause trouble if it is accompanied by an estimate of the error likely to be caused by the simplification. Stochastic features may be incorporated in either snapshot or dynamic models. To the data series and the projected future trends are added uncertainty bounds, which show the numerical confidence of the indicator. One subset of the general category of stochastic indicators is event-probability indicators. Since one of the uses of an energy indicator set is to assess vulnerability to exogenous events (such as an oil embargo or coal strike), a natural extension of the indicator concept involves indicators of the likelihood of such events occurring.

Under the guidelines for this project, a major effort has been devoted to setting up the data and computer systems to demonstrate indicators in the forst category -- i.e., "snapshot" indicators. This effort is discussed in Section 3. (Section 4 presents the work performed on Dynamic and Stochastic Indicators.)

In the initial conception of this project, there was an emphasis on "snapshot" indicators that would have a predictive or "leading" character in the sense normally attributed to the NBER set for the national economy. Investigation of this prospect has led to the conclusion that it probably is not possible to construct clear leading indicators of this type for the key energy aggregates. There are several reasons for this result. First, under the trial of the past year, even the leading indicators of the U.S. economy as a whole have not lived up to their hoped-for performance. In the recent downturn, the index of all leading indicators in fact lagged the key events [3].[1] To the extent that key energy aggregates (particularly those concerned with demand) are related to these larger economic processes, one cannot expect similar indicators for the energy sub-sector to do any better at predicting future events. Second, most of the important energy aggregates are subject to many influences which may be very important, but which are not taken into account in the construction of a simple "snapshot" indicator.

Therefore it is useful, when thinking of "snapshot indicators, to consider separately the problems of status indication and forecasting. There are many indicators which are of great value in describing the current circumstance, and the path by which it was achieved. But multiple influences and short-range impacts abound throughout the energy economy, and few of these indicators qualify as "leading" indicators, in the sense that they reliably forecast movements in other statistics. Often variables at the early stages of the processes of demand or supply (e.g., appliance sales or exploratory wells drilled) may be thought of as "leading influences", which forecast changes later in the chain, but only if they are not counteracted by other influences. And thus static or "snapshot" indicators, in general, can be expected to lead only when everything else remains constant (a rare circumstance). Technological change; economic imbalances; changes in weather, international relations, or government market regulation; and other unexpected events may counteract any single effect revealed by a leading indicator.

1 A good discussion of these "snapshot" indicators and their limitations, compared with methods based on more explicit models, is provided by M.K. Evans [2], especially chapter 16.

In short, for the problem of forecasting medium and short range movements in the energy sector, we find that the number of indicators of the conventional "snapshot" kind that have a strong "leading" characteristic is likely to prove limited, and to require quite sophisticated interpretation (i.e., a complex implicit model in the mind of the interpreter). In effect, the achievement of improvements in forecasting leads one to the formulation of dynamic indicators (models) which take into account multiple effects from both above and below the point of interest in the supply-demand chain, and from competing fuels and products.

It should be emphasized that these cautionary notes regarding "snapshot" indicators are directed to their use for prediction. The accurate interpretation of the current circumstance in the energy sector, and of recent trends, remains a critically important function to be provided by "snapshot" indicators. The key national income aggregates and indicators constructed from them, become no less valuable because they do not predict the future. The need for good indicators is particularly evident if one considers that, while available energy data may be incomplete, it is at the same time too voluminous. There are tradeoffs between completeness and complexity in energy data and one of the roles of indicators is to help resolve the seeming contradiction of having too much data, but not enough. A prime role of indicators is to condense and simplify data, so that some idea of the condition of the energy sector can be seen in a few clear graphs, without painstaking analysis.

For example, even a single data series may be difficult to interpret without some notion of what it should be like under normal supply-demand conditions. In the construction of many of the indicators proposed below, the problem of interpretation is reduced by normalizing the data in various ways so that the indicator will tend to remain roughly constant if the energy system is functioning normally. Often this simple step helps make it easier to tell at a glance whether a particular activity is going "well" or "badly" orver a given time interval, which is the purpose of constructing indicators in the first place.

The obvious problem created by such normalized indicators, of course, is that information is lost as the separate data series are combined into a single indicator. For example, it a stock/flow ratio (such as the number of days supply of fuel oil) declines, it could be due to a change in the stock or a change in the flow, or a change in both. If one wants to find out why the indicator is doing what it is doing, then detailed analysis becomes necessary, and this may require not only the raw data which were used to compute the indicator, but additional data as well. Thus, the information system which supplies the indicators should be prepared to supply a great deal of back-up material.

The dilemma between indicators that are easy to interpret, and back-up data that are complete, is best resolved in the obvious way: compute the indicators, but also keep available and on reserve a reasonably complete set of raw data for detailed analysis and research. This approach will be followed in Section 3 below.

2.2 Categories of Energy Indicators

Figure 2.1 presents a framework which can be used to sturcture discussion of various indicators of energy sector performance, and how they relate to one another and to the overall issue of energy sufficiency. The left side of the figure shows the processes of energy provision. Given the fact that

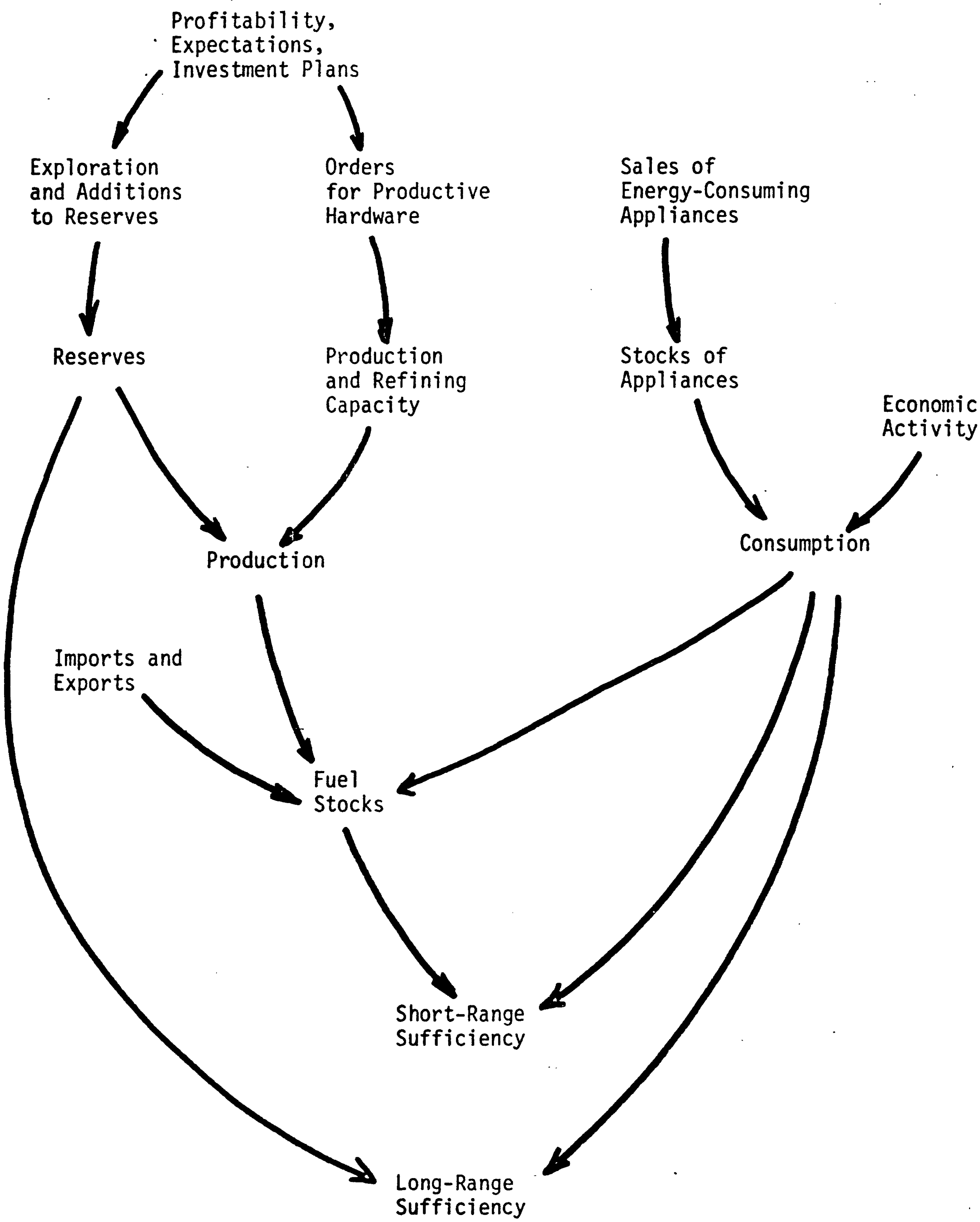


Figure 2.1 A Framework for Presenting Energy Indicators

energy supply in the United States is carried out primarily by private market institutions, the first step is the profitability (or expectations about the profitability) of investments in the energy sector, and resulting investment plans. Depending on plans and current expected return, investments will be made in exploration for oil and gas and other resources, which lead to increased reserves available to the domestic economy. Expectation of a return also leads to investment in hardware (such as coal mining equipment, drilling rigs, and electric power plants) which ultimately lead to increases in the production, processing, and transportation capacity of the energy sector.

The capacity thus created is used to produce domestic energy, which is supplemented by fuel imports (primarily oil) and decreased by exports (primarily coal). The buffer between this production process and the process of energy consumption is a set of fuel stocks at various points in the supply chain.

The right side of the figure shows the processes that lead to energy demand and consumption. Sales of energy-using appliances add to the stocks of devices. Their level of utilization, which in turn is influenced by the overall level of economic activity (and the availability of energy supply), results in energy consumption in the economy.

The concept of short-range sufficiency of energy as shown in the figure involves the interrelation of consumption patterns and available domestic production, augmented in the short run by fuel stocks. Long-range sufficiency involves the interpaly of longer term developments in demand and the development of reserves and the associated capacity to produce from them.

We have divided the various indicators into the following six groups, using Figure 2.1 as a framework.

2.2.1 Long-Run Domestic Supply

This category covers indicators in the top left corner of Figure 2.1, and includes profitability, expectations, and plans in energy industries; exploration and reserve additions, and the development of production, conversion, transport, and refining facilities.

2.2.2 Long Run Consumption

The determinants of energy consumption in the long-run include the available stocks of energy-using appliances, and the level of utilization of these appliances, as shown in the top right corner of Figure 2.1.

2.2.3 Short-Run Domestic Supply Adequacy

While long-run domestic supply of energy depends on exploration and investment in hardware, the sort-term supply must be drawn from existing domestic capacity, from stocks of fuels, or from world markets. Thus in Figure 2.1, indicators in this category are concerned with production, import, and exports, fuel stocks, and of course, current consumption. Most indicators in this area are of two types: (1) indications of the vulnerability of supply flows to disruption and the adequacy of stocks as a buffer, and (2) indication of actual supply shortage or constraints that cause consumption to be less than demand at current prices.

2.2.4 Prices

At each point in the diagram, relative prices are present implicitly. They determine the amount and composition of domestic supply and demand, the degree and direction of interfuel substitution, and the amount and composition of energy imports.

2.2.5 International Market

Since it is unlikely we shall eliminate fuel imports in the near future, it is important to have indicators of the condition of the international market, and the penetration of foreign supplies into the U.S. market. This point in the overall frame work of Figure 2.1 is noted "imports and exports."

2.2.6 Environmental and Social Impacts

There are environmental and social impacts associated with all phases of the energy network: supply, consumption and imports. Thus environmental and/or social indicators may be appropriate at many points in the framework of Figure 2.1.

3. SNAPSHOT INDICATORS

Among the many objectives of this project one of the more important sets of goals has been to

1) study possible sets of "snapshot" indicators,
2) develop prototype information systems for managing data and performing the needed analytical and graphical functions to produce energy indicators, and
3) demonstrate a prototype subset of the possible indicators using the information system chosen.

In this section, we present the results of the effort to develop snapshot indicators and information systems to support them. (The details of the information management systems themselves are presented in Section 5.)

3.1 Survey of Potential Indicators

One step in the process of meeting the objective laid out above was to pull together a complete list of all areas where suggestions had been made for the construction of indicators; this long list is presented as Appendix A. This list included items that spanned the range from excellent to uninteresting in terms of value, and from easy to impossible in terms of feasibility. It was reduced to a short-list by project personnel working closely with a representative to the Office of Energy Data Policy. The criteria by which items survived the cut were roughly the following:

1) Does it appear likely that the indicator or set of indicators can be defined in a precise manner?
2) Does it appear to answer an interesting question or illuminate an interesting issue, or is there an identifiable client for this information?

3) Is it likely that data for the indicator is readily available or might be obtained?

As a result of this selection process, the following list was developed.

1. Long-Run Domestic Supply

* 1.1 Profitability of energy companies
1.2 Amount of electricity generated by nuclear sources, and nameplate capacity.
* 1.3 Drilling rigs in operation; well completions.
* 1.4 Capital expenditures in petroleum and coal industry.
1.5 Distribution of sources of electrical energy, and projected distribution of planned generating capacity.
1.6 Back orders of drilling rigs, movable platforms, and drag lines.
* 1.7 New discoveries of each fuel (especially oil and gas) and natural gas wells drilled.
* 1.8 Drilling success rates, discoveries as a function of footage drilled and finding rates, and reserves of each fuel, especially oil, coal (by sulfur content), and gas.
1.9 Bottlenecks and construction lead times for new facilities
1.10 Availability (present units plus new units minus units replaced) of, domestic capacity to produce, and net exports of the following eight items:
    drilling rigs
    fixed drilling platforms
    mobile drilling platforms
    oil country tabular goods
    steel products
    steel pipe and tubing
    walking draglines
    steam turbine generators

* 1.11 Reserve/production ratio for crude oil and natural gas.
* 1.12 Corporate profits before taxes for industries in natural gas, crude petroleum, petroleum refining, coal, and electric utilities relative to all industries.
1.13 Revenues of utilities against expected revenues, given rate increases, consumption expectations and ability of utilities to raise funds on capital markets.
1.14 Energy investment as a percent of total business fixed investment and cumulative dollar investment for expansion of transportation network by 1985 in oil, gas, and coal industries.

2. <u>Long-Run Consumption</u>

2.1 Total energy consumed (BTU's per GNP dollar, current and constant dollars) and energy consumed, deleting energy used for heating.
2.2 Airline passenger load factors (BTU's/pass.mile).
* 2.3 Gasoline consumption of new autos sold (miles/gallon, monthly, sales-weighted-efficiency).
2.4 Natural gas curtailments
2.5 Rates of growth of consumption of different fuels (and categories like industrial, residential, commercial, transportation...)
* 2.6 Percentage share of energy consumption by all fuels
2.7 Percentage of total energy consumed as electricity
2.8 Average number of commuters per automobile
2.9 Home insulation consumption ($ sales and units sold)
2.10 Number of electric utility plants switched from oil & gas to coal
2.11 Number of current & new residential & commercial heating systems; by type of fuel used.

3. Short-Run Domestic Supply Adequacy

* 3.1 Days of supply remaining of petroleum, refined products, and coal.
* 3.2 Total domestic production of crude oil, refined products, and natural gas.

4. Price

* 4.1 Price of energy per BTU for all fuels (current & constant dollars).
  4.2 Average national BTU price (current & constant dollars)
  4.3 Transportation cost of coal (absolute and percent of total cost)
  4.4 World price of energy per BTU for all fuels (current and constant dollars).
* 4.5 Consumer and wholesale price indices for all fuels and electricity.

5. International Market

* 5.1 Excess production capacity among OPEC nations.
  5.2 International production and consumption for all fuels.
* 5.3 Total imports of crude petroleum and petroleum products

6. Environmental and Social Impacts

6.1 Environmental quality index for selected areas of the U.S.

6.2 For fossil fuel power plants, total number and percentage of the total of facilities with given types of pollution control devices, and current and projected installations of stack gas cleaning equipment, and any backlog of orders for such equipment.

6.3 Sulfur content of coal being mined, by region; and for the following residuals, the amount of environmental residual generated both by the energy sector and type of facility, and the amount of residual produced per amount of energy generated:

-acids (equivalent tons/day)

-bases (eqivalent tons/day)

-total dissolved solids (tons/day)

-suspended solids (tons/day)

-organics or oil spills

-thermal water pollution (BTU's/day)

-particulates (tons/day) ESP, lead, asbestos...

-nitrogen oxides (tons/day)

-ozone

-sulfur oxides (tons/day)

-hydrocarbons (tons/day)

-carbon monoxide (tons/day)

-aldehydes (tons/day)

-solids (tons/day)

-fixed land (acres/year) (alternative uses precluded for some time)

-incremental land (acres/year) (maximum excluded from alternative uses).

3.2 Development of Data Bank and Indicator Construction Facility

Using this list, an effort was made to investigate as many of the indicators as possible, given the limited time and resources available. A bank of energy data has been developed for the purpose of generating indicators. In the list above, an asterisk indicates each of the areas where actual indicators have been constructed for display here, or where all or a significant portion of the data needed to compute an indicator in the area are already in the data bank.

At present, the data are loaded on one of the two management information systems employed for this project. The bulk of the data are available on the TROLL system; a smaller subset of the data also are available on the GEMIS system (see Section 5). The basic data series now available on the system are documented in Appendix B.

It should be emphasized that not all the energy indicators that are available from the current data bank are discussed in this section:

(1) Many of the simple data series which are reported in Appendix B are themselves useful energy indicators for particular purposes. Indeed, the current FEA Monthly Energy Review contains little else but data series of this type. We single out only a few of these simple series for presentation in this section because of their special interest in relation to other indicators shown here or because they are of value in illustrating some point about the display of data of this type.

(2) Of the indicators that may be constructed by manipulating two or more data series, we have explored only a portion of those that might be constructed using the data bank already established. What has been provided is a facility whereby this task can be easily performed by FEA personnel as their interests lead them,

or as they respond to requests for information.

(3) The set of possible indicators can be greatly increased by adding more data series to the existing data bank.

In choosing the set of indicators to develop and document here, we have attempted a reasonable compromise between the desire to cover as many as possible of the most interesting issues in the energy area, the need in some areas for data searching and validation, and the limitations on time and resources. Clearly, the results shown here are only a step in the process of developing a complete set of indicators and establishing the procedures necessary for their maintenance, publication, and continuing refinement.

3.3 Prototype Indicators

In this section we present a prototype set of "snapshot" indicators. The indicators are grouped according to the breakdown of categories introduced in Section 2, and sample indicators are presented for five of the six categories. For each indicator, three types of information are provided:

(1) Documentation of the definition, interpretation and method of construction of the indicator

(2) A plot of the indicator

(3) A printout of the numerical values of the indicator itself and, in some cases, of some of the data series used to construct it.

It is suggested that this form of documentation and presentation would be very useful in any further efforts to expand or refine the indicator set developed here. The documentation is, of course, essential; the numerical data often prove very useful in interpreting and checking indicators and should be kept close at hand except, perhaps, where the limitation of mass publication forbid it.

3.3.1 Long-Run Supply

In the area of long-run supply, a set of 7 sample indicators are shown here. They focus on the stock of proved reserves of oil and natural gas, and its relation to current consumption. They include

- Crude Oil Reserve - Production Ratio (PET.RP)
- Natural Gas Reserve - Production Ratio (GAS.RP)
- Oil and Gas Reserve - Production Ratio (OG.RP)
- Crude Oil Additions to Reserves (PET.ADRS)
- Natural Gas Additions to Reserves (GAS.ADRS)
- Petroleum Reserve Adequacy
- Total Number of Rotary Drilling Rigs Running (RIGS60)

The reserve-production ratio is a conventional measure of long-run supply adequacy and a useful component of any set of sufficiency indicators. The strong decline in these indicators for oil and gas over the 1960's, a downward trend only broken by the large Alaskan finds, means that one of two things must happen. Either domestic output ultimately must also be reduced (for there is some minimum working inventory which must be maintained) or a dramatically increased rate of discovery must be attained, which in turn means that greatly increased incentives to exploration must be provided.

Another set of indicators shown in this section shows the pattern of additions to reserves over time. Here for petroleum the relative magnitude of the Alaskan finds in relation to the experience of recent years is particularly clearly demonstrated. Moreover, it is interesting to note how consistent the additions to oil reserves have been over the last 15 years (aside from Alaska), and how very poor the natural gas experience has been in the past 6 years (once again, aside from Alaska).

These phenomena are shown most dramatically by looking at the relationship between current reserves and the level of reserves that would be necessary to attain a prescribed level of self-sufficiency. The indicator of "Petroleum Reserve Adequacy" shows this comparison. The U.S. was roughly self-sufficient (in the sense that we could have held imports to no more than 15% of domestic needs) up until the late 1960's. But at that point, even with the Alaskan finds, the gap between needed reserves and actual began to grow. The extreme difficulty we will have in closing this gap is shown in an approximate way by the comparison of the size of this gap, and the normal rate of reserve additions shown in the former indicator.

Also shown below is one of the more common indicators of exploratory activity, the number of rotary drilling rigs running. This series is based on a long-term set of data published by the Hughes Tool Company, which dominates the market for drill bits.

Among the many possible directions of further work in this area, the following deserve high priority:

(1) Data and indicators regarding drilling effort in feet (a supplement to the indicator of total rigs running), drilling costs, and finding rates per foot drilled.

(2) Data and indicators on the effort expended in secondary and tertiary treatment methods, and the results.

(3) Data and indicators on the financial health of the electric utilities.

DOCUMENTATION FOR PET.RP

NAME: Crude Oil Reserve-Production Ratio

DEFINITION: The ratio of proved crude oil reserves at end of year to annual crude oil production.

INTERPRETATION: If one assumes that no new crude oil reserves are found, that crude production continues at a constant rate, and all other factors (prices, technology, etc.) remain unchanged, then this indicator shows how much time remains before proved reserves are fully depleted.

UNITS Years

FREQUENCY: Annual

INPUTS: PET.RSVS--Proved reserves of crude oil (estimated as of December 31 of any given year) are the estimated quantities of all liquids statistically defined as crude oil, which geological and engineering data demonstrate with reasonable certainty to be recoverable in future years from known reservoirs under existing economic and operating conditions. Source: API, AGA, CPA, Bluebook.

PET. PROD--Crude oil production is the volume of liquids statistically defined as crude oil, which is produced from oil reservoirs during a year. The amount of such production is generally established by measurement of volumes delivered from leased storage tanks (i.e., tne point of custody transfer) to pipelines, trucks, or other media for transport to refineries or terminals. Source: API, AGA, CPA, Bluebook.

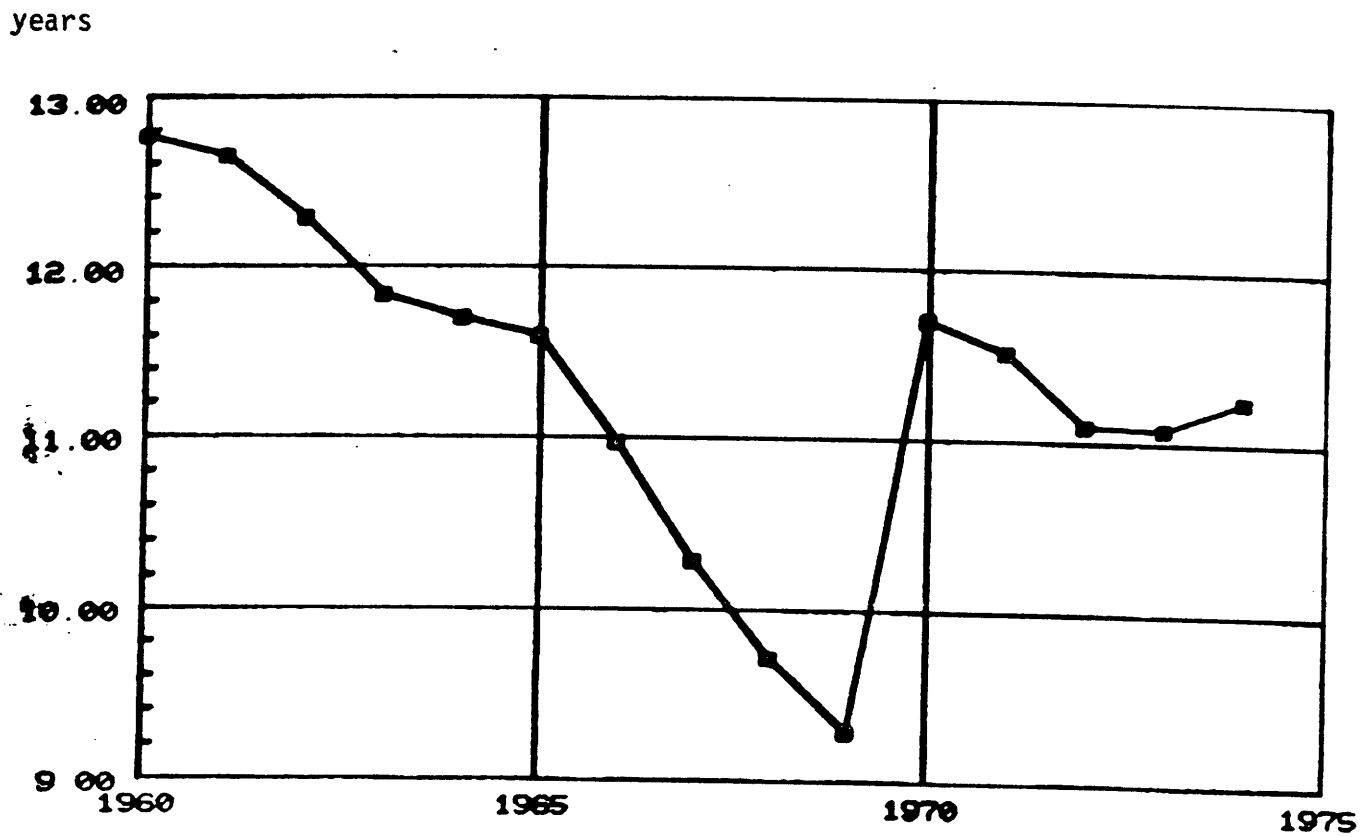
FORMULA:

$$\frac{\text{Proved Crude Oil Reserves}}{\text{Crude Oil Production}} = \frac{\text{PET.RSVS}}{\text{PET.PROD}} \; \frac{\text{(bbl)}}{\text{(bbl/year)}}$$

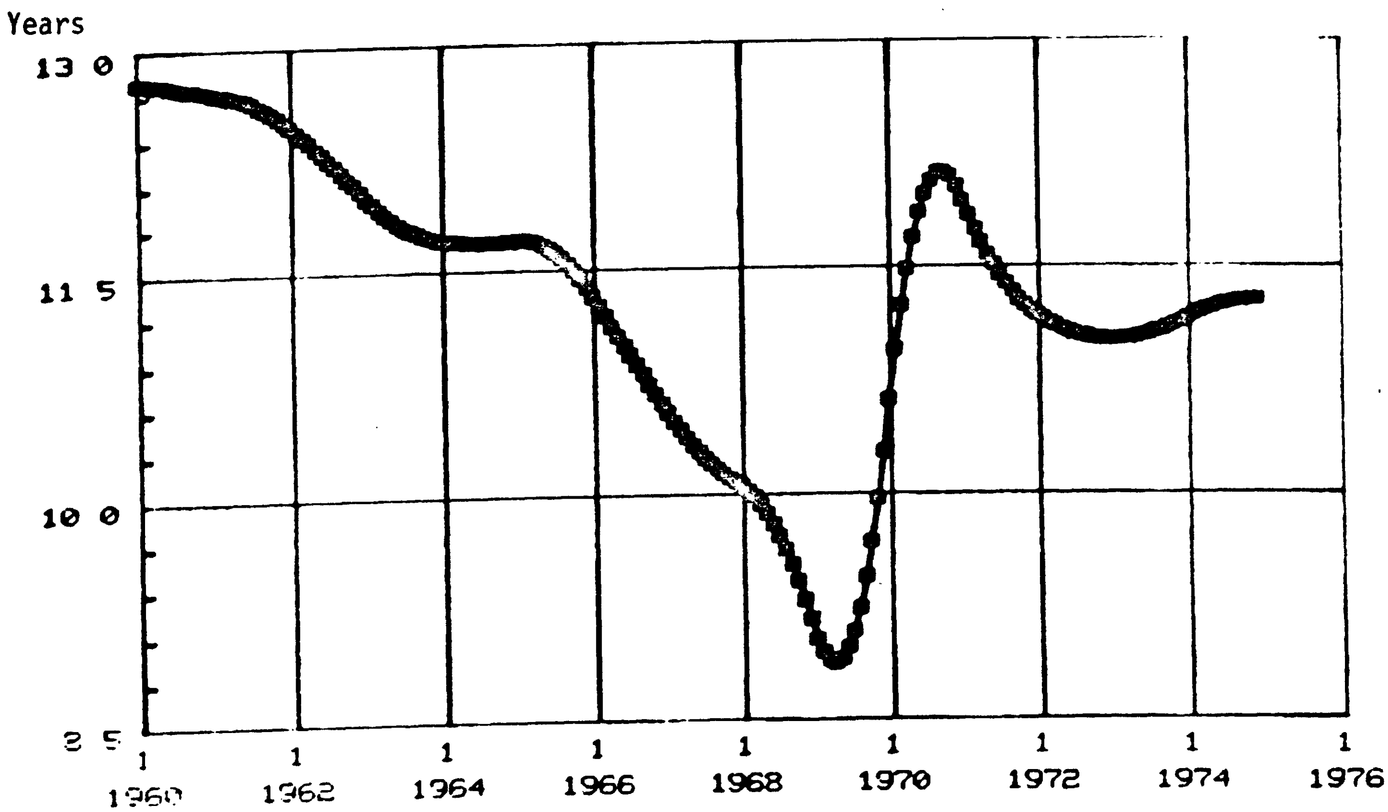
OUTPUT: Vertical axis on graph is in years and horizontal in time. Table of data is also given.

CONTACT: John Curtis, FEA, Office of Energy Data Policy, (202) 961-7986.

Crude Oil Reserve-Production Ratio



Crude Oil Reserve-Production Ratio (smoothed)



Crude Oil Reserve-Production Ratio

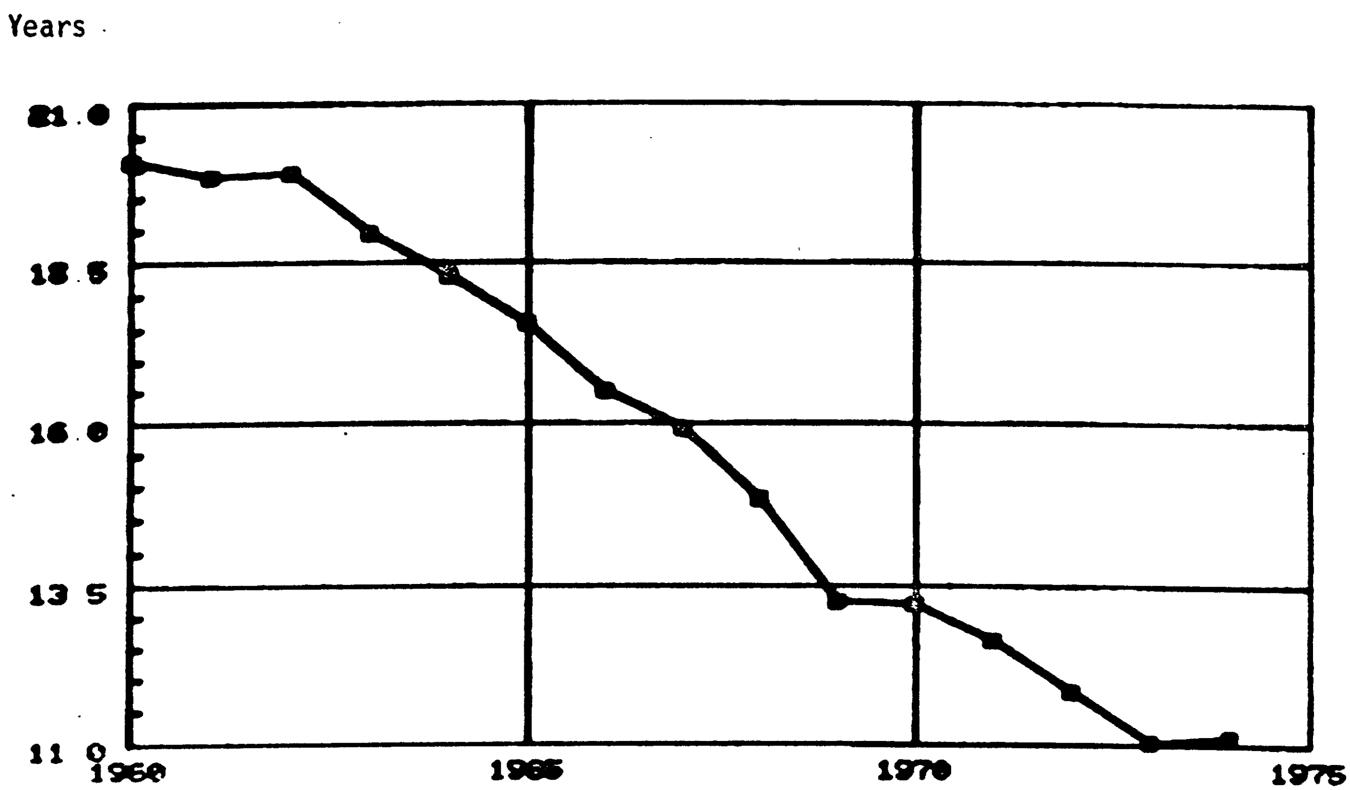
COLCOMB(PET.RSVS,PET.PROD,PET.RP)

	PET.RSVS	PET.PROD	PET.RP
1960	31613.	2473.	12.7833
1961	31758.	2508.	12.6627
1962	31389.	2552.	12.2998
1963	30969.	2613.	11.8519
1964	30990.	2645.	11.7164
1965	31352.	2699.	11.6162
1966	31452.	2864.	10.9818
1967	31376.	3047.	10.2973
1968	30707.	3161.	9.71433
1969	29631.	3189.	9.29163
1970	39001.	3328.	11.7191
1971	38062.	3299.	11.5374
1972	36339.	3274.	11.0993
1973	35299.	3185.	11.0829
1974	34250.	3043.	11.2553
	10^6 bbl	10^6 bbl/yr.	

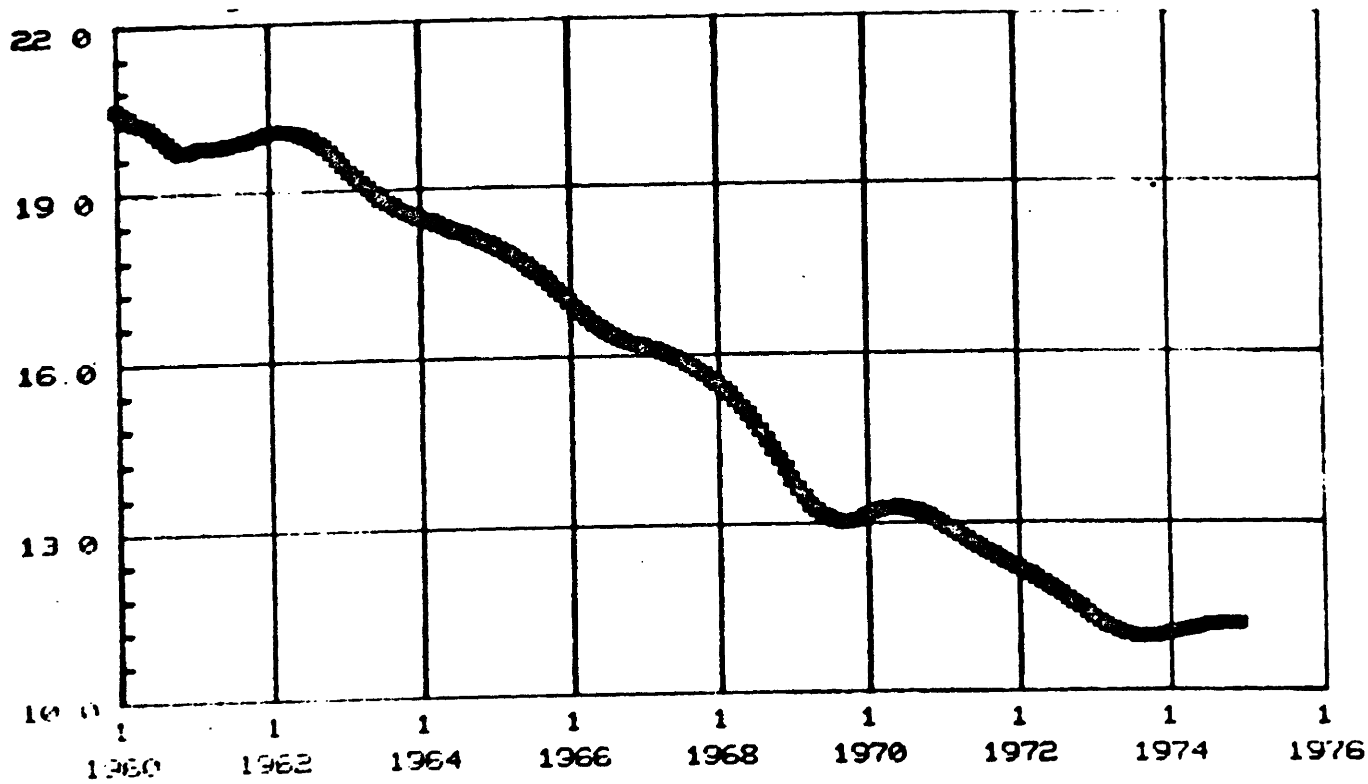
DOCUMENTATION FOR: GAS.RP

NAME:	Natural Gas Reserve-Production Ratio
DEFINITION:	The ratio of proved natural gas reserves at end of year to annual natural gas production.
INTERPRETATION:	If one assumes that no new natural gas reserves are found, that natural gas production continues at a constant rate, and all other factors (prices, technology, etc.) remain unchanged, then this indicator shows how much time remains before natural gas is fully depleted.
UNITS:	Years
FREQUENCY:	Annual
INPUTS:	GAS.RSVS - proved natural gas reserves GAS.PROD - production out of reserves of natural gas and natural gas liquids.
FORMULA:	$\frac{\text{GAS.RSVS}}{\text{GAS.PROD}} \quad \frac{ft^3}{ft^3/yr}$
SOURCE:	Internally generated. See documentation of inputs for their source information.
COMMENTS:	Data available from 1960 through 1973. The slowing of decline in 1970 results from the discovery of the Alaskan reserves.

Natural Gas Reserve-Production Ratio



Natural Gas Reserve-Production Ratio (smoothed)



Natural Gas Reserve-Production Ratio

COLCOMB(GAS.RSVS,GAS.PROD,GAS.RP)

	GAS.RSVS	GAS.PROD	GAS.RP
1960	262326.	13019.	20.1495
1961	266274.	13379.	19.9024
1962	272279.	13638.	19.9647
1963	276151.	14546.	18.9847
1964	281251.	15347.	18.3261
1965	286469.	16252.	17.6267
1966	289333.	17491.	16.5418
1967	292908.	18381.	15.9354
1968	287350.	19373.	14.8325
1969	275109.	20723.	13.2755
1970	290746.	21961.	13.2392
1971	278806.	22077.	12.6288
1972	266085.	22512.	11.8197
1973	249950.	22605.	11.0573
1974	237132.	21318.	11.1236
	10^9 ft^3	10^9 ft^3/yr.	

DOCUMENTATION FOR: OG.RP

NAME: Oil and Gas Reserve-Production Ratio

DEFINITION: This ratio of proved oil and natural gas reserves (in BTU) at the end of the year to the annual production (in BTU) of these fuels.

INTERPRETATION: If one assumes that no new reserves of crude oil and natural gas are found, and that production of these fuels in aggregate BTU's continue at a constant rate, then this indicator shows how much time remains before the aggregate of these fuels is fully depleted.

UNITS: Years

FREQUENCY: Annual

INPUTS: PET.RSVS, GAS,RSVS, PET.PROD, GAS.PROD

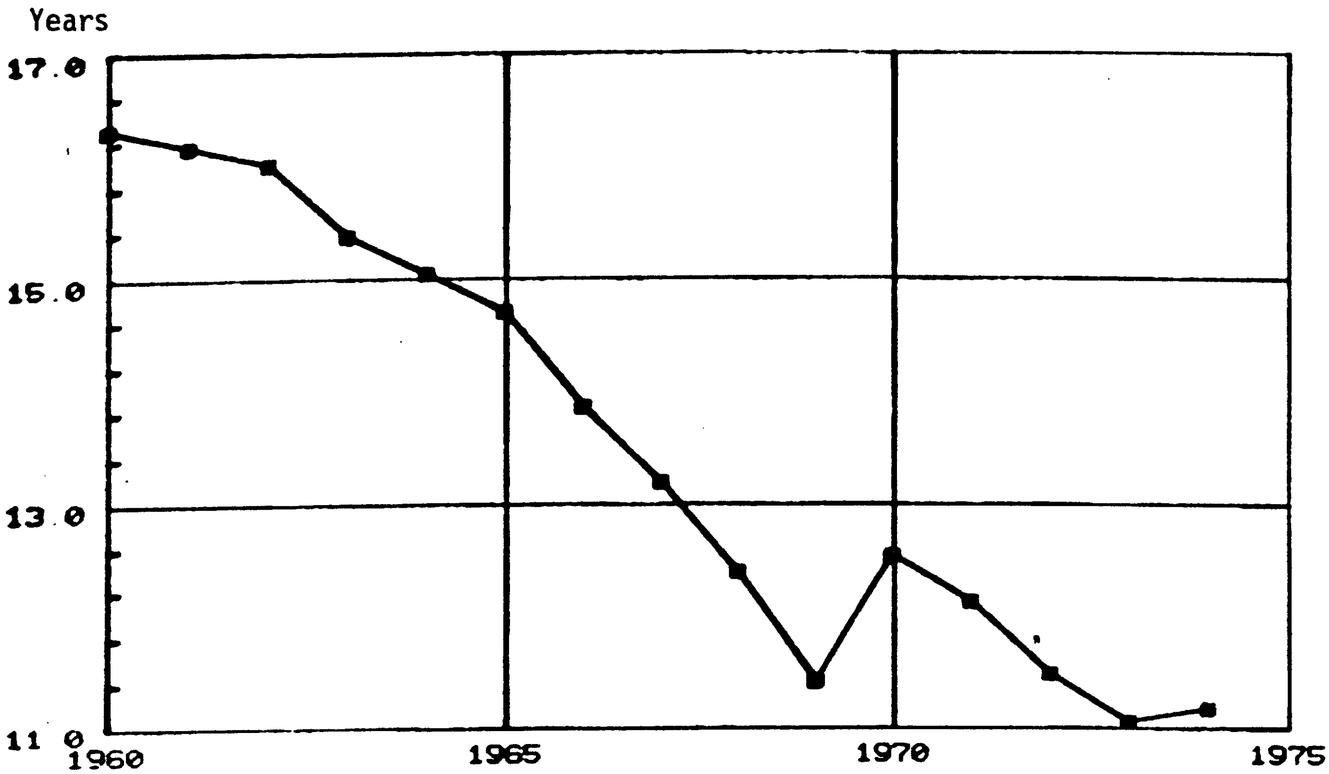
FORMULA:

$$\frac{\text{PET.RSVS} * 5.8 \times 10^{6} + \text{GAS.RSVS} * 1030}{\text{PET.PROD} * 5.8 \times 10^{6} + \text{GAS.PROD} * 1030} \quad \frac{\text{(BTU)}}{\text{(BTU/year)}}$$

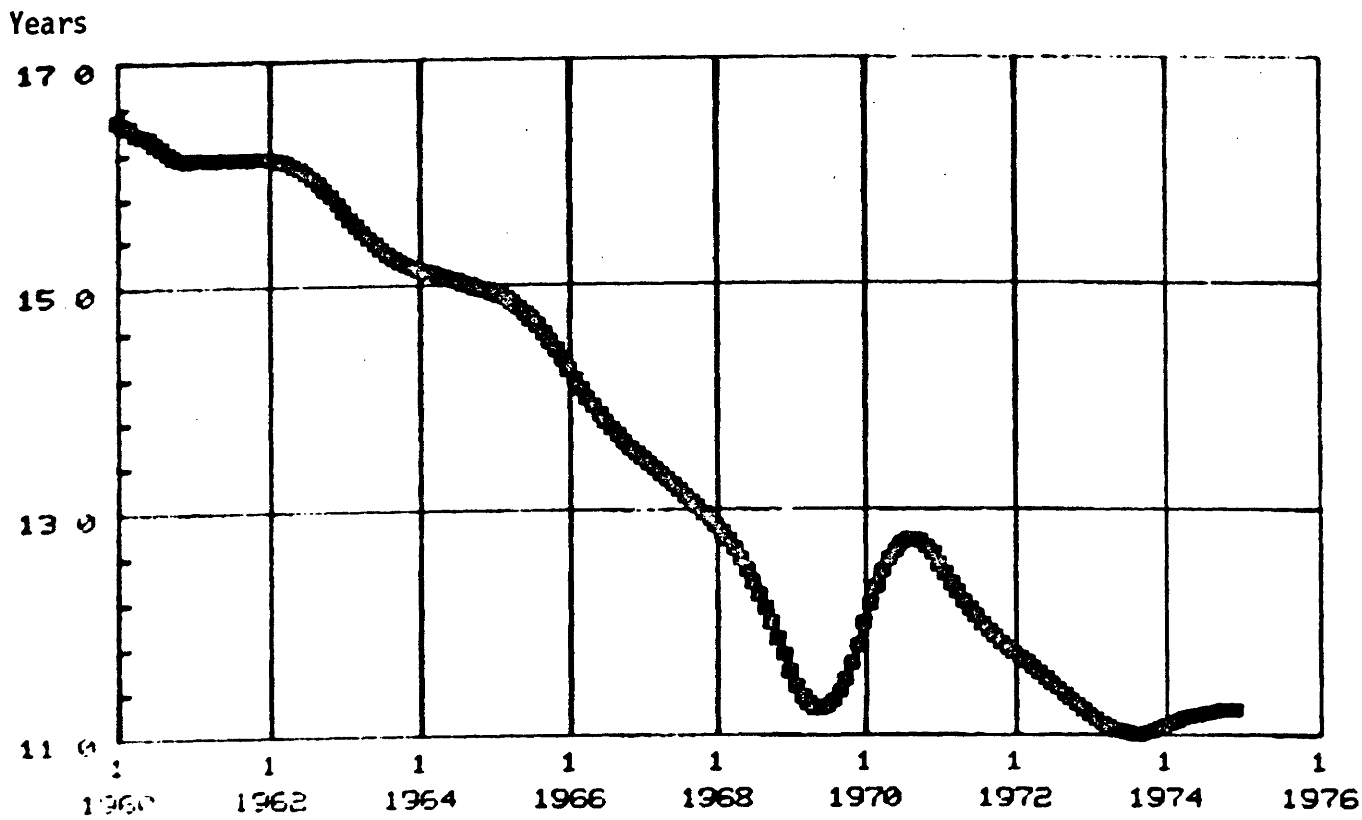
OUTPUT: Vertical Axis in years, horizontal in time.

CONTACT: John Curtis, FEA, Office of Energy Data Policy, (202) 961-8796.

Oil and Gas Reserve-Production Ratio



Oil and Gas Reserve-Production Ratio (smoothed)



Oil and Gas Reserve-Production Ratio

COLCOMB(OG.RSVS,OG.PROD,OG.RP)

	OG.RSVS	OG.PROD	OG.RP
1960	4.535510E+17	2.775296E+16	16.3424
1961	4.584583E+17	2.832677E+16	16.1846
1962	4.625034E+17	2.884873E+16	16.032
1963	4.640555E+17	3.013778E+16	15.3978
1964	4.694302E+17	3.114840E+16	15.0708
1965	4.769045E+17	3.239376E+16	14.7221
1966	4.804344E+17	3.462692E+16	13.8746
1967	4.836759E+17	3.660591E+16	13.2134
1968	4.740708E+17	3.828797E+16	12.3817
1969	4.552217E+17	3.984088E+16	11.426
1970	5.256740E+17	4.192222E+16	12.5393
1971	5.079296E+17	4.187349E+16	12.1301
1972	4.848336E+17	4.217655E+16	11.4953
1973	4.621826E+17	4.175613E+16	11.0686
1974	4.428959E+17	3.960693E+16	11.1823
	10^6 bbl	10^6 bbl/yr.	

DOCUMENTATION FOR: PET.ADRS

NAME: Crude Oil Additions To Reserves

DEFINITION: The incremental amount of crude reserves for a given rate of production in period t, where t = year.

INTERPRETATION: This indication gives a rough feeling for the amount and success of domestic exploratory efforts.

UNITS: Millions of barrels.

FREQUENCY: Yearly

INPUTS: PET.RSVS--Proved reserves of crude oil (estimated as of December 31 of any given year) are the estimated quantities of all liquids statistically defined as crude oil, which geological and engineering data demonstrate with reasonable certainty to be recoverable in future years from known reservoirs, under existing economic and operating conditions. Source: API, AGA, CPA, Bluebook.

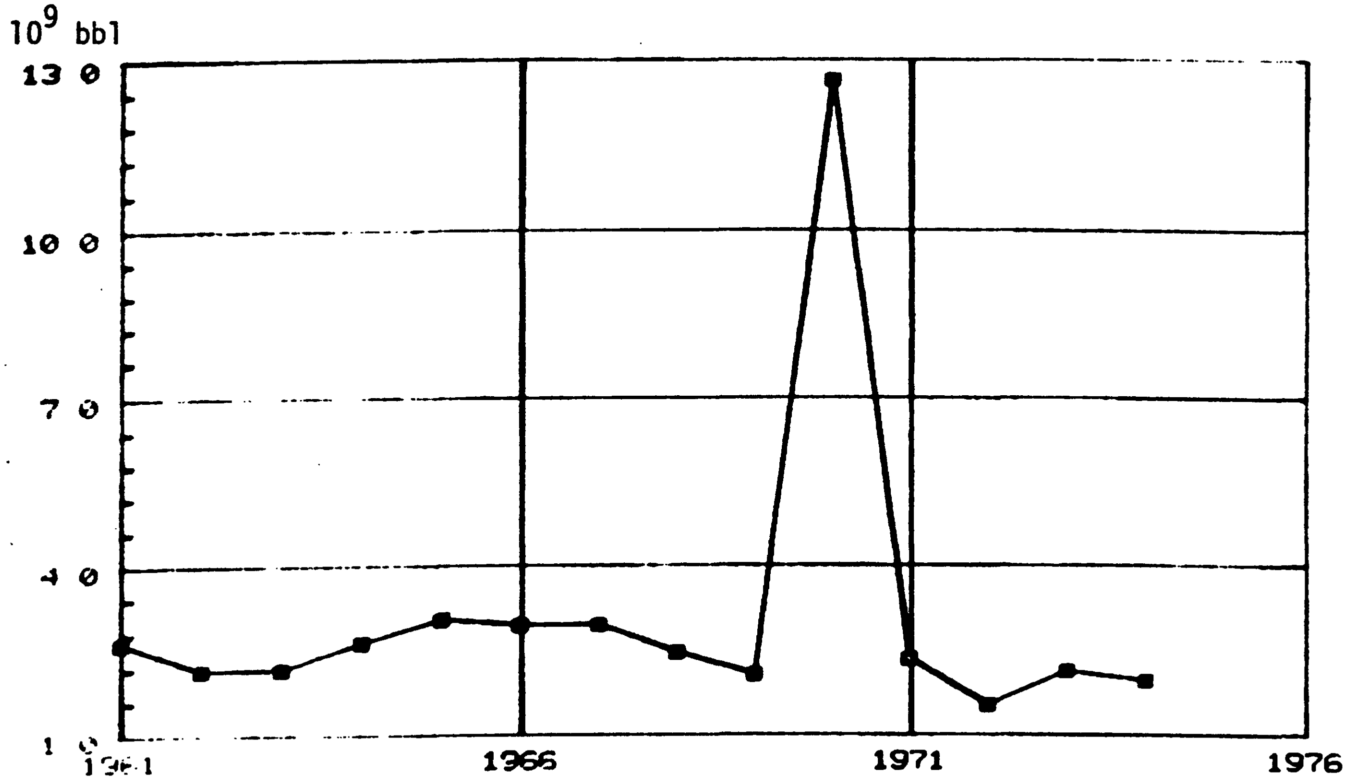
PET.PROD--Crude oil production is the volume of liquids statistically defined as crude oil, which is produced from oil reservoirs during a year. The amount of such production is generally established by measurement of volumes delivered from least storage tanks (i.e., the point of custody transfer) to pipelines, trucks, or other media for transport to refineries or terminals. Source: API, AGA, CPA, Bluebook.

FORMULA: $PET.RSVS_t - PET.RSVS_{t-1} + PET.PROD_t$

OUTPUT: Vertical axis is additions to reserves, horizontal axis is time in years.

CONTACT: John Curtis, FEA, Office of Energy Data Policy, (202) 961-8796.

Crude Oil Additions To Reserves



Crude Oil Additions to Reserves

PET.ADRS

	ANNUAL DATA
1961	2653.
1962	2183.
1963	2193.
1964	2666.
1965	3061.
1966	2964.
1967	2971.
1968	2492.
1969	2113.
1970	12698.
1971	2360.
1972	1551.
1973	2145.
1974	1994.

10^6 bbl

DOCUMENTATION FOR: GAS.ADRS

NAME: Natural Gas Additions To Reserves

DEFINITION: The incremental amount of gas reserves for a given rate or production in period t, where t = year.

INTERPRETATION: This indication gives a rough feeling for the amount and success of domestic exploratory efforts.

UNITS: Trillions of cubic feet.

FREQUENCY: Yearly

INPUTS: GAS.RSVS--Proved reserves of natural gas (estimated as of December 31 of any given year) which geological and engineering data demonstrate with reasonable certainty to be recoverable in future years from known reservoirs, under existing economic and operating conditions. Source: API, AGA, CPA, Bluebook.

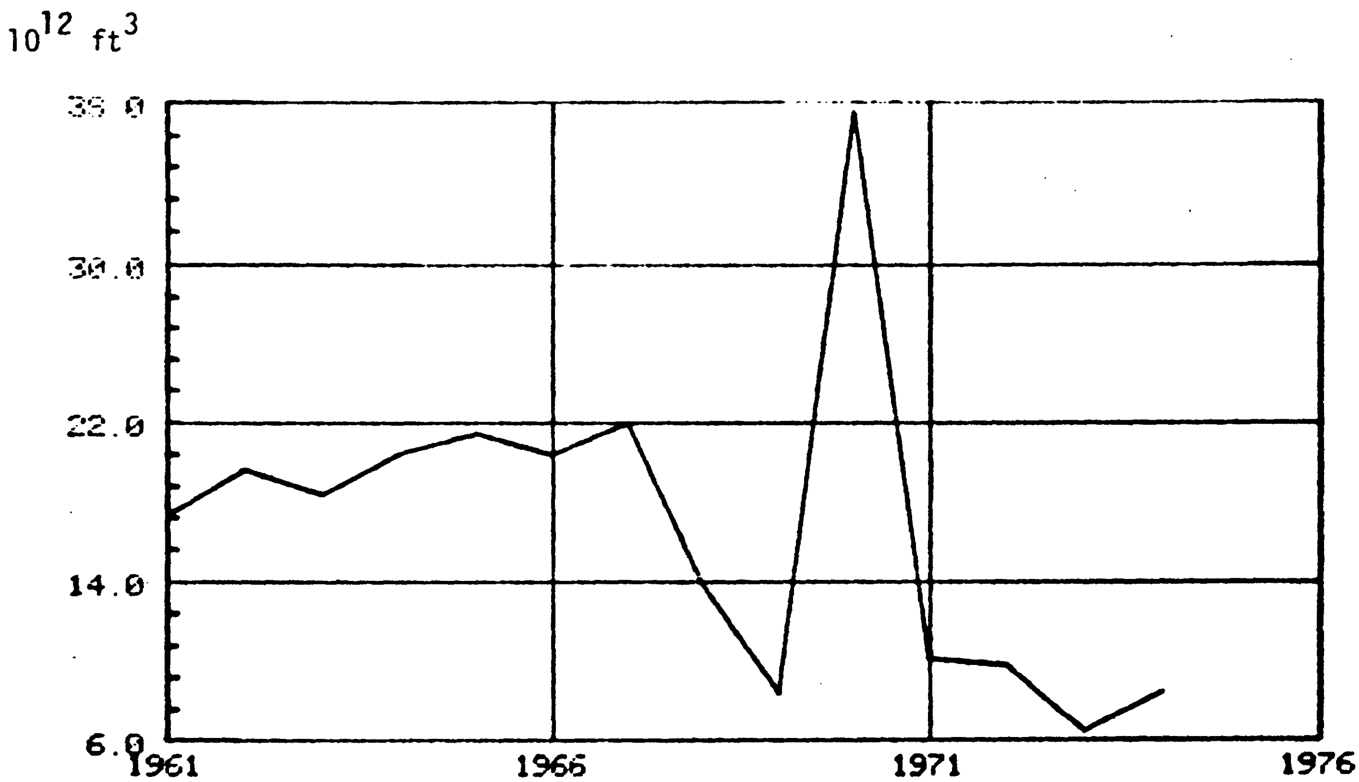
GAS.PROD--Crude oil production is the volume of gas which is produced during a year. Source: API, AGA, CPA, Bluebook.

FORMULA: $GAS.RSVS_t - GAS.RSVS_{t-1} + GAS.PROD_t$

OUTPUT: Vertical axis is additions to reserves, horizontal axis is time in years.

CONTACT: John Curtis, FEA, Office of Energy Data Policy, (202) 961-8796.

Natural Gas Additions To Reserves



Natural Gas Additions To Reserves

GAS.ADRS

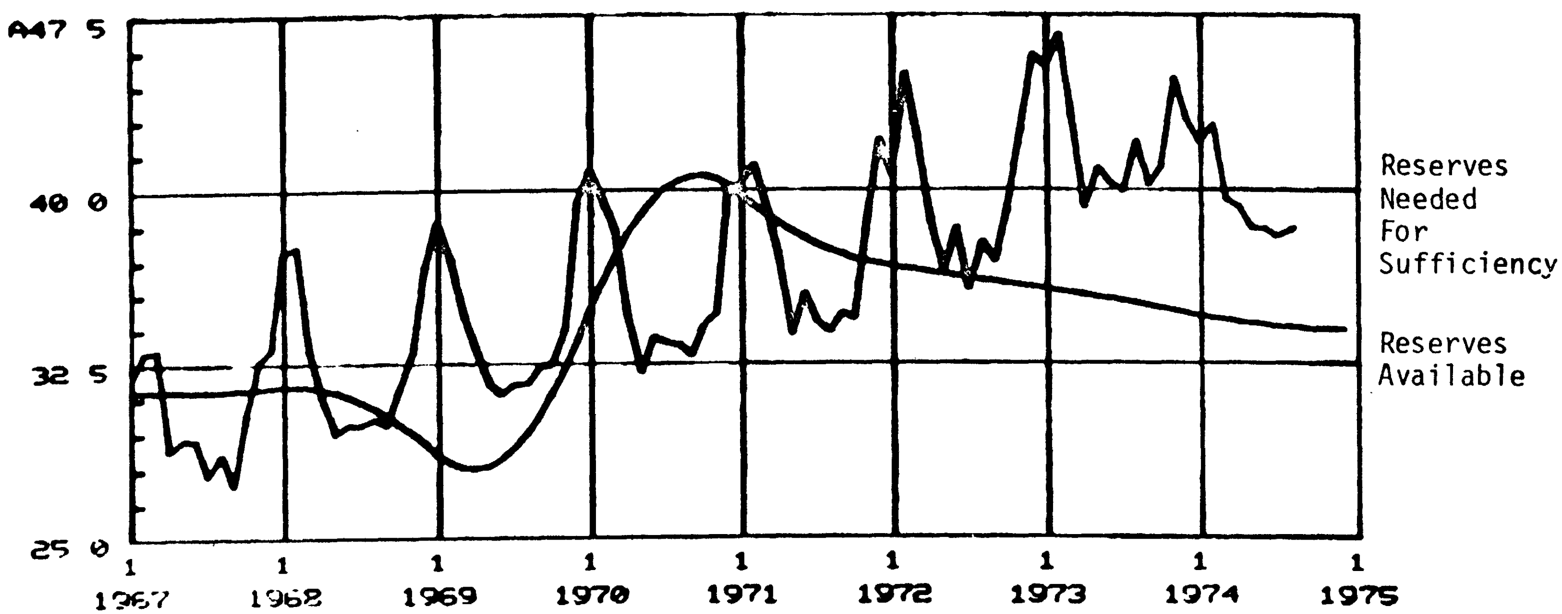
	ANNUAL DATA
1961	17327.
1962	19643.
1963	18418.
1964	20447.
1965	21470.
1966	20355.
1967	21956.
1968	13815.
1969	8482.
1970	37598.
1971	10137.
1972	9791.
1973	6470.
1974	8500.

10^9 ft^3

DOCUMENTATION FOR: PET.NEED, PET.RSVS

NAME:	Petroleum Reserve Adequacy
DEFINITION:	U.S. petroleum reserves compared with the reserves needed to maintain 85% domestic supply with a 10 to 1 reserve-production ratio.
INTERPRETATION:	Indicates long-run trends in the petroleum sector and shows the exploratory effort needed to gain sufficiency in this fuel.
UNITS:	Billions of barrels.
FREQUENCY:	Monthly
INPUTS:	PET.RSVS, DS.DMD, GS.DMD, RS.DMD, JT.DMD
FORMULA:	PET.NEED = (DS.DMD + GS.DMD + RS.DMD + JT.DMD) * .365 * 10 * .85
OUTPUT:	Vertical axis is 10^9 bbl, horizontal axis is time.
CONTACT:	John Curtis, FEA, Office of Energy Data Policy, (202) 961-8796.

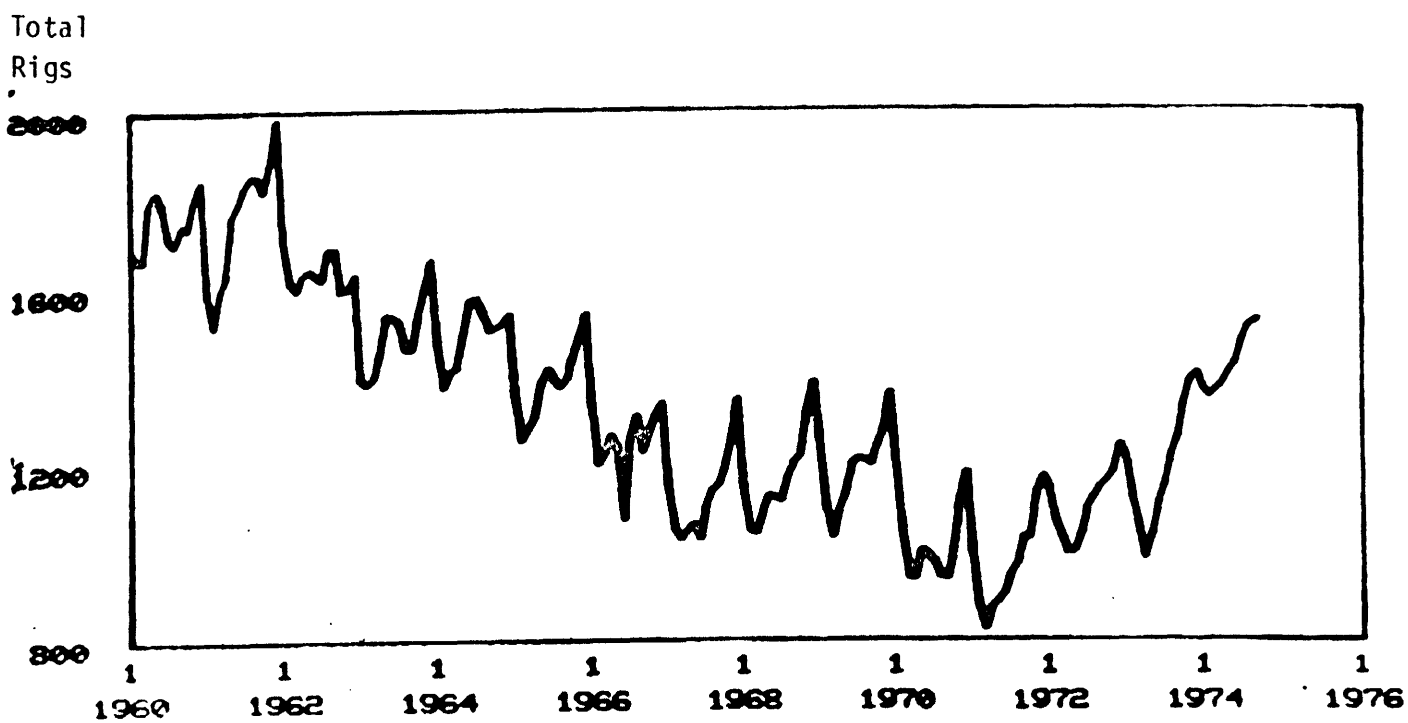
Petroleum Reserve Adequacy



DOCUMENTATION FOR: RIGS60

NAME:	Total Number of Rotary Drilling Rigs Running
DEFINITION:	Indicates total number of rotary drilling rigs for U.S., excluding: cable tools, stacked rigs and rigs moving to new locations.
INTERPRETATION:	Domestic discoveries of oil and gas and subsequent production cannot take place without the required drilling equipment represented by RIGS60.
UNITS:	Numbers of rigs
FREQUENCY:	Monthly
SOURCE:	Hughes Tool Co.
COMMENTS:	Data available from January, 1960 to May, 1975 (estimated).

Rotary Drilling Rigs Running



Rotary Drilling Rigs Running

TRANSP(RIGS60)

	1960	1961	1962
JANUARY	1692.	1595.	1711.
FEBRUARY	1670.	1518.	1619.
MARCH	1667.	1600.	1602.
APRIL	1793.	1631.	1640.
MAY	1822.	1765.	1646.
JUNE	1789.	1793.	1634.
JULY	1716.	1832.	1624.
AUGUST	1701.	1855.	1692.
SEPTEMBER	1746.	1853.	1689.
OCTOBER	1736.	1822.	1598.
NOVEMBER	1804.	1886.	1602.
DECEMBER	1843.	1983.	1633.

	1963	1964	1965
JANUARY	1395.	1481.	1359.
FEBRUARY	1383.	1378.	1256.
MARCH	1400.	1417.	1286.
APRIL	1455.	1420.	1308.
MAY	1548.	1506.	1385.
JUNE	1538.	1577.	1416.
JULY	1530.	1580.	1393.
AUGUST	1467.	1549.	1377.
SEPTEMBER	1466.	1510.	1393.
OCTOBER	1537.	1514.	1442.
NOVEMBER	1608.	1525.	1491.
DECEMBER	1664.	1548.	1546.

	1966	1967	1968
JANUARY	1336.	1160.	1151.
FEBRUARY	1200.	1058.	1054.
MARCH	1227.	1036.	1050.
APRIL	1266.	1055.	1092.
MAY	1237.	1070.	1133.
JUNE	1080.	1038.	1128.
JULY	1253.	1107.	1120.
AUGUST	1308.	1146.	1176.
SEPTEMBER	1230.	1156.	1206.
OCTOBER	1272.	1191.	1220.
NOVEMBER	1312.	1253.	1320.
DECEMBER	1337.	1345.	1386.

Rotary Drilling Rigs Running (Continued)

	1969	1970	1971
JANUARY	1260.	1207.	1007.
FEBRUARY	1103.	1045.	872.
MARCH	1039.	946.	828.
APRIL	1094.	946.	876.
MAY	1135.	1010.	892.
JUNE	1201.	997.	911.
JULY	1212.	984.	954.
AUGUST	1211.	949.	980.
SEPTEMBER	1200.	946.	1039.
OCTOBER	1238.	1004.	1037.
NOVEMBER	1276.	1120.	1138.
DECEMBER	1361.	1182.	1172.

	1972	1973	1974
JANUARY	1147.	1219.	1372.
FEBRUARY	1071.	1126.	1355.
MARCH	1034.	1049.	1367.
APRIL	1002.	993.	1381.
MAY	1005.	1046.	1412.
JUNE	1049.	1118.	1432.
JULY	1104.	1155.	1480.
AUGUST	1130.	1222.	1518.
SEPTEMBER	1152.	1266.	1527.
OCTOBER	1165.	1334.	NA
NOVEMBER	1186.	1390.	NA
DECEMBER	1241.	1405.	NA

	1975
JANUARY	NA
FEBRUARY	NA
MARCH	NA
APRIL	NA
MAY	NA
JUNE	NA
JULY	NA
AUGUST	NA
SEPTEMBER	NA
OCTOBER	NA
NOVEMBER	NA
DECEMBER	NA

3.3.2 Long-Run Consumption

As an example of the possible development of indicators of long-run consumption, two series are shown here. Both relate to gasoline consumption. The first indicates the level of sales of passenger cars at retail; it shows a growth in this market over a period of a decade and a half, and then rapid decline in recent months as a result of changes in the economy and the energy sector. Second is an indicator of the energy-using characteristics of these new vehicles. This indicator can be used to monitor one of the stated goals of current energy policy, which is to reduce the consumption of the incoming car fleet by 40% over a period of years.

There is a wide variety of indicators that might be constructed to reflect trends in long-run consumption. Many require the collection and manipulation of new data series, however, or of data series that were not compiled for purposes of studying energy phenomena. The following deserve the earliest attention:

(1) Improved indicators of automotive fuel use. With a moderate amount of additional data collection and the construction of a set of simple models, it should be possible to construct an indicator set which reveals something about the several determinants of gasoline consumption (fleet age, distribution of new vehicles, miles travelled) and how they are evolving over time. With a very simple forecasting model a very useful dynamic indicator set might be produced.

(2) An indicator set showing the intensity and efficiency of domestic air travel should not be difficult to construct.

(3) From detailed heating fuel supply data and degree-day data (or direct estimates of K-factors) from specific regions of the country (perhaps New England) it should prove possible to produce an indicator set that would show conservation trends in home heating.

DOCUMENTATION FOR: RCAR6D

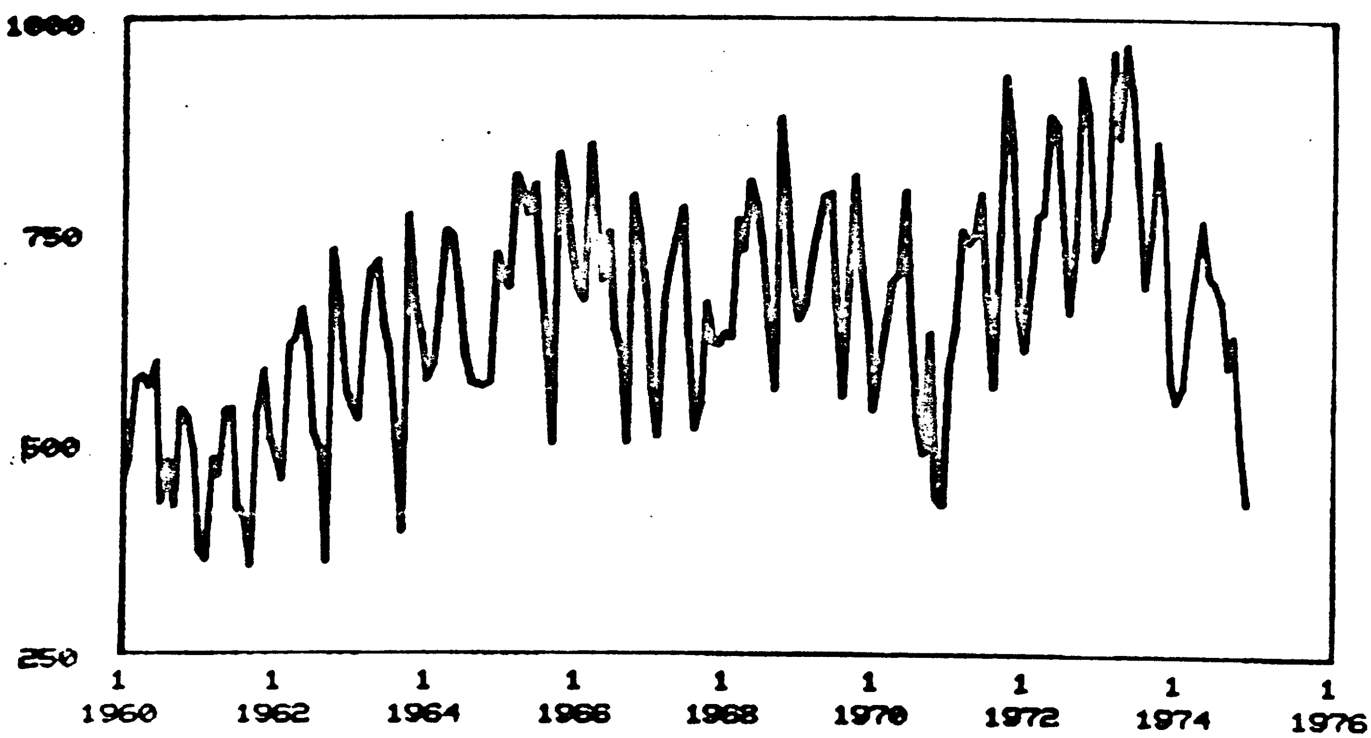
NAME: Number of New Domestic Passenger Cars Sold at Retail

UNITS: Thousands of autos

SOURCE: U.S. Department of Commerce, Social and Economic Statistics Administration, Bureau of Economic Analysis, "Survey of Current Business"

COMMENTS: Data available from 1958.

Passenger Cars Sold At Retail



Passenger Cars Sold At Retail

Thousands Of Cars

TRANSP(RCAR6D)

	1960	1961	1962
JANUARY	455.	370.	486.
FEBRUARY	483.	360.	457.
MARCH	576.	482.	616.
APRIL	579.	460.	621.
MAY	566.	540.	657.
JUNE	594.	539.	609.
JULY	428.	420.	512.
AUGUST	479.	410.	492.
SEPTEMBER	423.	351.	356.
OCTOBER	540.	535.	729.
NOVEMBER	531.	585.	657.
DECEMBER	488.	504.	561.

	1963	1964	1965
JANUARY	544.	573.	695.
FEBRUARY	527.	592.	684.
MARCH	650.	679.	817.
APRIL	704.	753.	800.
MAY	715.	743.	773.
JUNE	636.	676.	807.
JULY	606.	603.	712.
AUGUST	502.	571.	610.
SEPTEMBER	392.	568.	499.
OCTOBER	771.	566.	842.
NOVEMBER	664.	570.	801.
DECEMBER	624.	724.	722.

	1966	1967	1968
JANUARY	684.	564.	630.
FEBRUARY	668.	509.	624.
MARCH	854.	670.	767.
APRIL	765.	710.	729.
MAY	692.	745.	811.
JUNE	751.	780.	781.
JULY	635.	627.	737.
AUGUST	608.	517.	635.
SEPTEMBER	501.	547.	563.
OCTOBER	794.	665.	885.
NOVEMBER	746.	618.	785.
DECEMBER	678.	615.	679.

Passenger Cars Sold At Retail (Continued)

	1969	1970	1971
JANUARY	645.	539.	586.
FEBRUARY	662.	598.	637.
MARCH	722.	646.	756.
APRIL	754.	691.	737.
MAY	795.	699.	748.
JUNE	798.	800.	798.
JULY	662.	641.	668.
AUGUST	555.	526.	566.
SEPTEMBER	709.	489.	756.
OCTOBER	817.	630.	934.
NOVEMBER	706.	436.	848.
DECEMBER	639.	425.	649.

	1972	1973	1974
JANUARY	610.	736.	551.
FEBRUARY	698.	775.	568.
MARCH	772.	964.	654.
APRIL	774.	863.	703.
MAY	883.	972.	767.
JUNE	877.	909.	698.
JULY	769.	808.	691.
AUGUST	656.	686.	668.
SEPTEMBER	741.	754.	591.
OCTOBER	932.	858.	628.
NOVEMBER	891.	778.	506.
DECEMBER	719.	574.	430.

DOCUMENTATION FOR CAR.AMVI

NAME: Average Gasoline Mileage of Domestic New Cars Sold by U. S. Dealers

DEFINITION: The average mileage-per-gallon of domestic-make new cars sold by U. S. dealers. This average is computed monthly and is weighted according to monthly sales volumes of new automobiles.

INTERPRETATION: This indicator is useful in assessing the degree to which conservation of motor gasoline is being achieved by the introduction of more fuel-efficient automobiles into the domestic market.

The "Comments" listed below should be read and considered carefully, since they suggest possible deficiencies of the indicator, as previously calculated.

UNITS: Miles per gallon

FREQUENCY: Monthly

INPUTS: SALES--domestic-make new car sales by U. S. dealers, by make and model of car. Monthly, from Ward's Automotive Reports.

MPG--mileage-per-gallon of new cars. Annually, from Environmental Protection Agency's Gas Mileage Guide for New Car Buyers.

FORMULA: Let:

model = model of car

date = date (monthly)

mpg (model) = miles-per-gallon of the specified model of car;

for 01/74 $\leq$ date $\leq$ 10/74, mpg(model) is obtained by multiplying the miles-per-gallon listed in the EPA's 1974 Gas Mileage Guide for Car Buyers by an adjustment factor of 1.227. This adjustment factor is used to compensate for 45% highway driving, and to account for a 5% adjustment in EPA testing methods.

for 11/74 $\leq$ date $\leq$ 02/75, mpg(model) is obtained from the data in the EPA's 1975 Gas Mileage Guide for New Car Buyers, according to the following formula:

$$\text{mpg(model)} = \frac{1}{\dfrac{.55}{\text{urban mpg}} + \dfrac{.45}{\text{highway mpg}}} .$$

sales(date, model) = sales of given model of car during the month specified by date; from Ward's Automotive Reports.

The indicator itself is obtained by the following formula, for a given month specified by date:

$$\left[\frac{\sum_{\text{model}} \dfrac{\text{sales(date, model)}}{\text{mpg(model)}}}{\sum_{\text{model}} \text{sales(date, model)}} \right]^{-1}$$

However, see also the "Comments" below, since due to data limitations, the calculation of the indicator differs from the above formula.

COMMENTS: In producing the accompanying graph, several simplifications were made, as listed below:

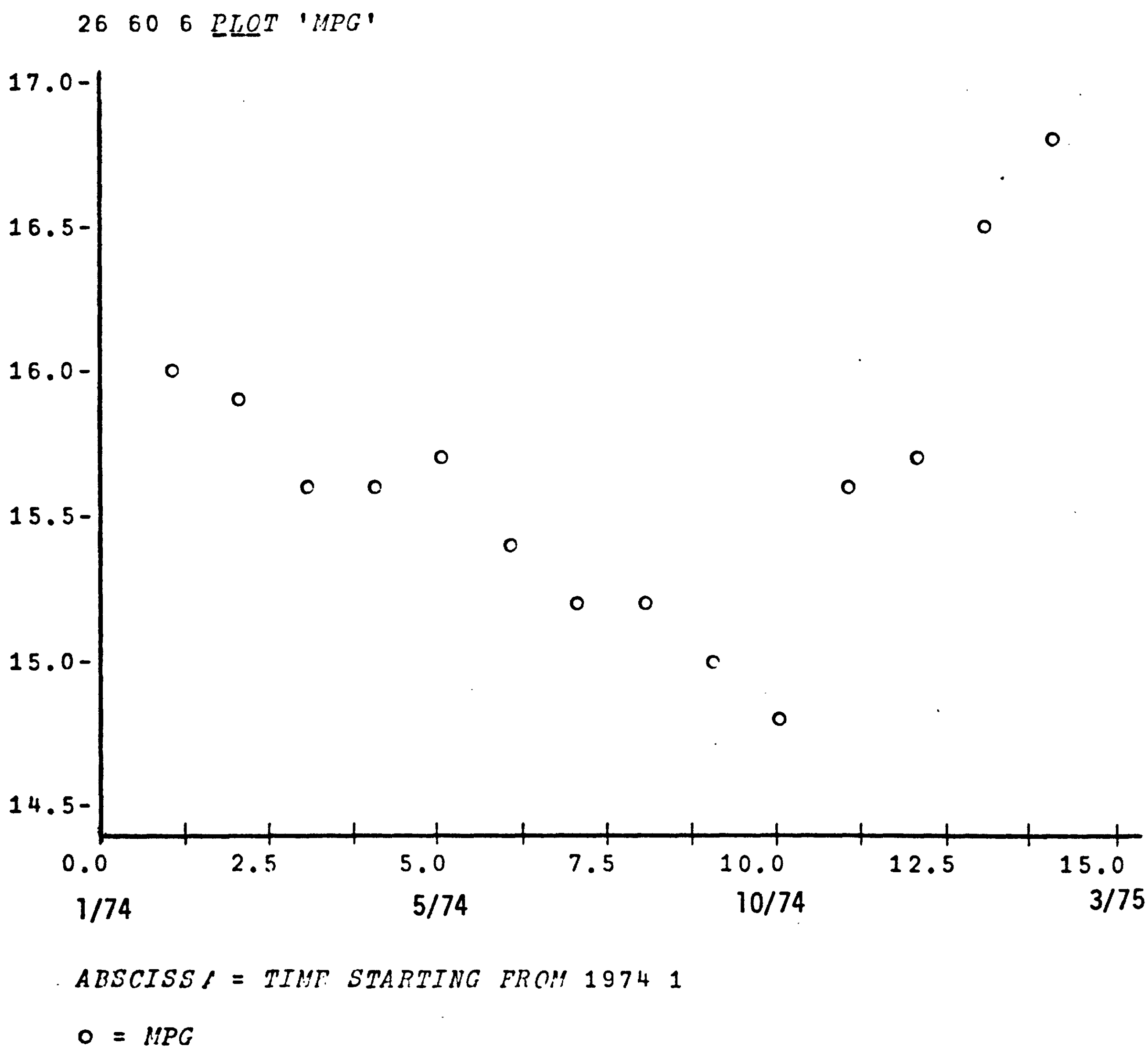
1. For all new cars sold from January 1974 through October 1974, the miles-per-gallon data for 1974 model cars were used, when available. For all new cars sold from November 1974 through February 1975, the miles-per-gallon data for 1975 model cars were used, when available.

2. In cases where miles-per-gallon data for a given model and year of car were not available, data were used for a different year (1974 or 1975) car, when available. (For example, if miles-per-gallon data on 1975 cars of a given make were not available, but 1974 data were available, then the 1974 miles-per-gallon data were used in place of the 1975 miles-per-gallon, in the formula for calculating the indicator.)

3. In cases where the miles-per-gallon data were not available for either the 1974 or 1975 cars of a given model, that model was omitted in the indicator calculation. Such omissions constituted about 5% of the total domestic car sales by U. S. dealers.

4. In cases where more than one miles-per-gallon figures were listed for a given model car of a given year, the data corresponding to the smallest engine size were used, when the data were listed according to engine size. In other such cases, the data for what was believed to be a common or similar version of the specified model were used.

5. The formula given for calculating the average miles-per-gallon weights cars on the basis that each car will be driven the same number of miles during a given time period. (Rather than, for instance, so that the cars will be driven so that each consume the same amount of gasoline, during a given time period.)



Average Gasoline Mileage of New Cars

3.3.3 Short-Run Domestic Supply Adequacy

A set of 18 indicators are shown in this area:

- Crude Oil Domestic Production (CR.PDD)
- Days Supply of Crude Oil (CR.STFL)
- Days Supply of Distillate Oil (DS.STFL)
- Days Supply of Gasoline (GS.STFL)
- Days Supply of Residual Fuel (RS.STFL)
- Days Supply of Jet Fuel (JT.STFL)
- Days Supply of Petroleum Products (PET.STFL)
- Average Daily Imports of Crude Oil (CR.IMD)
- Average Daily Imports of Distillate Oil (DS.IMD)
- Average Daily Imports of Motor Gasoline (GS.IMD)
- Average Daily Imports of Residual Fuel (RS.IMD)
- Average Daily Imports of Jet Fuel (JT.IMD)
- Fraction of Crude Oil Imported (CR.FI)
- Days Supply of Crude Imports (CR.STIM)
- Percent of Imports from Insecure Sources (AR.IMD, CE.IMD)
- Days Supply of Petroleum Imports (PET.STIM)
- Days Supply of Insecure Imports
- Import Dependence

The first of these indicators simply represents the domestic production of crude oil and shows that after peaking out at about the turn of the decade, domestic crude oil production has declined in recent years. It differs from the series published in the January Monthly Energy Review only in that the graphic presentation of the data seems more clear, and the series is longer.

The next five indicators show stock-flow ratios for the various petroleum products and for crude oil and petroleum products as a whole. These are once again conventional measures of short-run supply adequacy and represent the ability of the economy to absorb a short-run shock or disruption in either domestic or foreign energy supplies. As with the stock-import indicators below, these series do suffer from one serious flaw. They show only primary stocks of petroleum products; the secondary stocks held by consumers are not included. Since secondary stocks are significant in some sectors (e.g., residual fuel oil, heating oil), such indicators both understate the stocks available and, in some circumstances, fail to reveal significant stock changes which are taking place.

Next are a set of indicators, which show the average daily imports of various fuels and of crude oil, are no more than revised presentations of data presented in the Monthly Energy Review.

Next, is a very useful indicator of domestic energy sufficiency, when taken in concert with other series--the fraction of crude oil imported. This series shows how much of the domestic refining capacity is dependent on imports for feed stocks. As one might expect, it has risen significantly over the last ten years or so. Then there is a set of indicators that is probably the most useful of all given current concern about international petroleum markets. They show the ratios of petroleum primary stocks to imports and to insecure imports based on data about the composition of imports provided by FEA.

Aside from the obvious need to understand more about the characteristics of secondary stocks, there are several priority areas for work on short-term supply indicators:

(1) Natural gas curtailments and/or unsatisfied demand

(2) Reserve ratios in electric power systems

(3) Stock-flow conditions in various parts of the coal industry

(4) Expressions of vulnerability of particular regions of the country (e.g., the Northeast to disruption of imports).

DOCUMENTATION FOR: CR.PDD

NAME: Crude Oil Domestic Production

DEFINITION: The average daily volume of crude oil, including lease condensate, flowing out of the ground at the wellhead, each month.

INTERPRETATION: This indicator provides the absolute volume of domestic crude oil production. It is a basis for a variety of indicators of domestic self-sufficiency in petroleum production.

UNITS: Million bbls./day

FREQUENCY: Monthly

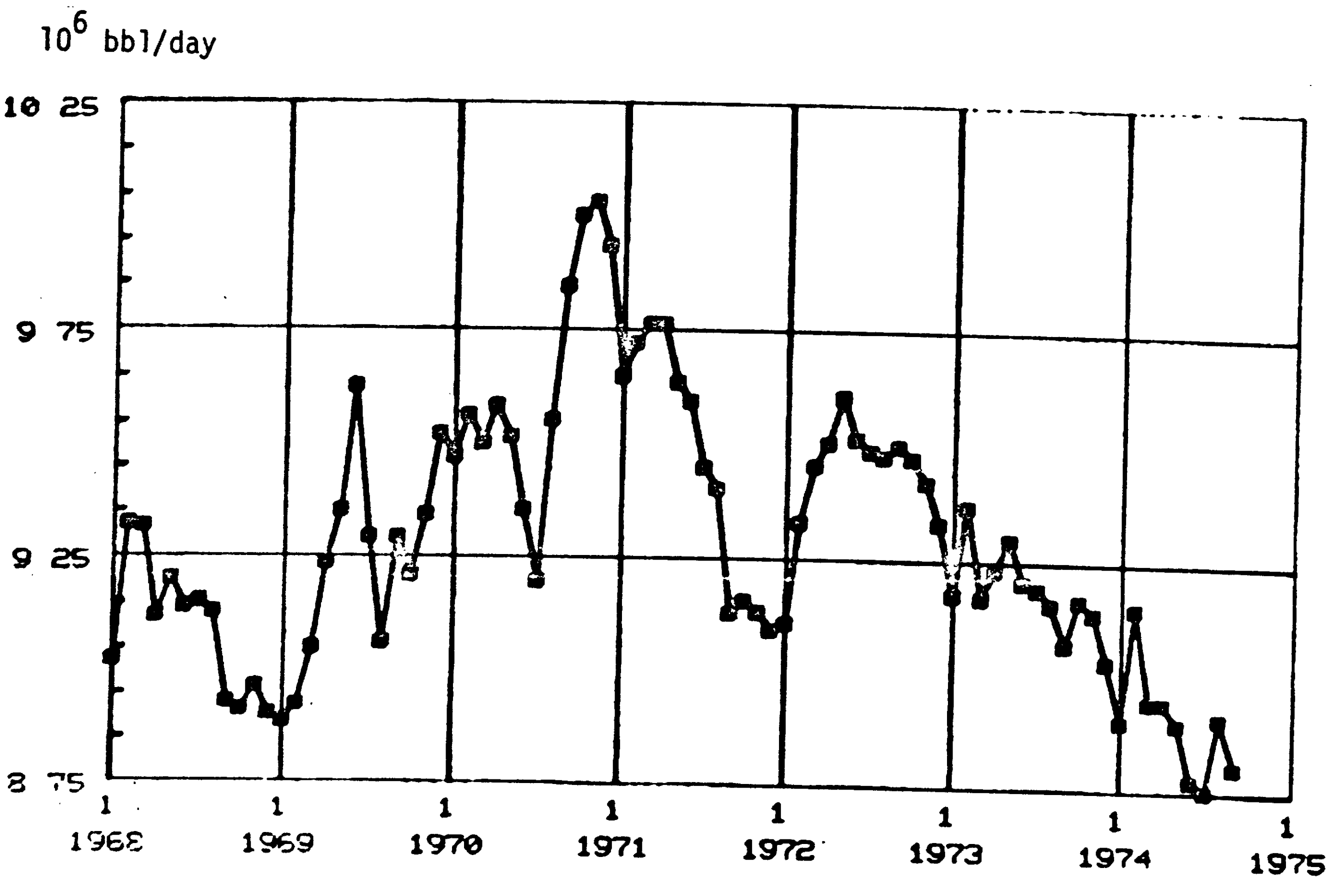
FORMULA: $\frac{\text{Monthly Crude Oil Domestic Production}}{\text{\# of calender days/month}}$

OUTPUT: CR.PDD is graphed where vertical axis is thousand bbls/day and horizontal axis is time.

SOURCE: Directly from MER. Source is BOM through April 1974, FEA from May 1974 thereafter.

CONTACT: John Curtis, FEA, Office of Energy Data Policy, (202) 961-7986.

Crude Oil Domestic Production



Crude Oil Domestic Production

10^3 bbl/day

	1968	1969	1970
JANUARY	9027.58	8887.77	9478.45
FEBRUARY	9325.	8927.82	9570.14
MARCH	9319.06	9055.32	9507.77
APRIL	9122.9	9238.46	9591.23
MAY	9205.03	9355.55	9523.13
JUNE	9143.96	9631.4	9358.83
JULY	9156.32	9295.48	9200.87
AUGUST	9133.87	9066.9	9560.19
SEPTEMBER	8932.4	9295.07	9853.1
OCTOBER	8916.19	9212.64	10013.2
NOVEMBER	8969.2	9345.57	10044.4
DECEMBER	8905.52	9528.16	9944.09

	1971	1972	1973
JANUARY	9655.	9114.29	9179.
FEBRUARY	9723.71	9336.	9373.
MARCH	9768.03	9461.64	9175.
APRIL	9768.93	9512.96	9233.
MAY	9644.52	9614.29	9303.
JUNE	9603.9	9521.53	9209.
JULY	9456.58	9496.29	9195.
AUGUST	9410.64	9482.52	9161.
SEPTEMBER	9135.57	9508.3	9077.
OCTOBER	9162.32	9481.58	9172.
NOVEMBER	9139.	9426.43	9144.
DECEMBER	9099.61	9334.61	9041.

	1974
JANUARY	8907.
FEBRUARY	9156.
MARCH	8950.
APRIL	8952.
MAY	8903.
JUNE	8777.
JULY	8754.
AUGUST	8913.
SEPTEMBER	8809.
OCTOBER	NA
NOVEMBER	NA
DECEMBER	NA

DOCUMENTATION FOR: CR.STFL

NAME: Days supply of crude oil

DEFINITION: Number of days that U.S. could continue to consume crude oil at the demand rate in period t given a level of stocks in period t, where t = month.

INTERPRETATION: This indicator provides a rough guide to how vulnerable crude oil stocks are to any exogenous shocks either domestic or foreign. The indicator should be used and applied judiciously because movement in it can be associated with either the numerator or denominator of the ratio.

UNITS: Days

FREQUENCY: Monthly

INPUTS: CR.DMD--Domestic demand for crude oil in thousand bbl/day. From MER. Source: BOM through April 1974, FEA from May 1974 forward.

CR.ST-- Primary stocks of crude oil held by producers in thousands of bbl. From MER. Source: BOM through April 1974, FEA from May 1974 thereafter.

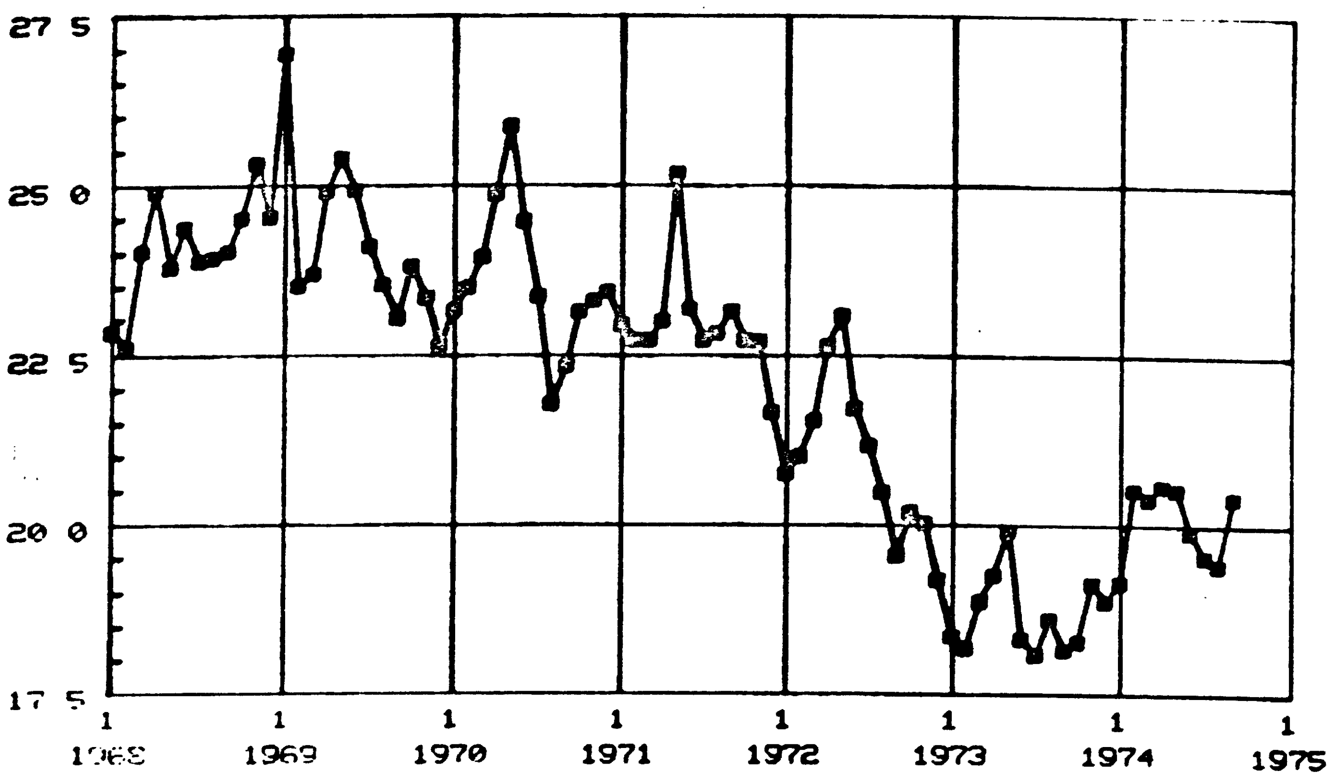
FORMULA:

$$\frac{\text{Stocks of crude}_t}{\text{Domestic demand for crude}_t} = \frac{\text{CR.ST}}{\text{CR.DMD}} \quad \frac{\text{(bbl)}}{\text{(bbl/day)}}$$

OUTPUT: CR.STFL is graphed where vertical axis is days and horizontal axis is time. Seasonal adjustment also shown. Table of data used to produce graph is also shown.

CONTACT: John Curtis, FEA, Office of Energy Data Policy, (202) 961-7986

Days Supply of Crude Oil



3-23b

Days Supply of Crude Oil

	1968	1969	1970
JANUARY	22.8653	26.9897	23.1934
FEBRUARY	22.6312	23.5481	23.5226
MARCH	24.067	23.713	23.983
APRIL	24.9538	24.9544	24.9173
MAY	23.8106	25.4447	25.9176
JUNE	24.4192	24.9717	24.5209
JULY	23.9104	24.1296	23.3838
AUGUST	23.9489	23.553	21.8207
SEPTEMBER	24.0629	23.0714	22.3655
OCTOBER	24.5555	23.8367	23.1745
NOVEMBER	25.3756	23.3632	23.3226
DECEMBER	24.5808	22.6638	23.4554

	1971	1972	1973
JANUARY	22.9717	20.7913	18.3797
FEBRUARY	22.7542	21.032	18.2069
MARCH	22.7486	21.5826	18.9084
APRIL	23.018	22.6647	19.2811
MAY	25.2246	23.115	19.9318
JUNE	23.2056	21.7546	18.3362
JULY	22.7453	21.1953	18.0987
AUGUST	22.8457	20.5057	18.6507
SEPTEMBER	23.1698	19.5742	18.1757
OCTOBER	22.7468	20.2132	18.3041
NOVEMBER	22.7223	20.0417	19.1529
DECEMBER	21.7008	19.2197	18.8899

	1974
JANUARY	19.1681
FEBRUARY	20.5372
MARCH	20.4055
APRIL	20.6113
MAY	20.5482
JUNE	19.9078
JULY	19.5582
AUGUST	19.4333
SEPTEMBER	20.4234
OCTOBER	NA
NOVEMBER	NA
DECEMBER	NA

DOCUMENTATION FOR: DS.STFL

NAME: Days Supply of Distillate Oil

DEFINITION: Number of days that U.S. could continue to consume distillate oil at the demand rate in period t given a level of stocks in period t, where t = month.

INTERPRETATION: This indicator provides a rough guide to how vulnerable distillate oil stocks are to any exogenous shocks either domestic or foreign. The indicator should be used and applied judiciously because movement in it can be associated with either the numerator or denominator of the ratio.

UNITS: Days

FREQUENCY: Monthly

INPUTS:

DS.DMD--Domestic demand for distillate oil in thousand bbl/day. From MER. Source: BOM through April 1974, FEA from May 1974 forward.

DS.ST-- Primary stocks of distillate oil held by producers in thousand bbl. From MER. Source: BOM through April 1974, FEA from May 1974 thereafter.

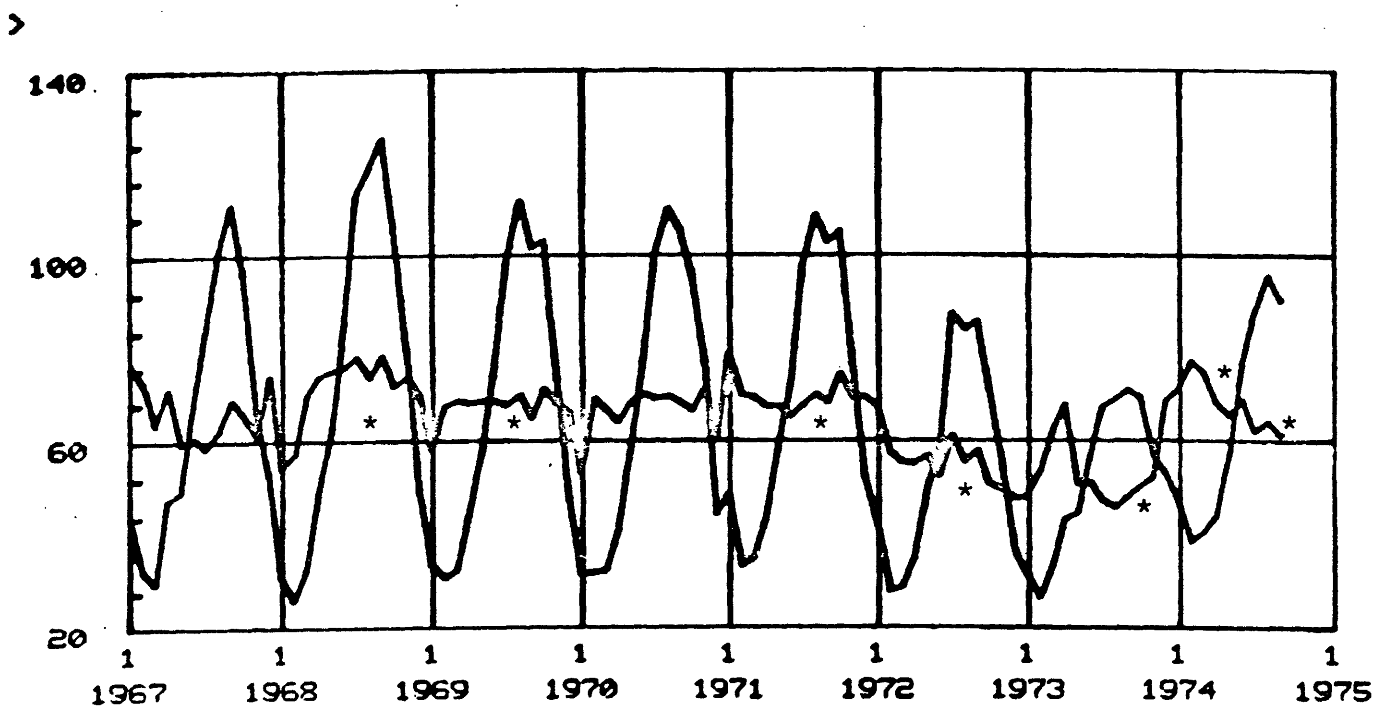
FORMULA:

$$\frac{\text{Stocks of distillate}_t}{\text{Domestic demand for distillate}_t} = \frac{\text{DS.ST}}{\text{DS.DMD}} \frac{\text{(bbl)}}{\text{(bbl/day)}}$$

OUTPUT: DS.STFL is graphed where vertical axis is days and horizontal axis is time. Seasonal adjustment shown. Table of data used to produce graph is also shown.

CONTACT: John Curtis, FEA, Office of Data Policy, (202) 961-7986

Days Supply of Distillate Oil



*indicates seasonally adjusted series

Days Supply of Distillate Oil

TRANSP(DS.STFL)

	1967	1968	1969
JANUARY	43.974	31.101	33.954
FEBRUARY	32.8845	26.2403	31.0119
MARCH	29.9091	33.276	32.8673
APRIL	47.7801	49.4183	44.7183
MAY	49.4176	63.2951	58.5316
JUNE	68.8645	86.3134	77.0693
JULY	84.6361	113.325	98.7586
AUGUST	101.304	119.562	112.027
SEPTEMBER	110.997	125.649	101.899
OCTOBER	96.0193	105.26	103.35
NOVEMBER	64.5284	78.5784	72.7165
DECEMBER	52.5435	49.6187	47.5382

	1970	1971	1972
JANUARY	31.8619	48.9263	42.9841
FEBRUARY	32.2674	33.5561	28.333
MARCH	32.6714	35.2769	29.1841
APRIL	41.2685	43.1256	35.3938
MAY	59.5224	59.3451	50.167
JUNE	78.4673	72.7592	58.696
JULY	100.812	98.245	87.9553
AUGUST	110.288	108.823	84.6424
SEPTEMBER	105.213	103.048	86.2989
OCTOBER	95.9351	105.33	70.8844
NOVEMBER	77.0681	75.4032	53.9814
DECEMBER	44.699	52.0084	36.4648

	1973	1974
JANUARY	31.6867	47.439
FEBRUARY	26.7052	38.8949
MARCH	33.5845	40.9704
APRIL	43.5381	44.0966
MAY	44.9212	57.8536
JUNE	57.1596	77.2072
JULY	67.6339	88.127
AUGUST	69.4221	95.8309
SEPTEMBER	71.5071	90.6133
OCTOBER	69.6159	NA
NOVEMBER	57.0747	NA
DECEMBER	53.3137	NA

DOCUMENTATION FOR GS.STFL

NAME: Days Supply of Gasoline (GS.STFL)

DEFINITION: Number of days that U.S. could continue to consume gasoline at the demand rate in period t given a level of stocks in period t, where t = month.

INTERPRETATION: This indicator provides a rough guide to how vulnerable motor gasoline stocks are to any exogenous shocks either domestic or foreign. The indicator should be used and applied judiciously because movement in it can be associated with either the numerator or denominator of the ratio.

UNITS: Days

FREQUENCY: Monthly

INPUTS: GS.DMD--Domestic demand for motor gasoline in thousand bbl/day. From MER. Source: BOM through April 1974, FEA from May 1974 forward.

GS.ST --Primary stocks of motor gasoline held by producers in thousand bbl. From MER. Source: BOM through April 1974, FEA from May 1974 thereafter.

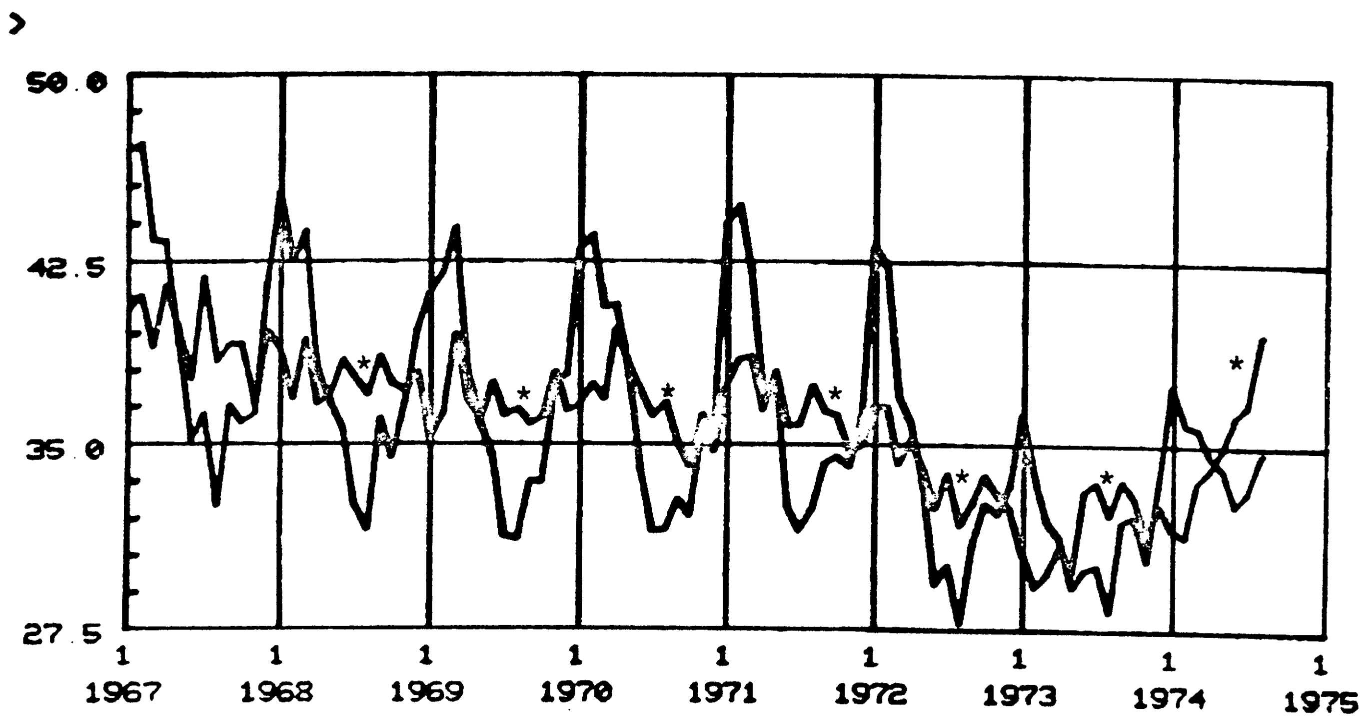
FORMULA;

$$\frac{\text{Stocks of gasoline}_t}{\text{Domestic demand for gasoline}_t} = \frac{\text{GS.ST}}{\text{GS.DMD}} \quad \frac{(\text{bbl})}{(\text{bbl/day})}$$

OUTPUT: GS.STFL is graphed where vertical axis is days and horizontal axis is time. Seasonal adjustment shown. Table of data used to produce graph is also shown.

CONTACT: John Curtis, FEA, Office of Data Policy, (202) 961-7986.

Days Supply of Gasoline



*indicates seasonally adjusted series

Days Supply of Gasoline

TRANSP(GS.STFL)

	1967	1968	1969
JANUARY	46.9909	45.3397	41.2099
FEBRUARY	47.202	42.5161	42.0991
MARCH	43.3876	43.7499	43.896
APRIL	43.3416	38.0777	38.0441
MAY	39.0794	36.6574	35.7616
JUNE	35.1128	35.6452	34.53
JULY	36.2241	32.6068	31.3194
AUGUST	32.5352	31.5484	31.1982
SEPTEMBER	36.5733	36.0658	33.4995
OCTOBER	35.9058	34.4948	33.5201
NOVEMBER	36.3421	36.8003	37.568
DECEMBER	41.6961	39.7217	37.8475

	1970	1971	1972
JANUARY	43.0026	44.0909	43.2369
FEBRUARY	43.6064	44.7853	42.3127
MARCH	40.6642	42.0735	36.9918
APRIL	40.767	37.1635	35.8965
MAY	37.7618	37.5634	33.3735
JUNE	33.9349	32.4881	29.3687
JULY	31.5134	31.5623	30.1184
AUGUST	31.59	32.4359	27.8118
SEPTEMBER	32.7945	34.2542	30.9846
OCTOBER	32.1504	34.5624	32.664
NOVEMBER	35.7168	34.1783	32.2642
DECEMBER	35.9447	36.1353	33.3919

	1973	1974
JANUARY	36.2809	37.4655
FEBRUARY	33.6336	35.9111
MARCH	31.8953	35.7529
APRIL	31.3199	34.653
MAY	29.2737	34.0793
JUNE	29.936	32.6541
JULY	30.125	33.172
AUGUST	28.3038	34.7078
SEPTEMBER	31.9667	NA
OCTOBER	32.1421	NA
NOVEMBER	30.4001	NA
DECEMBER	33.6626	NA

DOCUMENTATION FOR: RS.STFL

NAME: Days Supply of Residual Oil

DEFINITION: Number of days that U.S. could continue to consume residual oil at the demand rate in period t given a level of stocks in period t, where t = month.

INTERPRETATION: This indicator provides a rough guide to how vulnerable residual oil stocks are to any exogenous shocks either domestic or foreign. The indicator should be used and applied judiciously because movement in it can be associated with either the numerator or denominator of the ratio.

UNITS: Days

FREQUENCY: Monthly

INPUTS: RS.DMD--Domestic demand for residual oil in thousand bbl/day. From MER. Source: BOM through April 1974, FEA from May 1974 forward.

RS.ST-- Primary stocks of residual oil held by producers in thousand bbl. From MER. Source: BOM through April 1974, FEA from May 1974 thereafter.

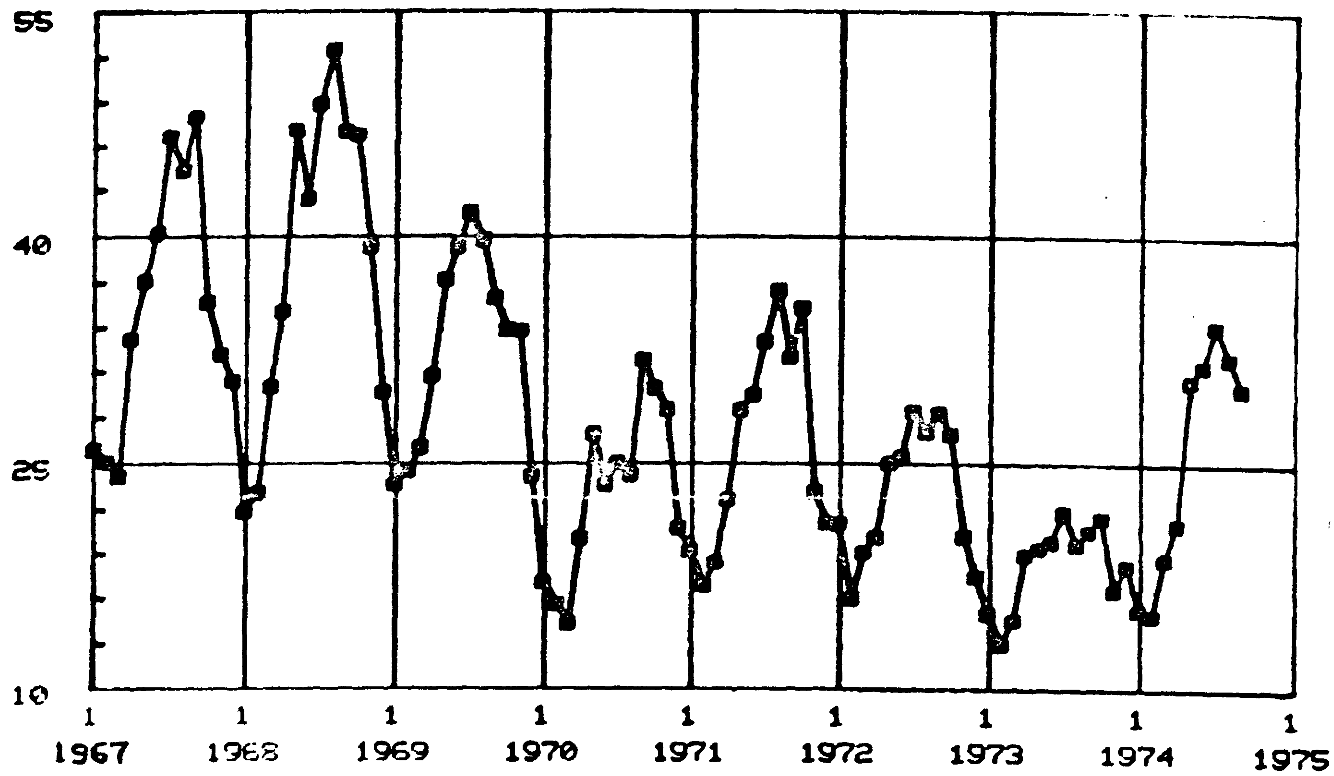
FORMULA:

$$\frac{\text{Stocks of residual}_t}{\text{Domestic demand for residual}_t} = \frac{\text{RS.ST}}{\text{RS.DMD}} \quad \frac{\text{(bbl)}}{\text{(bbl/day)}}$$

OUTPUT: RS.STFL is graphed where vertical axis is days and horizontal axis is time. Seasonal adjustment shown. Table of data used to produce graph is also shown.

CONTACT: John Curtis, FEA, Office of Energy Data Policy, (202) 961-7986

Days Supply of Residual Oil



Days Supply of Residual Oil

	1967	1968	1969
JANUARY	25.9981	21.8809	23.7202
FEBRUARY	25.2321	23.147	24.6332
MARCH	24.2337	30.1272	26.0879
APRIL	33.3781	35.2567	30.8074
MAY	37.2337	47.2884	37.3746
JUNE	40.3663	42.658	39.4989
JULY	46.8139	49.0849	41.721
AUGUST	44.5382	52.7072	39.9143
SEPTEMBER	48.1909	47.2205	36.1102
OCTOBER	35.8196	46.8916	33.9571
NOVEMBER	32.2948	39.4801	33.8522
DECEMBER	30.4499	29.7611	24.2824

	1970	1971	1972
JANUARY	17.1192	19.3169	21.1154
FEBRUARY	15.6982	16.9607	16.0489
MARCH	14.3254	18.5574	19.2267
APRIL	20.1711	22.7155	20.223
MAY	27.0026	28.5979	25.1232
JUNE	23.7213	29.6123	25.5505
JULY	25.0898	33.1464	23.5585
AUGUST	24.3748	36.658	27.1677
SEPTEMBER	31.9544	32.095	28.472
OCTOBER	30.0868	35.5136	27.0161
NOVEMBER	28.5984	23.3019	20.2962
DECEMBER	20.8196	21.1659	17.5289

	1973	1974
JANUARY	15.0687	15.338
FEBRUARY	13.0281	14.9515
MARCH	14.5591	18.767
APRIL	19.0307	21.113
MAY	19.5421	30.577
JUNE	19.912	31.5324
JULY	21.9601	34.2229
AUGUST	19.7443	32.0992
SEPTEMBER	20.6565	30.019
OCTOBER	21.5799	NA
NOVEMBER	16.6725	NA
DECEMBER	18.378	NA

DOCUMENTATION FOR JT.STFL

NAME: Days Supply of Jet Fuel

DEFINITION: Number of days that U.S. could continue to consume jet fuel at the demand rate in period t given a level of stocks in period t, where t = month.

INTERPRETATION: This indicator provides a rough guide to how vulnerable jet fuel oil stocks are to any exogenous shocks either domestic or foreign. The indicator should be used and applied judiciously because movement in it can be associated with either the numerator or denominator of the ratio.

UNITS: Days

FREQUENCY: Monthly

INPUTS: JT.DMD--Domestic demand for residual oil in thousand bbl/day. From MER. Source: BOM through April 1974, FEA from May 1974 forward.

JT.ST-- Primary stocks of residual oil held by producers in thousand bbl. From MER. Source: BOM through April 1974, FEA from May 1974 thereafter.

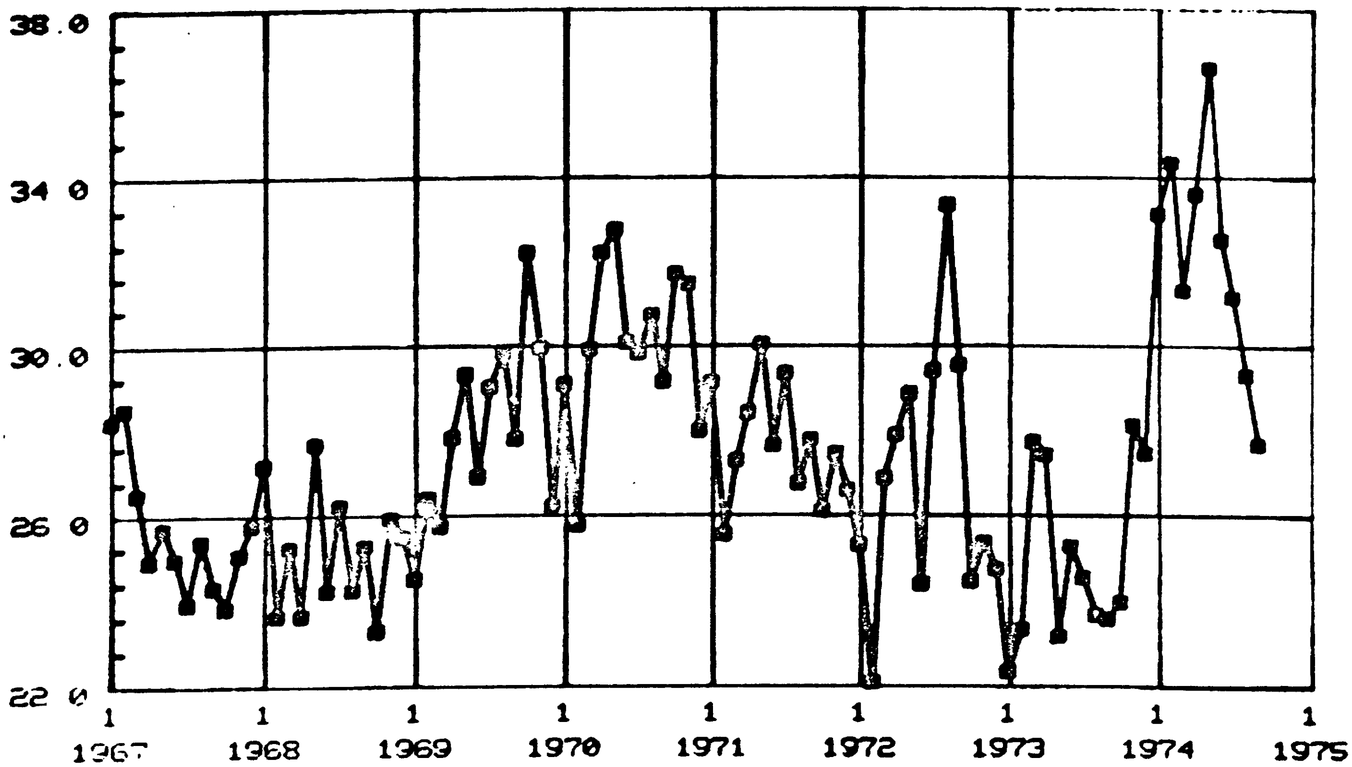
FORMULA:

$$\frac{\text{Stocks of jet fuel}_t}{\text{Domestic demand for jet fuel}_t} = \frac{\text{JT.ST}}{\text{JT.DMD}} \quad \frac{\text{(bbl)}}{\text{(bbl/day}}$$

OUTPUT: JT.STFL is graphed where vertical axis is days and horizontal axis is time. Seasonal adjustment shown. Table of data used to produce graph is also shown.

CONTACT: John Curtis, FEA, Office of Energy Data Policy, (202) 961-8796.

Days Supply of Jet Fuel



Days Supply of Jet Fuel

TRANSP(JT.STFL)

	1967	1968	1969
JANUARY	28.2851	27.2182	24.5213
FEBRUARY	28.5486	23.676	26.6471
MARCH	26.5436	25.2384	25.7717
APRIL	24.9523	23.6824	27.9003
MAY	25.7366	27.7386	29.3401
JUNE	25.0078	24.2418	26.975
JULY	23.9614	26.2762	29.0492
AUGUST	25.4228	24.2647	29.8972
SEPTEMBER	24.3479	25.3001	27.8683
OCTOBER	23.8614	23.3138	32.2862
NOVEMBER	25.0989	25.951	29.9868
DECEMBER	25.7882	25.5149	26.2562

	1970	1971	1972
JANUARY	29.1336	29.1729	25.3252
FEBRUARY	25.8054	25.5812	22.1122
MARCH	29.9411	27.3569	26.9315
APRIL	32.2365	28.4622	27.3594
MAY	32.8379	30.0801	28.9139
JUNE	30.1252	27.6888	24.3818
JULY	29.8575	29.3671	23.429
AUGUST	30.7743	26.823	33.4556
SEPTEMBER	29.1923	27.8186	29.5623
OCTOBER	31.7855	26.1314	24.4517
NOVEMBER	31.54	27.5008	25.381
DECEMBER	28.0341	26.6545	24.7505

	1973	1974
JANUARY	22.3549	33.2201
FEBRUARY	23.3367	34.4384
MARCH	27.7794	31.3766
APRIL	27.469	33.7141
MAY	23.203	36.6929
JUNE	25.2701	32.6063
JULY	24.556	31.2316
AUGUST	23.6902	29.3624
SEPTEMBER	23.6141	27.7411
OCTOBER	23.9934	NA
NOVEMBER	28.1727	NA
DECEMBER	27.499	NA

DOCUMENTATION FOR: PET.STFL

NAME: Days Supply of Petroleum Products

DEFINITION Number of days the U.S. could continue to consume petroleum products at the rate in period t using only the stocks available at that time.

INTERPRETATION: A rough guide to the ability of the overall petroleum supply system to absorb shocks.

UNITS: Days

FREQUENCY: Monthly

FORMULA:

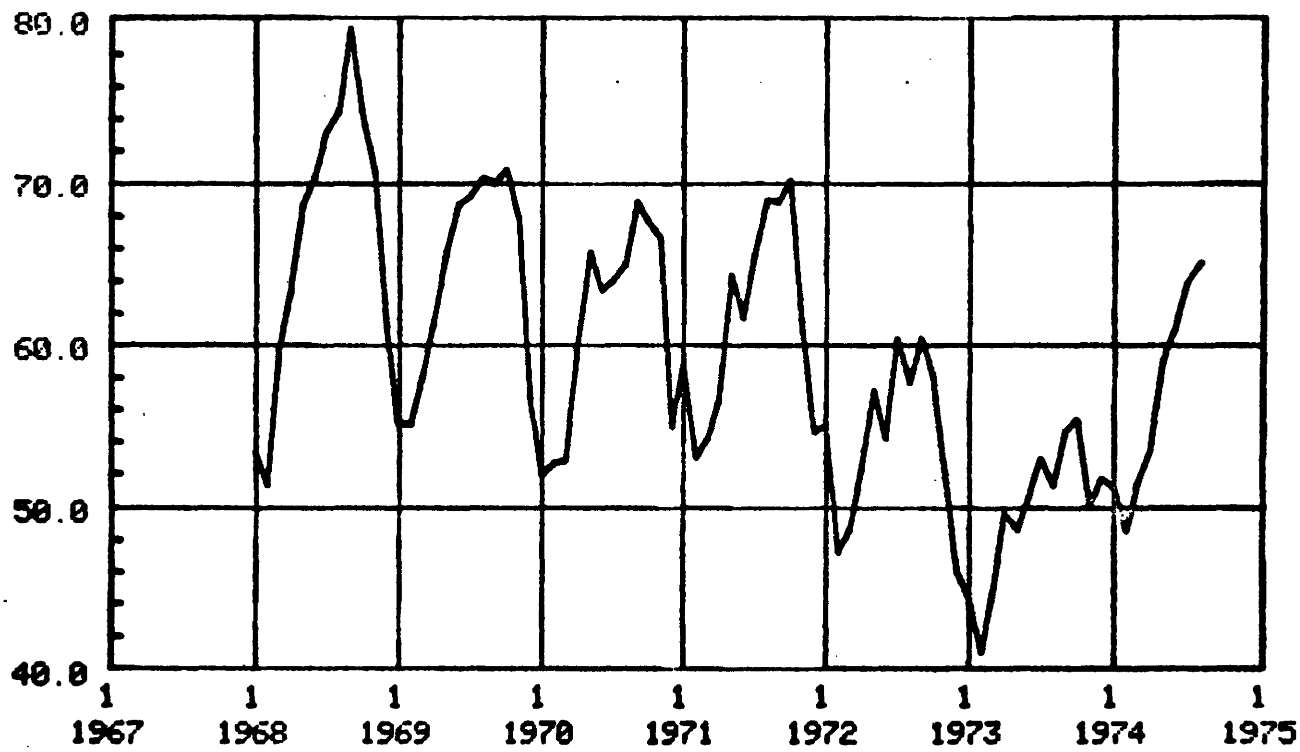
$$\frac{CR.ST + DS.ST + GS.ST + RS.ST + JT.ST}{DS.DMD + GS.DMD + RS.DMD + JT.DMD} \quad \frac{(bbl)}{(bbl/day)}$$

OUTPUT: Vertical axis is days supply; horizontal axis is time.

CONTACT: John Curtis, FEA, Office of Energy Data Policy, (202) 961-8796.

COMMENTS: Fuels are volume wieghted. BTU weighting would yield similar results.

Days Supply of Petroleum Products



Days Supply of Petroleum Products

TRANSP(PET.STFL)

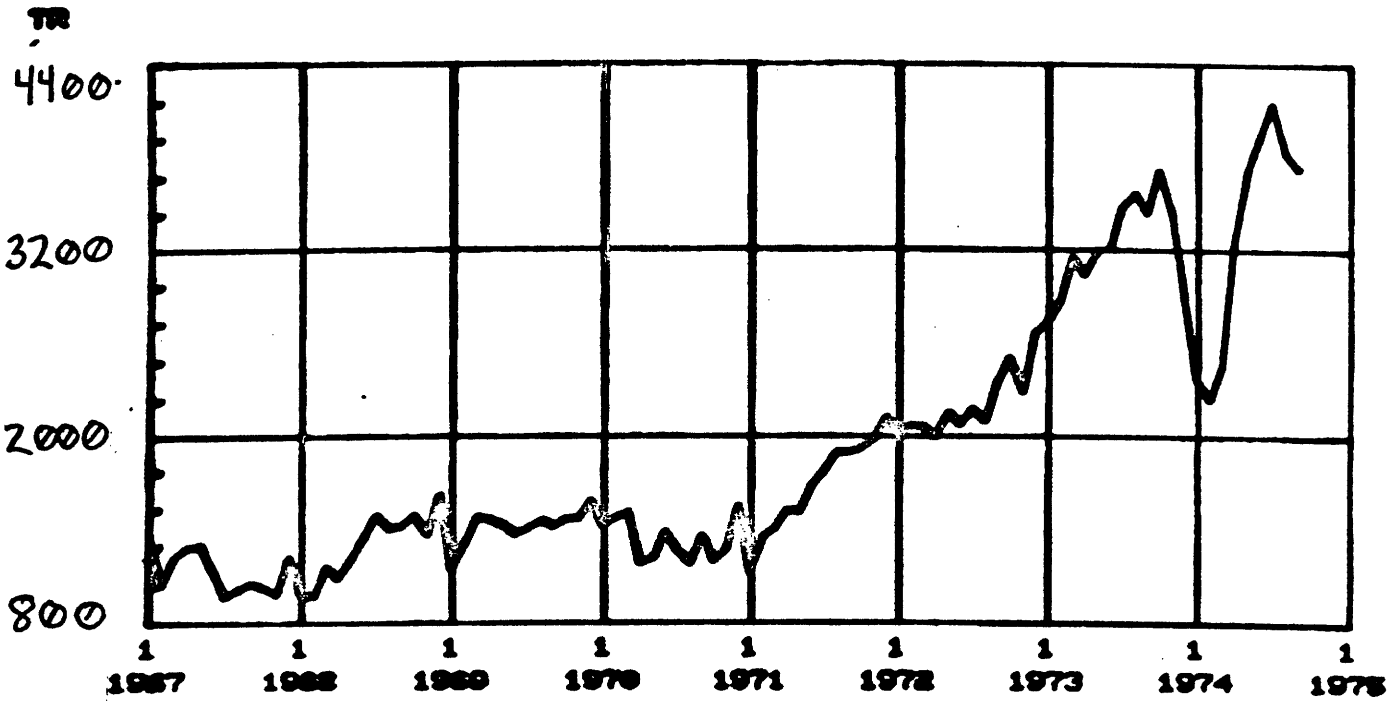
	1964	1965	1966	1967	1968	1969	1970
JANUARY	53.457	55.1924	52.0031	58.8486	55.082	44.515	51.2921
FEBRUARY	51.3613	55.1871	52.7563	53.1019	47.2828	41.141	48.5943
MARCH	59.649	58.1666	52.911	54.2283	48.7552	44.7791	51.4965
APRIL	63.3853	61.4995	59.9858	56.5607	52.3849	49.7423	53.3297
MAY	68.635	65.9349	65.8638	64.3848	57.1951	48.6071	59.0613
JUNE	70.3536	68.708	63.4781	61.8057	54.2332	50.7857	69.9766
JULY	73.2146	69.2325	64.1149	65.6748	60.4406	52.9813	63.8514
AUGUST	74.4802	70.4119	65.0697	68.9875	57.7516	51.3463	64.9631
SEPTEMBER	73.5672	70.039	68.8504	68.8635	60.488	54.6606	NA
OCTOBER	74.3815	70.8622	67.6225	70.2105	58.1746	55.4048	NA
NOVEMBER	70.7248	67.8905	66.6709	69.8951	51.8745	50.1427	NA
DECEMBER	61.3166	56.617	54.9688	54.6816	46.0311	51.7807	NA

DOCUMENTATION FOR: CR.IMD

NAME:	Average Daily Imports of Crude Oil
DEFINITION:	The average daily volume each month of crude oil imported which is reported at receiving refineries, including crude oil entering the U.S through pipelines from Canada.
INTERPRETATION:	This indicator shows the absolute volume of imports of crude oil. It is a basis for a variety of indicators of domestic dependence on foreign crude oil.
UNITS:	Thousand bbl/day
FREQUENCY:	Monthly
FORMULA:	$\frac{\text{Monthly imports of crude oil}}{\text{\# of calendar days/month}}$
OUTPUT:	CR.IMD is graphed where vertical axis is thousand bbl/day and horizontal axis is time. Table of data used to produce graph is also shown.
SOURCE:	Directly from MER. Source BOM through April 1974; FEA from May 1974 forward.
CONTACT:	John Curtis, FEA, Office of Energy Data Policy, (202) 961-8796.

Average Daily Imports of Crude Oil

10^3 bbl/day



Average Daily Imports of Crude Oil

(10^3 bbl/day)

TRANSP(CR.IMD)

	1967	1968	1969
JANUARY	1326.03	985.064	1143.29
FEBRUARY	1043.57	970.8	1304.89
MARCH	1212.42	1145.35	1472.71
APRIL	1273.97	1081.97	1452.27
MAY	1286.45	1208.45	1423.77
JUNE	1121.33	1340.4	1365.33
JULY	970.709	1474.74	1393.84
AUGUST	1014.77	1394.94	1447.97
SEPTEMBER	1048.6	1415.8	1412.13
OCTOBER	1028.71	1481.03	1452.97
NOVEMBER	987.4	1359.3	1459.17
DECEMBER	1203.97	1608.71	1565.77

	1970	1971	1972
JANUARY	1412.74	1121.68	2045.77
FEBRUARY	1468.36	1344.82	2081.
MARCH	1495.16	1394.68	2066.65
APRIL	1171.4	1506.9	2004.3
MAY	1208.65	1501.29	2159.94
JUNE	1372.6	1677.9	2084.8
JULY	1259.87	1766.71	2181.77
AUGUST	1177.9	1895.32	2111.71
SEPTEMBER	1343.67	1900.4	2363.63
OCTOBER	1189.52	1923.13	2516.23
NOVEMBER	1260.5	1985.23	2299.27
DECEMBER	1537.9	2127.81	2667.32

	1973	1974
JANUARY	2732.03	2381.9
FEBRUARY	2872.61	2247.86
MARCH	3161.97	2462.23
APRIL	3048.63	3267.03
MAY	3214.65	3748.
JUNE	3220.43	3957.
JULY	3500.97	4167.
AUGUST	3592.52	3852.
SEPTEMBER	3470.57	3758.
OCTOBER	3738.87	NA
NOVEMBER	3452.33	NA
DECEMBER	2891.39	NA

DOCUMENTATION FOR: DS.IMD

NAME: Average Daily Imports of Distillate Fuel Oil

DEFINITION: The average daily volume each month of distillate fuel oil imported into the U.S.

INTERPRETATION: This indicator shows the absolute volume of imports of distillate fuel oil. It is a basis for a variety of indicators of domestic dependence on foreign distillate fuel oil.

UNITS: thousand bbl/day

FREQUENCY: Monthly

FORMULA:

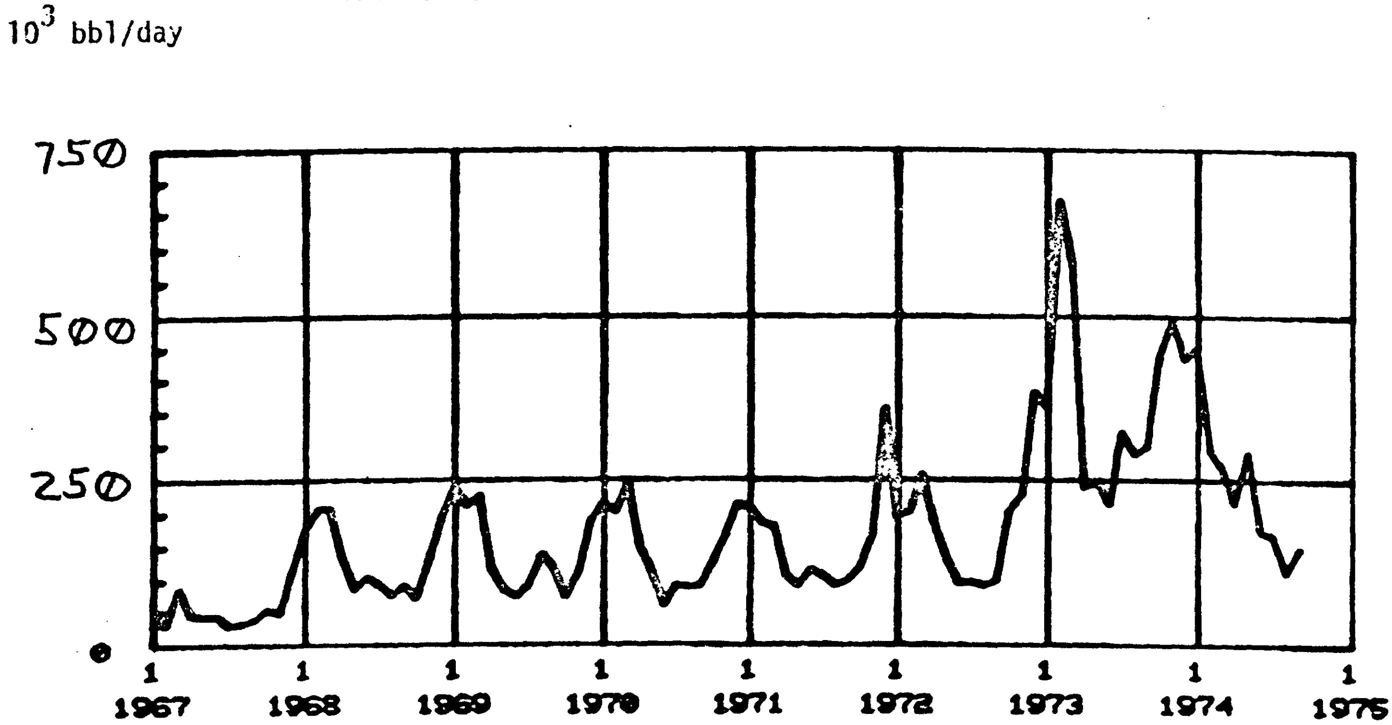
$$\frac{\text{Monthly imports of distillate fuel oil}}{\text{\# of calendar days/month}}$$

OUTPUT: DS.IMD is graphed where vertical axis is thousand bbl/day and horizontal axis is time. Table of data used to produce graph is also shown.

SOURCE: Directly from MER. Source BOM from Department of Commerce data, except for imports of bonded bunkers, distillate fuel oil for onshore military use, and receipts from Puerto Rico, Guam, and the Virgin Islands based on data reported to the Oil Import Administration of the FEA.

CONTACT: John Curtis, FEA, Office of Energy Data Policy, (202) 961-8796.

Average Daily Imports of Distillate Fuel Oil



Average Daily Imports of Distillate Fuel Oil

(10^3 bbl/day)

TRANSP(DS.IMD)

	1967	1968	1969
JANUARY	37.0323	172.645	236.806
FEBRUARY	31.9643	208.	213.321
MARCH	36.9677	207.871	226.871
APRIL	45.9333	137.133	117.433
MAY	42.	85.3226	84.2903
JUNE	44.2333	104.533	74.0333
JULY	28.8064	94.5161	91.1613
AUGUST	34.	76.8387	138.581
SEPTEMBER	38.5	91.1667	115.733
OCTOBER	54.2258	72.5484	74.0968
NOVEMBER	47.8333	134.933	113.433
DECEMBER	113.806	196.129	189.839

	1970	1971	1972
JANUARY	215.677	209.161	196.968
FEBRUARY	203.821	185.714	204.
MARCH	244.548	180.161	257.129
APRIL	151.667	107.9	188.733
MAY	109.161	92.7097	131.806
JUNE	62.4	116.967	96.1
JULY	91.6129	107.323	97.3548
AUGUST	89.0968	91.6452	92.3226
SEPTEMBER	92.5667	98.7667	98.7667
OCTOBER	128.581	118.677	203.194
NOVEMBER	167.267	169.767	227.333
DECEMBER	214.935	354.387	382.226

	1973	1974
JANUARY	359.806	449.
FEBRUARY	672.036	293.
MARCH	579.129	267.
APRIL	240.367	216.
MAY	247.29	288.
JUNE	215.367	175.
JULY	318.709	168.
AUGUST	286.323	112.
SEPTEMBER	298.167	146.
OCTOBER	436.484	NA
NOVEMBER	493.133	NA
DECEMBER	434.323	NA

DOCUMENTATION FOR: GS.IMD

NAME: Average Daily Imports of Motor Gasoline

DEFINITION: The average daily volume each month of Motor Gasoline imported into the U.S.

INTERPRETATION: This indicator shows the absolute volume of imports of Motor Gasoline. It is a basis for a variety of indicators of domestic dependence on foreign Motor Gasoline.

UNITS: Thousand bbl/day

FREQUENCY: Monthly

FORMULA:

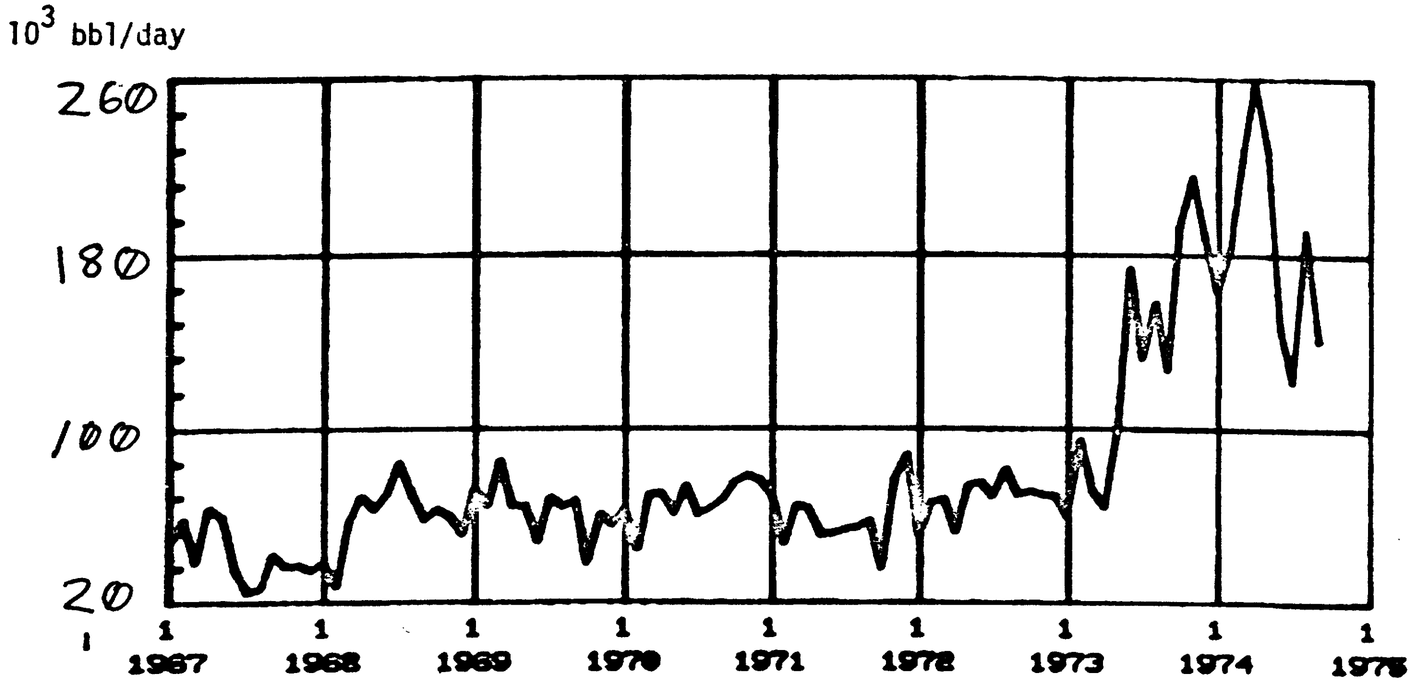
$$\frac{\text{Monthly imports of Motor Gasoline}}{\text{\# of calendar days/month}}$$

OUTPUT: GS.IMD is graphed where vertical axis is thousand bbl/day and horizontal axis is time. Table of data used to produce graph is also shown.

SOURCE: Directly from MER. Source BOM from Department of Commerce data, except for imports of bonded bunkers, and receipts from Puerto Rico, Guam, and the Virgin Islands based on data reported to the Oil Import Administration of the FEA.

CONTACT: John Curtis, FEA, Office of Energy Data Policy, (202) 961-8796.

Average Daily Imports of Motor Gasoline



Average Daily Imports of Motor Gasoline

TRANSP(GS.IMD)

(10^3 bbl/day)

	1967	1968	1969
JANUARY	47.4339	38.129	71.3871
FEBRUARY	57.9643	28.	64.0357
MARCH	38.7742	56.8064	85.0968
APRIL	63.0333	68.2	64.3
MAY	58.6452	62.4193	63.2903
JUNE	35.4333	69.4	47.8
JULY	24.871	84.1935	67.2903
AUGUST	26.5484	68.7097	63.6774
SEPTEMBER	41.4667	57.9667	66.2
OCTOBER	36.4193	62.2903	38.1935
NOVEMBER	36.7	58.6667	59.8
DECEMBER	34.7742	51.5806	55.3548

	1970	1971	1972
JANUARY	62.4516	69.6129	50.7742
FEBRUARY	44.2857	46.75	66.
MARCH	69.2258	64.2258	66.9677
APRIL	69.5333	63.1	52.3
MAY	60.4193	50.5161	73.7742
JUNE	73.3333	51.3667	74.8
JULY	60.0645	53.2903	68.9032
AUGUST	62.871	54.0645	81.0323
SEPTEMBER	66.9333	57.2667	69.4667
OCTOBER	74.7419	35.9677	70.8064
NOVEMBER	78.3	76.8	69.3333
DECEMBER	75.9355	88.2258	68.6129

	1973	1974
JANUARY	59.3871	162.806
FEBRUARY	95.25	184.
MARCH	70.7419	225.
APRIL	63.4	260.
MAY	101.484	228.
JUNE	173.8	145.
JULY	132.581	122.
AUGUST	157.129	192.
SEPTEMBER	127.2	140.
OCTOBER	194.194	NA
NOVEMBER	216.4	NA
DECEMBER	188.194	NA

DOCUMENTATION FOR: RS.IMD

NAME: Average Daily Imports of Residual Fuel Oil

DEFINITION: The average daily volume each month of residual fuel oil imported into the U.S.

INTERPRETATION: This indicator shows the absolute volume of imports of residual fuel oil. It is a basis for a variety of indicators of domestic dependence on foreign residual oils.

UNITS: Thousand bbl/day

FREQUENCY: Monthly

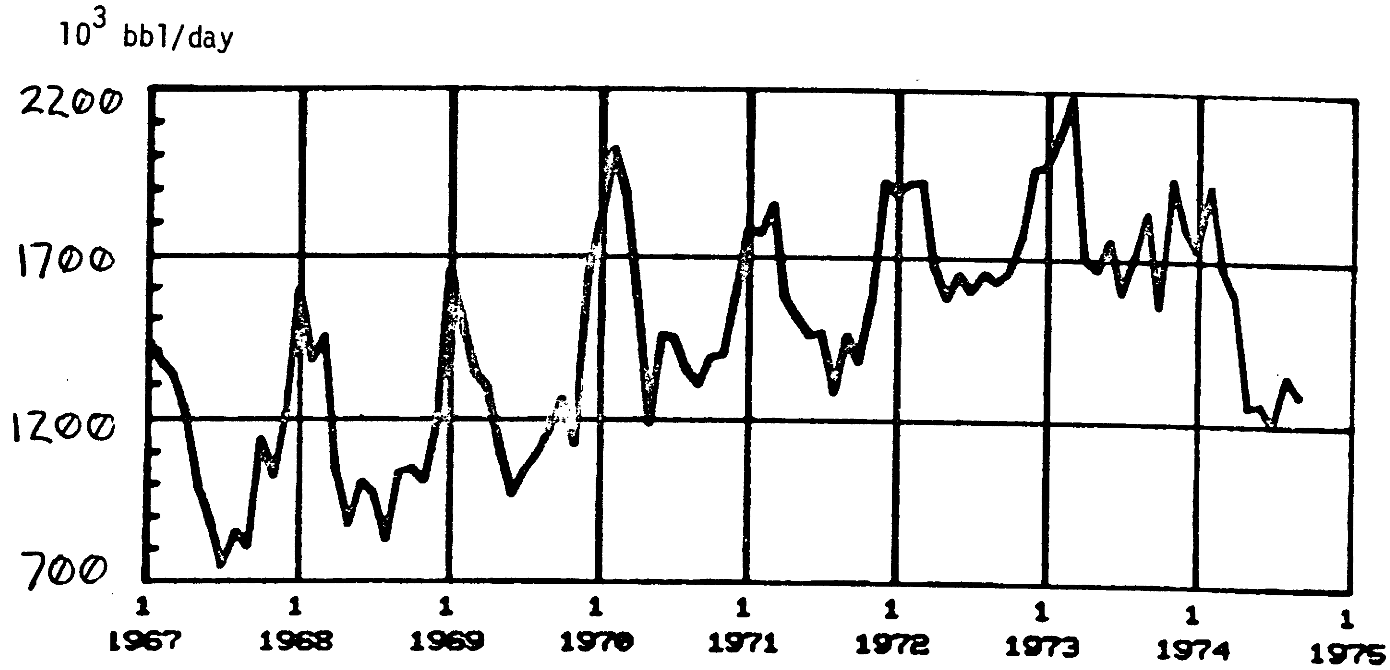
FORMULA: $\frac{\text{Monthly imports of residual fuel oil}}{\text{\# of calendar days/month}}$

OUTPUT: RS.IMD is graphed where vertical axis is thousand bbl/day and horizontal axis is time. Table of data used to produce graph is also shown.

SOURCE: Directly from MER. Source BOM from Department of Commerce data, except for imports of bonded bunkers, residual fuel oil for onshore military use, and receipts from Puerto Rico, Guam, and the Virgin Islands based on data reported to the Oil Import Administration of the FEA.

CONTACT: John Curtis, FEA, Office of Energy Data Policy, (202) 961-8796.

Average Daily Imports to Residual Fuel Oil



Average Daily Imports of Residual Fuel Oil

(10^3 bbl/day)

TRANSP(RS.IMD)

	1967	1968	1969
JANUARY	1430.32	1595.48	1657.55
FEBRUARY	1366.93	1371.	1472.5
MARCH	1326.74	1445.03	1334.65
APRIL	1213.07	1046.	1294.47
MAY	994.806	876.774	1103.87
JUNE	886.233	1010.37	969.8
JULY	746.064	974.774	1041.58
AUGUST	853.935	830.806	1100.16
SEPTEMBER	806.933	1037.63	1170.
OCTOBER	1142.13	1049.77	1258.1
NOVEMBER	1029.5	1013.77	1125.03
DECEMBER	1224.	1192.81	1652.55

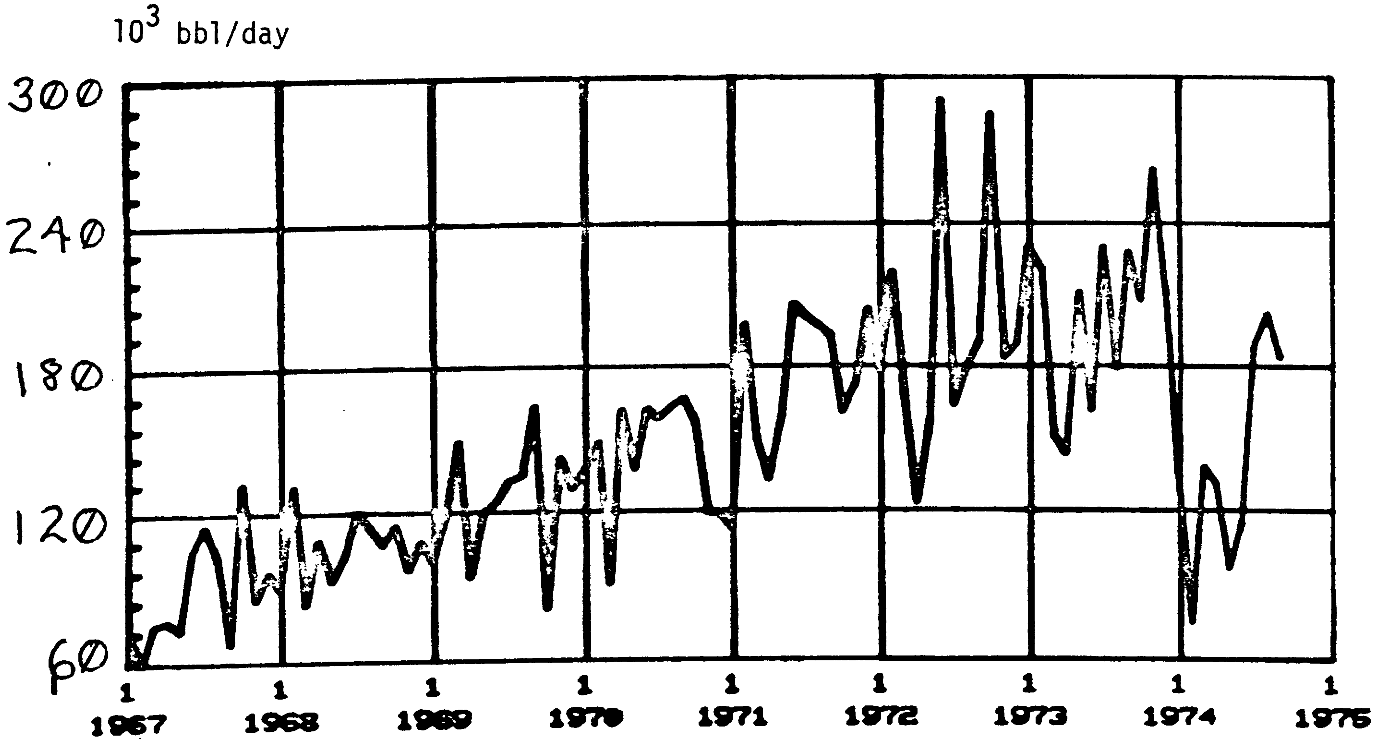
	1970	1971	1972
JANUARY	1306.39	1780.42	1892.19
FEBRUARY	2018.89	1772.68	1923.
MARCH	1886.71	1858.	1926.39
APRIL	1576.8	1574.3	1675.5
MAY	1187.9	1504.32	1573.23
JUNE	1452.7	1451.33	1648.5
JULY	1441.48	1459.06	1594.06
AUGUST	1343.64	1281.97	1653.03
SEPTEMBER	1303.97	1451.5	1624.53
OCTOBER	1384.74	1373.58	1654.94
NOVEMBER	1392.57	1570.	1769.17
DECEMBER	1579.1	1924.94	1967.74

	1973	1974
JANUARY	1977.1	1732.
FEBRUARY	2072.32	1923.
MARCH	2185.23	1674.
APRIL	1702.97	1587.
MAY	1666.35	1250.
JUNE	1757.2	1260.
JULY	1597.26	1197.
AUGUST	1720.84	1342.
SEPTEMBER	1841.6	1287.
OCTOBER	1555.97	NA
NOVEMBER	1941.6	NA
DECEMBER	1793.39	NA

DOCUMENTATION FOR: JT.IMD

NAME:	Average Daily Imports of Jet Fuel
DEFINITION:	The average daily volume each month of jet fuel imported into the U.S.
INTERPRETATION:	This indicator shows the absolute volume of imports of jet fuel. It is a basis for a variety of indicators of domestic dependence on foreign jet fuels.
UNITS:	Thousand bbl/day
FREQUENCY:	Monthly
FORMULA:	$\frac{\text{Monthly imports of jet fuel}}{\text{\# of calendar days/months}}$
OUTPUT:	JT.IMD is graphed where vertical axis is thousand bbl/day and horizontal axis is time. Table of data used to produce graph is also shown.
SOURCE:	Directly from MER. Source BOM from Department of Commerce data, except for imports of bonded bunkers, jet fuel for onshore military use, and receipts from Puerto Rico, Guam, and the Virgin Islands based on data reported to the Oil Import Administration of the FEA.
CONTACT:	John Curtis, FEA, Office of Energy Data Policy, (202) 961-8796.

Average Daily Imports of Jet Fuel



Average Daily Imports of Jet Fuel

TRANSP(JT.IMD) (10^3 bbl/day)

	1967	1968	1969
JANUARY	71.6129	89.6129	100.677
FEBRUARY	60.3214	131.071	119.679
MARCH	75.6129	83.4193	150.032
APRIL	76.8333	108.767	93.6667
MAY	73.2903	92.5484	120.548
JUNE	104.7	102.333	124.567
JULY	114.806	120.645	133.161
AUGUST	101.806	113.645	136.097
SEPTEMBER	68.2667	107.1	164.567
OCTOBER	133.032	114.516	81.3548
NOVEMBER	85.2333	96.7667	143.233
DECEMBER	96.	107.806	129.968

	1970	1971	1972
JANUARY	138.226	113.29	179.
FEBRUARY	149.464	197.286	220.
MARCH	91.	150.516	167.
APRIL	163.	134.067	124.
MAY	138.71	161.71	159.
JUNE	163.4	205.967	292.
JULY	160.161	201.387	165.
AUGUST	164.806	196.806	181.
SEPTEMBER	167.9	192.7	190.
OCTOBER	158.548	162.677	286.
NOVEMBER	119.6	174.067	184.
DECEMBER	119.323	203.484	189.

	1973	1974
JANUARY	231.	136.
FEBRUARY	221.	75.
MARCH	152.	139.
APRIL	145.	132.
MAY	211.	97.
JUNE	163.	115.
JULY	231.	188.
AUGUST	180.	202.
SEPTEMBER	229.	184.
OCTOBER	208.	NA
NOVEMBER	263.	NA
DECEMBER	210.	NA

DOCUMENTATION FOR: CR.FI

NAME: Fraction of Crude Oil Imported

DEFINITION: Fraction of monthly crude runs to stills supplied by imported crude

INTERPRETATION: Indicates the level of dependence of the domestic refinery sector on imported crude oil

UNITS: Dimensionless

FREQUENCY: Monthly

INPUTS: CR.IMD - Total U.S. imports of crude oil in bbl/day

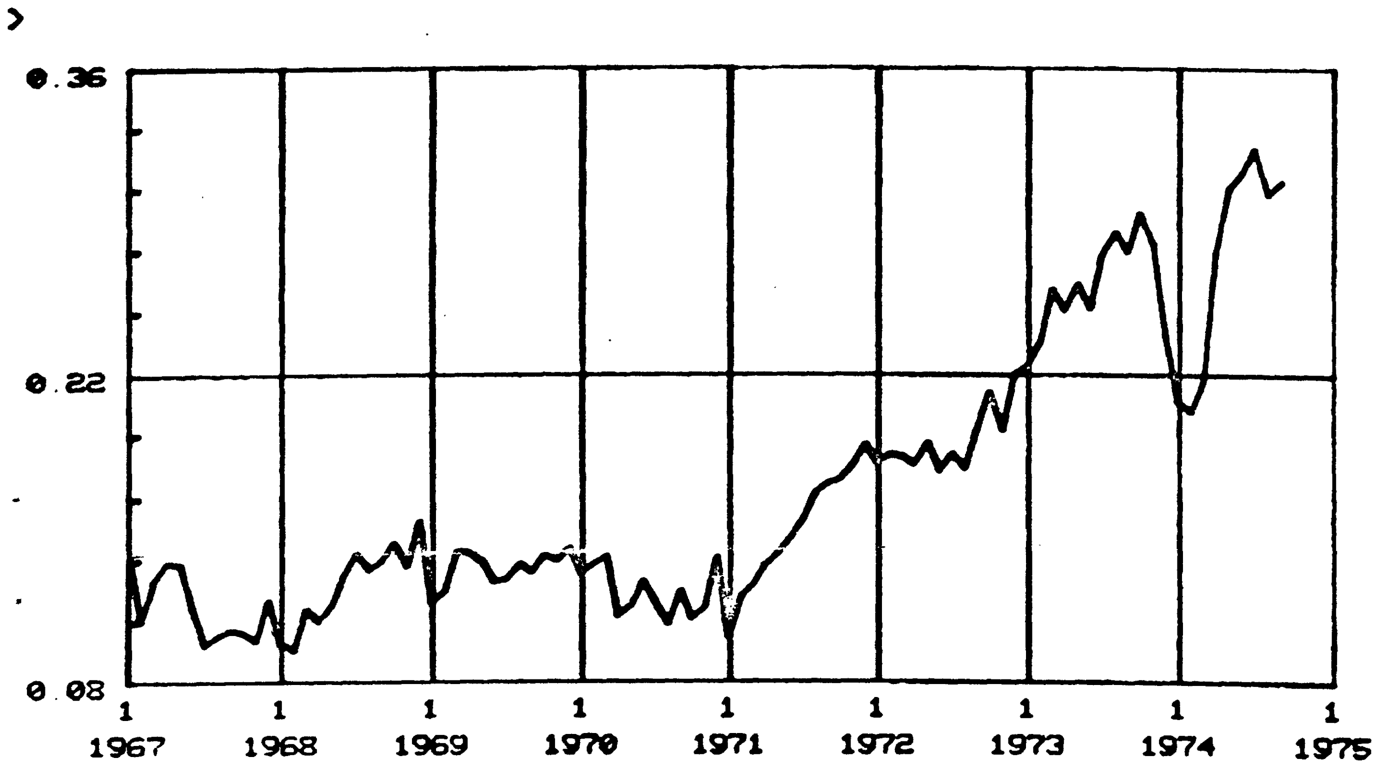
FORMULA:

$$\frac{\text{Av. Daily Imports of Crude Oil}}{\text{Av. Daily Crude Runs to Stills}}$$

OUTPUT: Graph, with fraction of crude runs imported on vertical axis; horizontal axis is time. Graphed monthly, with time axis labeled by year

CONTACT: John Curtis, FEA, Office of Energy Data Policy, (202) 961-8796.

Fraction of Crude Oil Imported



TRANSP(CR.FI)

	1967	1968	1969
JANUARY	0.139621	0.097602	0.116646
FEBRUARY	0.108377	0.094782	0.122069
MARCH	0.12695	0.113521	0.140164
APRIL	0.135114	0.108374	0.139606
MAY	0.134234	0.115595	0.135324
JUNE	0.114196	0.129645	0.126162
JULY	0.097062	0.139328	0.127397
AUGUST	0.101562	0.131654	0.134043
SEPTEMBER	0.10418	0.135973	0.130607
OCTOBER	0.102561	0.143705	0.138261
NOVEMBER	0.099044	0.133806	0.136105
DECEMBER	0.117818	0.153577	0.14171

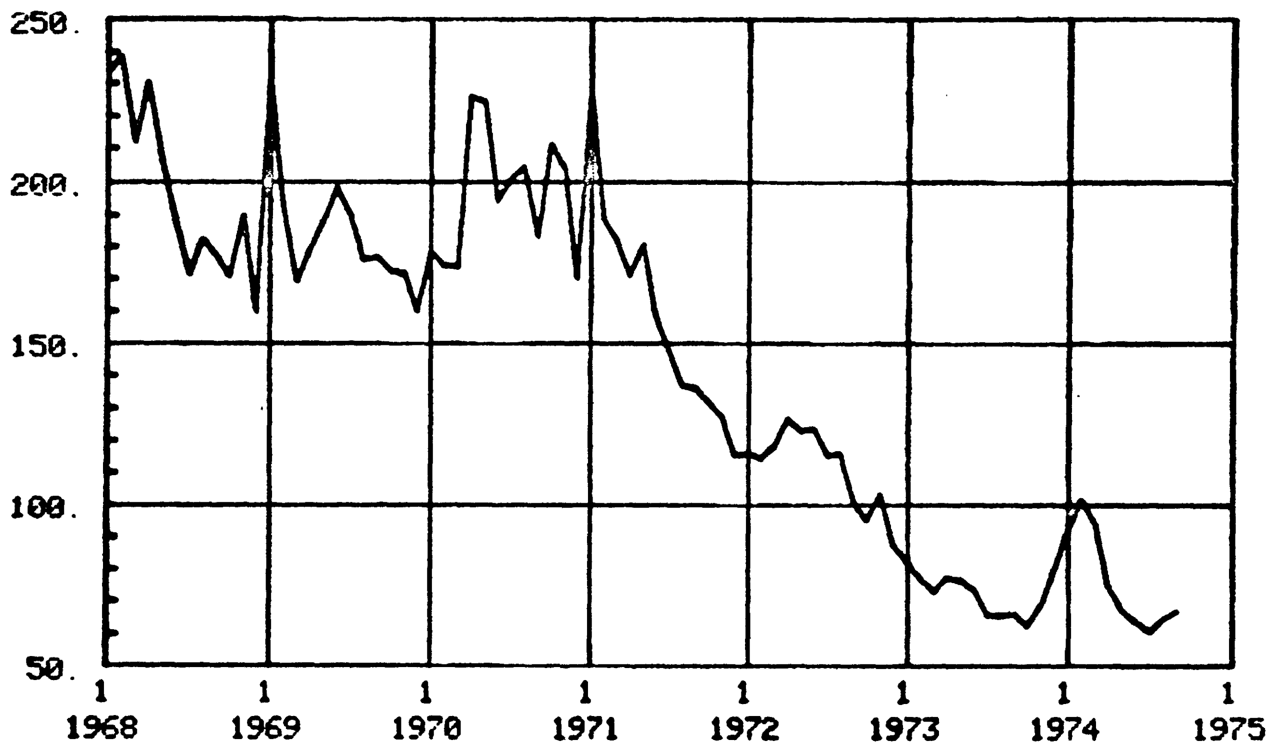
	1970	1971	1972
JANUARY	0.129986	0.100807	0.179639
FEBRUARY	0.134897	0.120586	0.183219
MARCH	0.137998	0.125291	0.182159
APRIL	0.11038	0.134456	0.179204
MAY	0.11565	0.139844	0.188187
JUNE	0.126249	0.146114	0.176063
JULY	0.116418	0.154273	0.183569
AUGUST	0.106362	0.166726	0.177225
SEPTEMBER	0.122043	0.170677	0.195146
OCTOBER	0.109538	0.172545	0.211965
NOVEMBER	0.114373	0.178549	0.19401
DECEMBER	0.137524	0.187669	0.220208

	1973	1974
JANUARY	0.224113	0.207284
FEBRUARY	0.235705	0.202473
MARCH	0.259163	0.216841
APRIL	0.249725	0.276329
MAY	0.261764	0.305286
JUNE	0.250378	0.311354
JULY	0.274595	0.322898
AUGUST	0.28432	0.302568
SEPTEMBER	0.276327	0.306725
OCTOBER	0.293066	NA
NOVEMBER	0.278996	NA
DECEMBER	0.237982	NA

DOCUMENTATION FOR: CR.STIM

NAME:	Days Supply of Crude Imports
DEFINITION:	Number of days the U.S. could continue to consume crude oil at the demand rate in period t if crude imports were stopped.
INTERPRETATION:	A rough guide to the capacity of the fuel sector to absorb shocks in the international supply system.
UNITS:	Days
FREQUENCY:	Monthly
FORMULA:	$\frac{CR.ST}{CR.IMD}$ $\frac{(bbl)}{(bbl/day)}$
OUTPUT:	Vertical axis in days supply, horizontal axis in time
CONTACT:	John Curtis, FEA, Office of Energy Data Policy, (202) 961-8796.

Days Supply of Crude Imports



Days Supply of Crude Imports

TRANSP(CR.STIM)

	1968	1969	1970
JANUARY	234.27	231.382	178.43
FEBRUARY	238.77	192.908	174.374
MARCH	212.005	169.18	173.792
APRIL	230.257	178.75	225.74
MAY	205.983	188.028	224.104
JUNE	188.353	197.934	194.227
JULY	171.613	189.404	200.862
AUGUST	181.908	175.712	204.195
SEPTEMBER	176.969	176.648	183.259
OCTOBER	170.874	172.404	211.566
NOVEMBER	189.644	171.656	203.917
DECEMBER	160.056	159.931	170.555

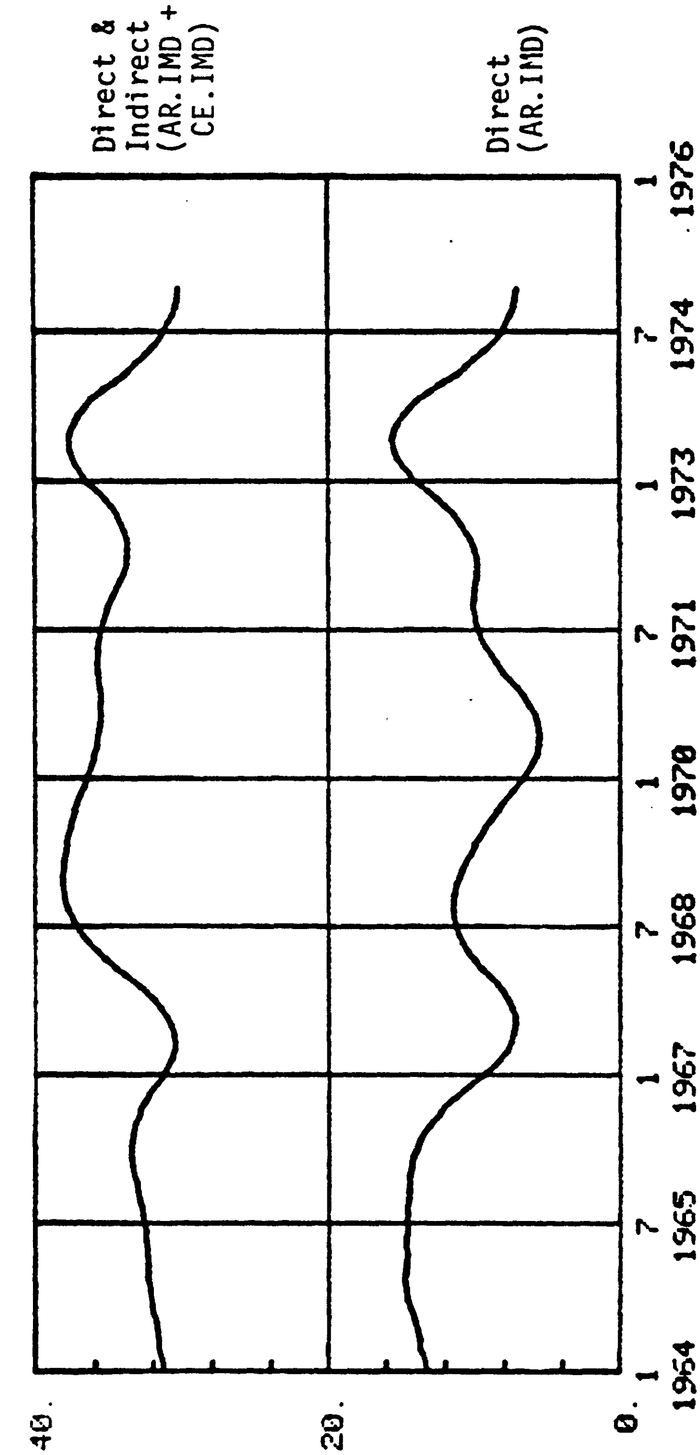
	1971	1972	1973
JANUARY	227.877	115.739	82.0107
FEBRUARY	188.696	114.792	77.2445
MARCH	181.567	118.482	72.9596
APRIL	171.194	126.474	77.2094
MAY	180.377	122.83	76.1443
JUNE	158.818	123.561	73.2342
JULY	147.435	115.462	65.9103
AUGUST	137.026	115.704	65.5975
SEPTEMBER	135.752	100.305	65.776
OCTOBER	131.832	95.3607	62.4574
NOVEMBER	127.261	103.302	68.6495
DECEMBER	115.633	87.2796	79.375

	1974
JANUARY	92.4727
FEBRUARY	101.432
MARCH	94.1039
APRIL	74.5897
MAY	67.3079
JUNE	63.9393
JULY	60.5709
AUGUST	64.2279
SEPTEMBER	66.5851
OCTOBER	NA
NOVEMBER	NA
DECEMBER	NA

DOCUMENTATION FOR: AR.IMD, CE.IMD

NAME:	Percent of Imports from Insecure Sources
DEFINITION:	AR.IMD - Percent of U.S. imports originating in Arab Persian Gulf countries.
	CE.IMD - Percent of U.S. imports originating in European or Caribbean sources (almost all are product imports).
FREQUENCY:	Annual. For calculations, a monthly series is needed. The corrected series is shown below.
SOURCE:	Mr. Gil Rodgers, FEA, Office of Energy Data Policy, (202) 961-8624.

Percent of Imports from Insecure Sources



Percent of Imports From Insecure Sources

TRANSP(AR.IMD)

	1964	1965	1966
JANUARY	13.218	14.7811	14.5381
FEBRUARY	13.3674	14.7407	14.4458
MARCH	13.5107	14.7254	14.293
APRIL	13.7544	14.7291	14.0797
MAY	13.8154	14.7288	13.8058
JUNE	13.9375	14.7184	13.4713
JULY	14.1115	14.6994	13.076
AUGUST	14.3013	14.6783	12.6202
SEPTEMBER	14.4994	14.6567	12.1038
OCTOBER	14.6879	14.6338	11.5267
NOVEMBER	14.8027	14.6084	10.889
DECEMBER	14.8274	14.5797	10.1907

	1967	1968	1969
JANUARY	9.46065	9.1707	11.0608
FEBRUARY	8.81516	9.70441	10.819
MARCH	8.2331	10.1696	10.5553
APRIL	7.36447	10.5662	10.2693
MAY	7.55928	10.8944	9.9624
JUNE	7.36757	11.1541	9.63354
JULY	7.28906	11.3452	9.28125
AUGUST	7.32417	11.4678	8.90918
SEPTEMBER	7.47276	11.522	8.5144
OCTOBER	7.73465	11.5076	8.09619
NOVEMBER	8.11011	11.4246	7.65894
DECEMBER	8.59891	11.2732	7.19897

Percent of Imports From Insecure Sources
(continued)

TRANSP (AR.IMD)

	1970	1971	1972
JANUARY	6.73315	7.65967	9.91699
FEBRUARY	6.33765	8.15771	9.86426
MARCH	6.0293	8.6001	9.88257
APRIL	5.80444	8.98315	9.96973
MAY	5.66895	9.31128	10.1316
JUNE	5.61841	9.58374	10.363
JULY	5.65356	9.79687	10.6648
AUGUST	5.77661	9.95508	11.0398
SEPTEMBER	5.98535	10.0576	11.4851
OCTOBER	6.27979	10.1008	12.0007
NOVEMBER	6.66211	10.0906	12.5903
DECEMBER	7.13013	10.0225	13.2495

	1973	1974
JANUARY	13.9446	12.0663
FEBRUARY	14.5393	11.2676
MARCH	14.9971	10.511
APRIL	15.3179	9.8231
MAY	15.5024	9.22559
JUNE	15.5508	8.69604
JULY	15.4614	8.24048
AUGUST	15.2366	7.86401
SEPTEMBER	14.8762	7.56152
OCTOBER	14.3767	7.33374
NOVEMBER	13.7424	7.18213
DECEMBER	12.9727	7.10742

Percent of Imports From Insecure Sources

TRANSP (AR.IMD + CE.IMD)

	1964	1965	1966
JANUARY	31.2269	32.3256	33.323
FEBRUARY	31.3237	32.3503	33.4265
MARCH	31.4205	32.3792	33.4796
APRIL	31.5732	32.4139	33.4823
MAY	31.6142	32.4621	33.4351
JUNE	31.6963	32.5254	33.3374
JULY	31.812	32.603	33.1892
AUGUST	31.9328	32.6935	32.991
SEPTEMBER	32.0514	32.7964	32.7426
OCTOBER	32.1625	32.9112	32.4437
NOVEMBER	32.2469	33.0392	32.0946
DECEMBER	32.2993	33.1799	31.6952

	1967	1968	1969
JANUARY	31.2708	33.9046	38.1868
FEBRUARY	30.9245	34.628	38.1581
MARCH	30.6813	35.2858	38.1001
APRIL	30.5416	35.8743	38.0383
MAY	30.5044	36.3981	37.9526
JUNE	30.5729	36.8548	37.8464
JULY	30.7419	37.2436	37.7175
AUGUST	31.0151	37.567	37.5732
SEPTEMBER	31.3937	37.824	37.4099
OCTOBER	31.8723	38.0117	37.2231
NOVEMBER	32.4565	38.1353	37.0225
DECEMBER	33.1438	38.1919	36.8005

Percent of Imports From Insecure Sources (continued)

TRANSP (AR.IMD + CE.IMD)

	1970	1971	1972
JANUARY	36.564	35.7598	34.2012
FEBRUARY	36.3508	35.8096	33.9675
MARCH	36.1048	35.8286	33.8115
APRIL	36.0015	35.8096	33.7288
MAY	35.8696	35.7605	33.728
JUNE	35.7634	35.6807	33.8042
JULY	35.6821	35.5627	33.9529
AUGUST	35.6301	35.4155	34.1851
SEPTEMBER	35.6045	35.2368	34.4912
OCTOBER	35.6045	35.0208	34.8728
NOVEMBER	35.6338	34.7761	35.3379
DECEMBER	35.6395	34.4978	35.8762

	1973	1974
JANUARY	36.4578	34.6267
FEBRUARY	36.9565	33.8884
MARCH	37.3359	33.2153
APRIL	37.5945	32.6074
MAY	37.7373	32.0698
JUNE	37.7607	31.6003
JULY	37.6626	31.1931
AUGUST	37.4495	30.8584
SEPTEMBER	37.1177	30.5903
OCTOBER	36.6636	30.3867
NOVEMBER	36.0937	30.2512
DECEMBER	35.4067	30.186

DOCUMENTATION FOR: PET.STIM

NAME: Days Supply of Petroleum Imports

DEFINITION: Number of days the U.S. could continue to consume crude oil and petroleum products at the demand rate in period t if all petroleum imports were stripped.

INTERPRETATION: Indicates the ability of the economy to absorb shocks in the international oil supply system.

UNITS: Days

FREQUENCY: Monthly

FORMULA:

$$\frac{CR.ST + DS.ST + GS.ST + RS.ST + JT.ST}{CR.IMD + DS.IMD + GS.IMD + RS.IMD + JT.IMD} \quad \frac{(bbl)}{(bbl/day)}$$

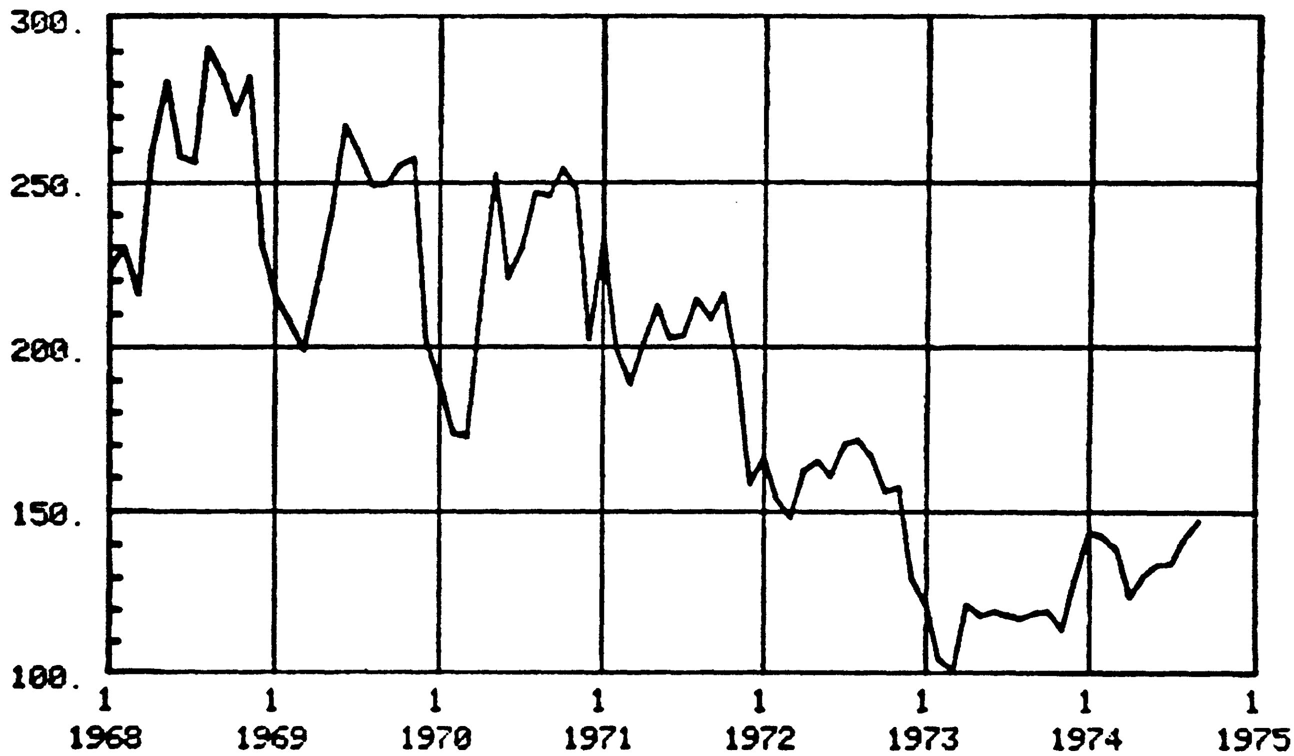
OUTPUT: Vertical axis is days supply, horizontal axis is time.

CONTACT: John Curtis, FEA, Office of Energy Data Policy, (202) 961-8796.

COMMENTS: Two graphs are shown, one where the fuels are volume weighted and one where they are BTU weighted. The difference is insignificant.

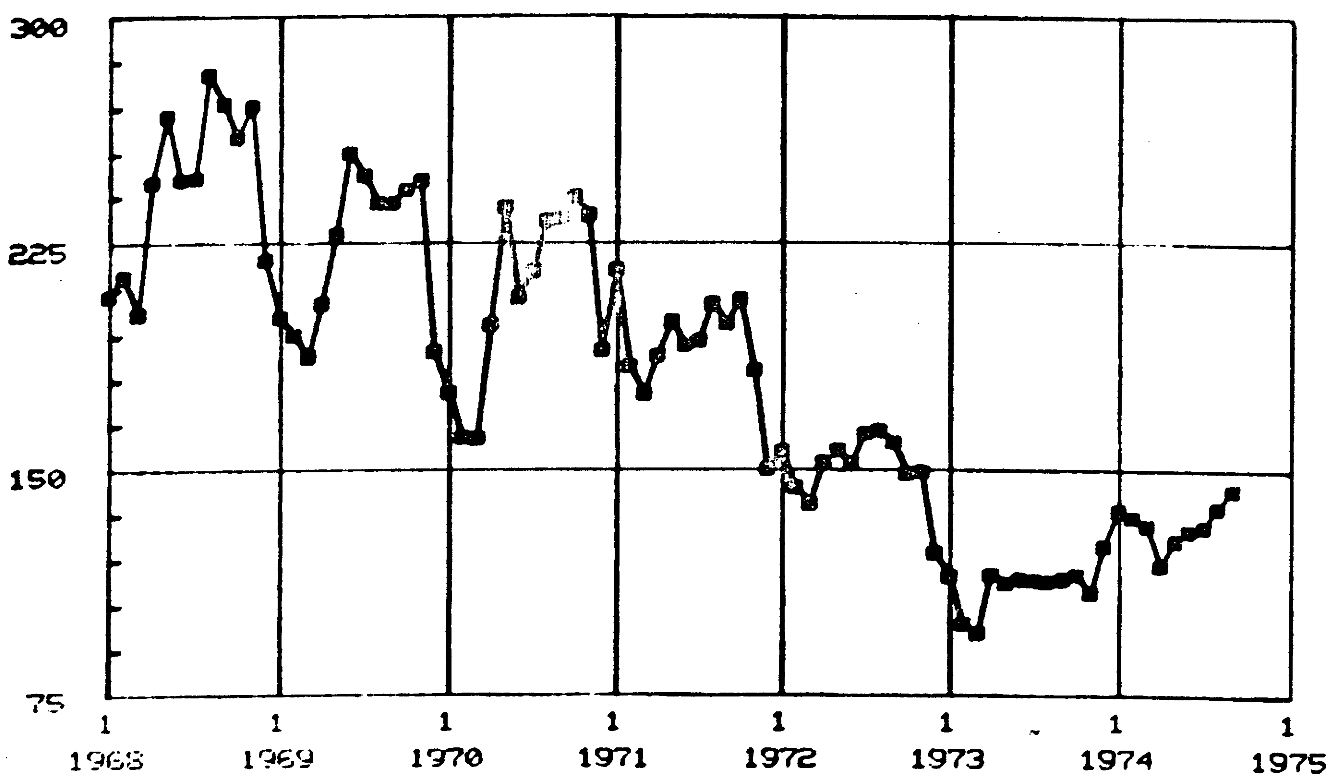
A third graph shows the level of stocks to total imports, and to "insecure" imports under two different definitions (see the documentation for AR.IMD, CE.IMD).

Days Supply of Petroleum Imports
(volume weighted)

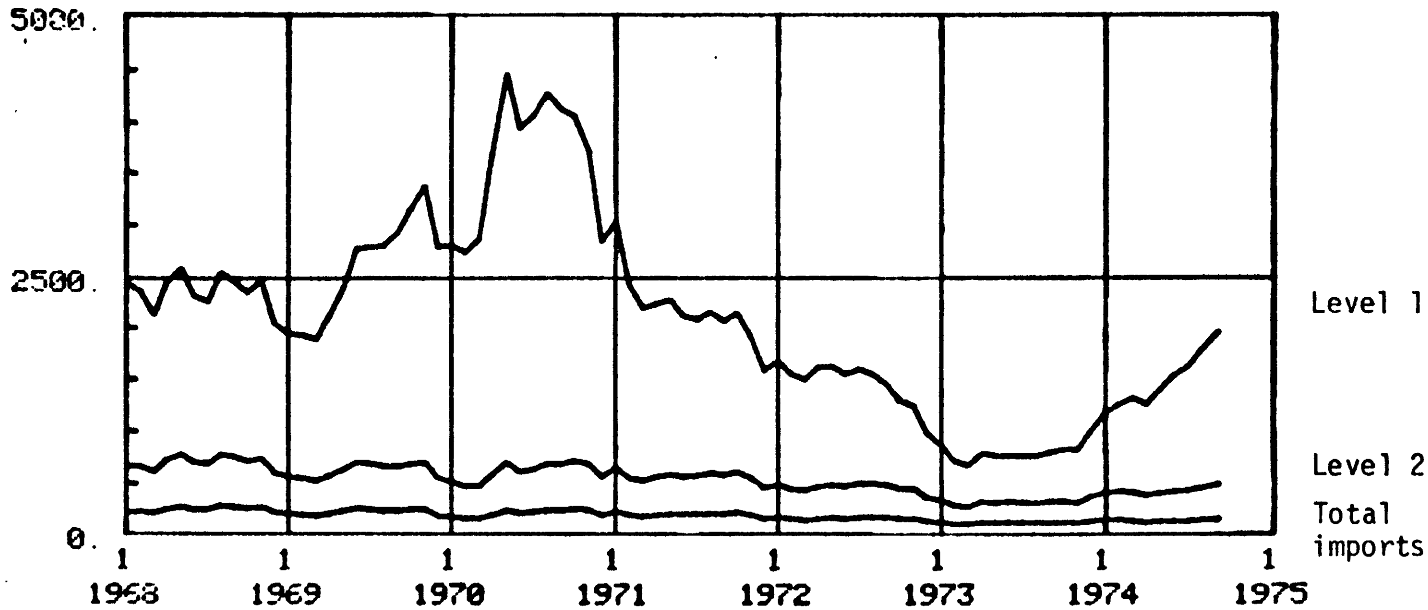


Days Supply of Petroleum Imports

(BTU weighted)



Days Supply of Petroleum Imports



Level 1 indicates stocks in relation to direct Arab Persian Gulf imports.

Level 2 indicates stocks in relation to imports directly from Arab Persian Gulf sources, and from European and Caribbean refining centers which may originate from Arab Persian Gulf sources.

Days Supply of Petroleum Imports
(volume weighted)

TRANSP(PET.STIM)

	1968	1969	1970
JANUARY	223.815	214.677	188.423
FEBRUARY	230.051	207.676	173.234
MARCH	216.223	199.142	172.591
APRIL	260.079	217.425	213.537
MAY	280.892	240.239	252.165
JUNE	257.986	267.391	220.799
JULY	256.561	258.835	229.418
AUGUST	291.222	248.793	246.312
SEPTEMBER	282.755	249.388	245.612
OCTOBER	270.935	255.266	254.935
NOVEMBER	282.523	257.243	247.404
DECEMBER	239.208	200.863	201.813

	1971	1972	1973
JANUARY	232.054	165.431	121.465
FEBRUARY	198.382	152.967	104.527
MARCH	188.638	147.747	101.157
APRIL	200.765	161.777	121.128
MAY	212.096	164.784	117.84
JUNE	202.455	159.954	119.252
JULY	203.916	169.984	118.925
AUGUST	214.349	171.166	117.334
SEPTEMBER	208.225	166.02	118.844
OCTOBER	216.015	155.422	119.291
NOVEMBER	192.213	156.849	113.903
DECEMBER	158.209	129.067	130.04

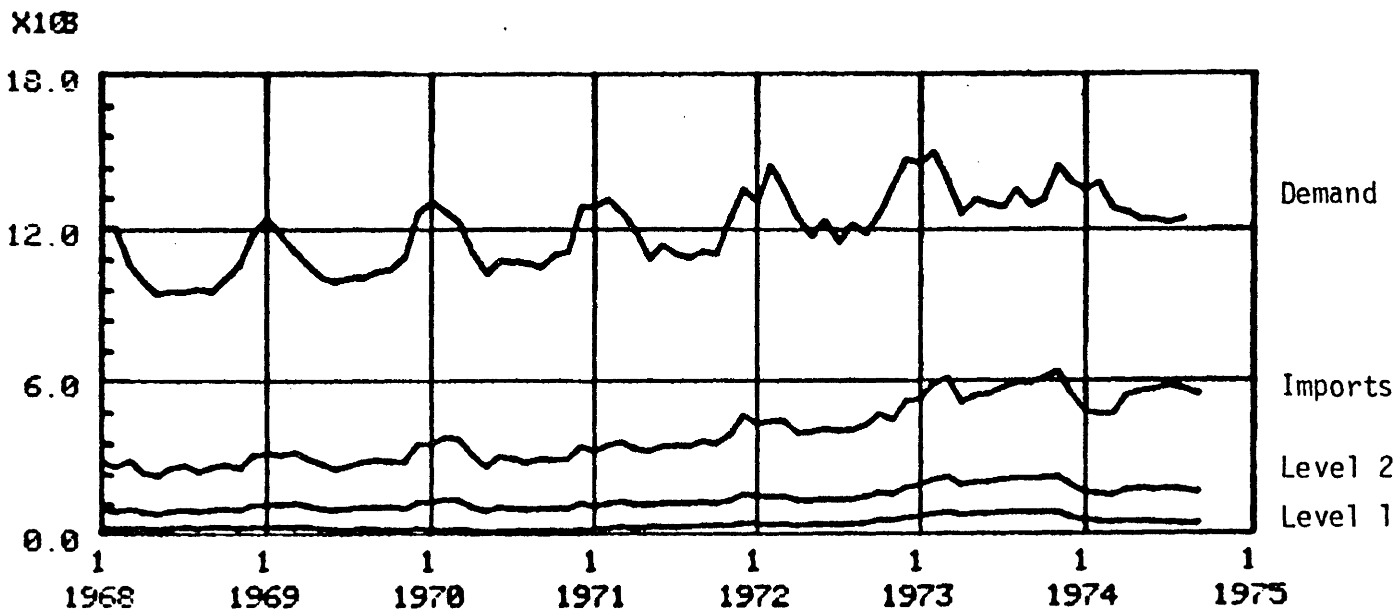
	1974
JANUARY	142.999
FEBRUARY	142.942
MARCH	138.043
APRIL	123.78
MAY	139.389
JUNE	133.594
JULY	134.075
AUGUST	140.993
SEPTEMBER	146.912
OCTOBER	NA
NOVEMBER	NA
DECEMBER	NA

TROLL COMMAND:

DOCUMENTATION FOR: (See Below)

NAME:	Import Dependence
DEFINITION:	Total demand for petroleum products, and total imports from all sources and from insecure sources
INTERPRETATION:	Shows the composition of U.S. petroleum supply, domestic and foreign.
UNITS:	bbl/day
FREQUENCY:	Monthly
FORMULA:	See plot
OUTPUT:	10^6 bbl/day on vertical axis, time on the horizontal.
CONTACT:	John Curtis, FEA, Office of Energy Data Policy, (202) 961-8796.

Import Dependence



Level 1 indicates imports directly from Arab Persian Gulf sources.

Level 2 indicates imports directly from Arab Persian Gulf sources, and from European and Caribbean refining centers which may originate from Arab Persian Gulf sources.

3.3.4 Prices

In this area, three indicators of the relative price of selected fuels and electric power have been prepared. They show the clear shift in relative prices in 1973-74, and the adjustment in the gasoline price in the months that followed.

Clearly much more work could be done on this group of indicators. The following merit high priority:

(1) Indices like the ones shown here, only for a weighted sum of fuels. Very likely this could be worked out with the cooperation of the Department of Commerce and/or the Bureau of Labor Statistics.

(2) Indices showing the effects of price control, such as the variation in prices of natural gas from region to region, or the components of the average oil price, showing the "old oil" and "new oil" contributions and the shifts in these shares over time.

DOCUMENTATION FOR: CPIF.DS

NAME: The ratio of the number 2 home heating oil consumer price index to the all items consumer price index.

INTERPRETATION: This ratio shows the movement in retail home heating oil prices relative to the general level of all items in the CPI. The fraction rises rapidly in the fall of 1973 following the Arab embargo. Since the beginning of 1974, this indicator has flattened out even though the distillate CPI has been increasing. This is because the all items CPI has been increasing at a faster rate. The index value is 1.0 in 1967.

UNITS: Percent

FREQUENCY: Monthly

INPUTS: CPI.DS--The CPI for number 2 home heating oil.

PC--the all items CPI.

FORMULA: $$CPIF.DS = \frac{CPI.DS}{PC}$$

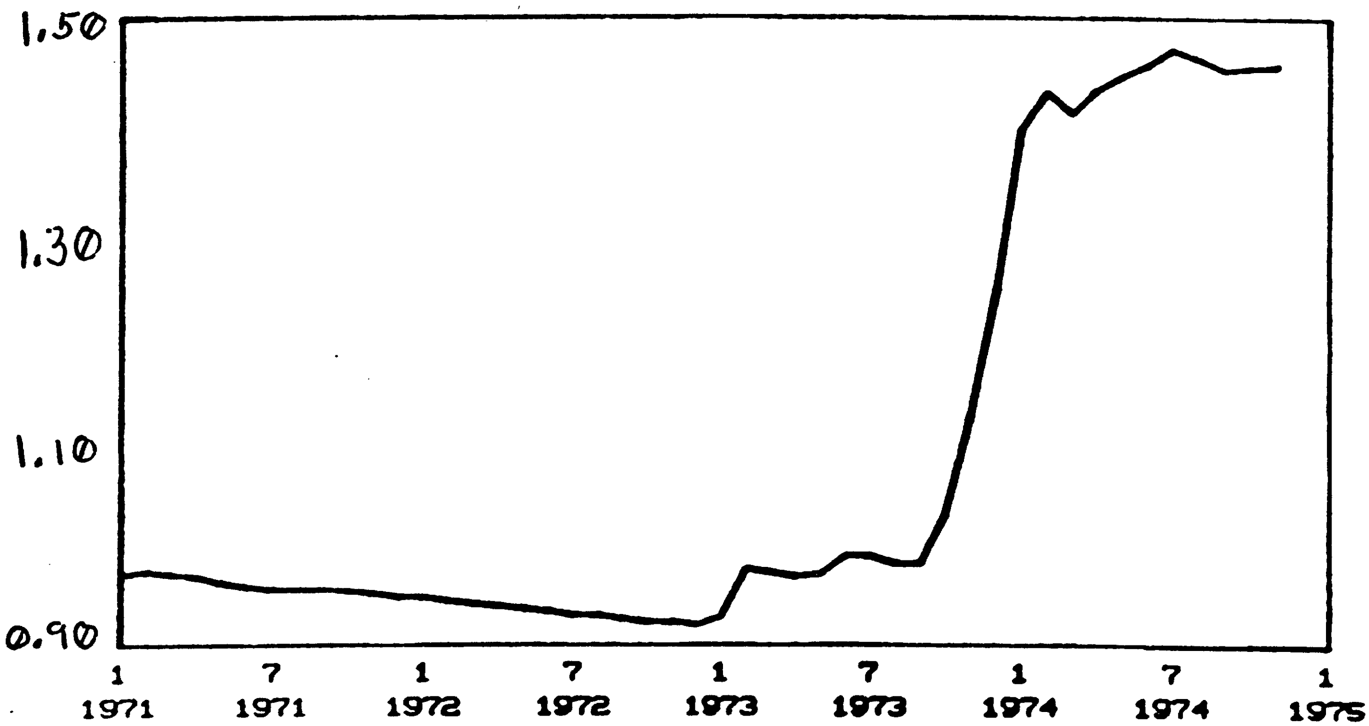
SOURCE: Bureau of Labor Statistics

CONTACT: John Curtis, FEA, Office of Data Policy, (202) 961-7986.

COMMENTS: See cautions listed with PC and PW documentation.

Relative Price of Distillate Oil

(value in 1967 = 1.0)



Relative Price of Distillate Oil

TRANSP(CPIF.DS)

	1971	1972	1973
JANUARY	0.967282	0.945617	0.92639
FEBRUARY	0.969849	0.941034	0.972784
MARCH	0.968281	0.939516	0.968413
APRIL	0.965053	0.937249	0.96557
MAY	0.959437	0.934242	0.968061
JUNE	0.955556	0.932	0.986405
JULY	0.953202	0.928287	0.985682
AUGUST	0.953317	0.927605	0.977054
SEPTEMBER	0.952537	0.92393	0.977122
OCTOBER	0.95098	0.921801	1.02489
NOVEMBER	0.949429	0.921986	1.12064
DECEMBER	0.945573	0.919089	1.24043

	1974
JANUARY	1.39227
FEBRUARY	1.42827
MARCH	1.40811
APRIL	1.43085
MAY	1.44467
JUNE	1.45541
JULY	1.47095
AUGUST	1.46164
SEPTEMBER	1.45089
OCTOBER	1.45294
NOVEMBER	1.45496
DECEMBER	NA

TROLL COMMAND: >

DOCUMENTATION FOR: CPIF.EL

NAME: The ratio of the electricity consumer price index to the all items consumer price index.

INTERPRETATION: This ratio shows the movement in retail electricity prices relative to the general level of all items in the CPI. Beginning early in 1974 the rate of increase in electricity prices exceeds that of the general price level. Also, the relative electricity price has not receded in recent months. The index value is 1.0 in 1967.

UNITS: Percent

FREQUENCY: Monthly

INPUTS: CPI.EL--The CPI for electricity

PC--The all items CPI

FORMULA: $CPIF.EL = \frac{CPI.EL}{PC}$

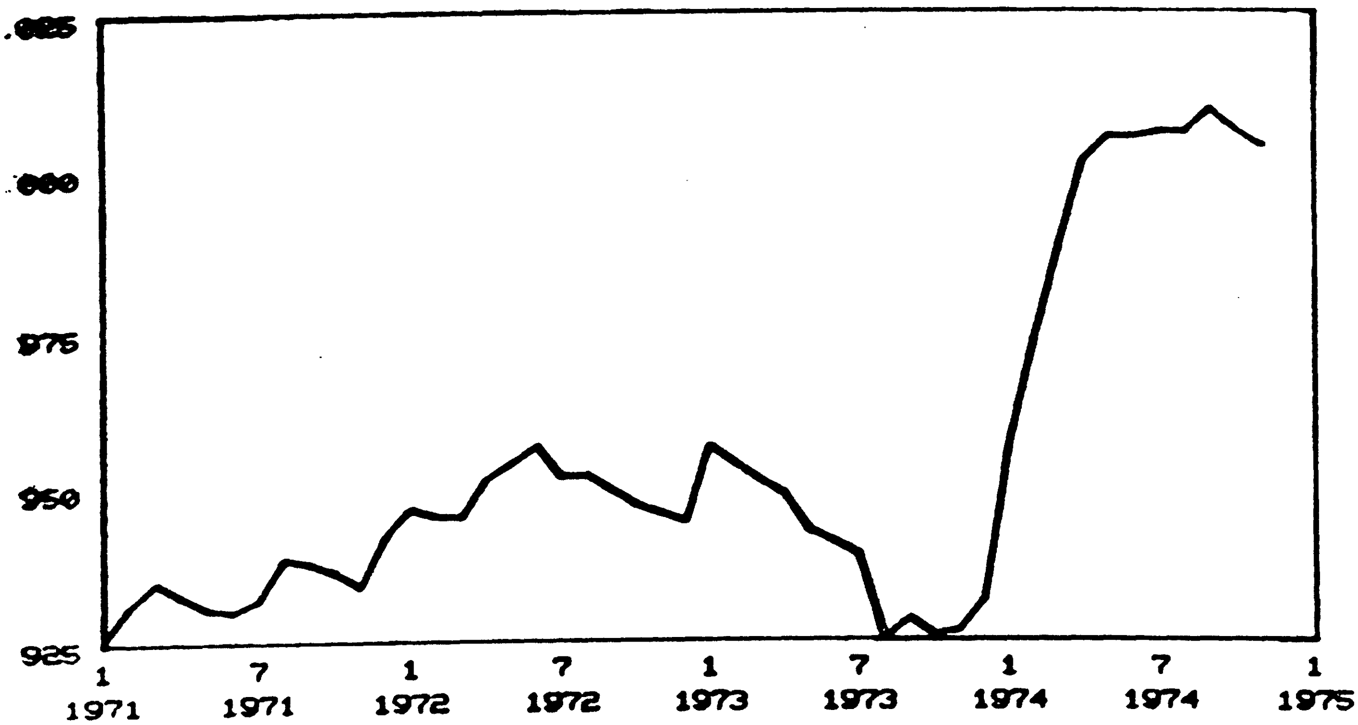
SOURCE: Bureau of Labor Statistics

CONTACT: John Curtis, FEA, Office of Data Policy, (202) 961-7986

COMMENTS: See cautions listed with PC.

Relative Price of Electricity

(value in 1967 = 1.0)



Relative Price of Electricity

TRANSP(CPIF.EL)

	1971	1972	1973	1974
JANUARY	0.926174	0.946429	0.956147	0.955619
FEBRUARY	0.931323	0.945073	0.953344	0.973145
MARCH	0.934892	0.945161	0.950693	0.98812
APRIL	0.932612	0.950925	0.948738	1.00139
MAY	0.930464	0.953488	0.942966	1.0055
JUNE	0.930041	0.956	0.941088	1.00545
JULY	0.931856	0.951394	0.93896	1.00608
AUGUST	0.938575	0.951472	0.925241	1.006
SEPTEMBER	0.937807	0.949287	0.928413	1.00989
OCTOBER	0.936275	0.947077	0.926062	1.00654
NOVEMBER	0.933932	0.945626	0.926599	1.00389
DECEMBER	0.942323	0.944226	0.931408	NA

DOCUMENTATION FOR: CPIF.MG

NAME: The ratio of the motor gasoline consumer price index to the all items consumer price index.

INTERPRETATION: This ratio shows the movement in retail gasoline prices relative to the general level of all items in the CPI. In the fall of 1973, following the OAPEC embargo, gasoline prices began to rise at a rate much faster than the all items index. In 1974, the rate of increase in CPI.MG has been much slower than increases in the general level of prices so that the curve tends to fall in the summer of 1974. The index value is 1.0 in 1967.

UNITS: Percent

FREQUENCY: Monthly

INPUTS: CPI.MG--The CPI for Motor Gasoline

PC--The all items CPI

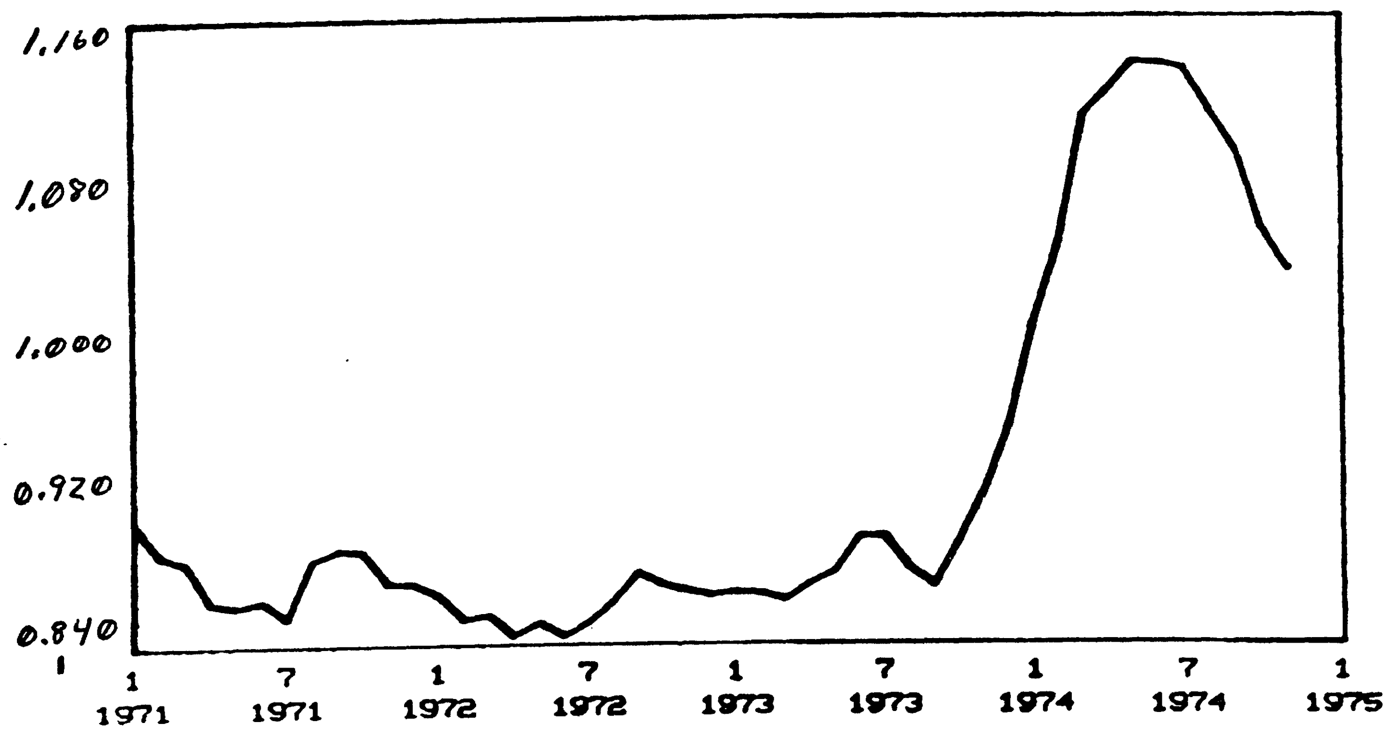
FORMULA: $CPIF.MG = \frac{CPI.MG}{PC}$

SOURCE: Bureau of Labor Statistics

CONTACT: John Curtis, FEA, Office of Data Policy, (202) 961-7986.

COMMENTS: See cautions listed with PC and PW documentation.

Relative Price of Gasoline



Relative Price of Gasoline
(Value is 1.0 in 1967)

TRANSP(CPIF.MG)

	1971	1972	1973	1974
JANUARY	0.903523	0.866071	0.866875	1.00286
FEBRUARY	0.887772	0.853796	0.866252	1.04452
MARCH	0.883139	0.855645	0.862866	1.10901
APRIL	0.862729	0.844731	0.870696	1.12161
MAY	0.860927	0.851644	0.876806	1.13677
JUNE	0.863374	0.8448	0.89426	1.13615
JULY	0.85468	0.851793	0.895252	1.13311
AUGUST	0.883702	0.862371	0.878608	1.11074
SEPTEMBER	0.889525	0.876387	0.869373	1.09031
OCTOBER	0.888889	0.870458	0.891654	1.05163
NOVEMBER	0.871941	0.867612	0.917878	1.03046
DECEMBER	0.871649	0.865672	0.952347	NA

3.3.5 International Market

The one indicator displayed in this area is concerned with the level of excess capacity in the OPEC nations--a key variable in charting the history of any international cartel. Other high priority items in this area include:

(1) Exploratory activity worldwide, and reserves and capacity data for oil and gas.

(2) Capacity expansion in the non-OPEC nations, and among the "competitive fringe" within OPEC.

(3) Price series on international sales of oil out of particular countries (as opposed to landed cost, which is useful but is clouded by transport cost and financial phenomena).

(4) Series which describe the international flows of nuclear material in various stages of the fuel cycle.

DOCUMENTATION FOR OPEC

NAME: OPEC Excess Capacity

DEFINITION: Net productive capacity of OPEC members compared with total production.

INTERPRETATION: A rough estimate of cartel excess capacity which is one measure of the market pressure the cartel is currently withstanding.

FREQUENCY: Capacity-annual; production-monthly.

SOURCE: Manipulation of data from
World Oil (WO), August 15 issue
Oil and Gas Journal (OGJ), Worldwide Oil Issue
Petroleum Industry Trends (PIT)

FORMULA: The method involves taking the previous peak production for each country and adjusting it by (1) the new capacity added by new drilling, yielding an estimate of gross capacity and (2) deduction of capacity loss due to normal decline, yielding an estimate of net capacity. Taking Saudi Arabia as an example, the method is the following.

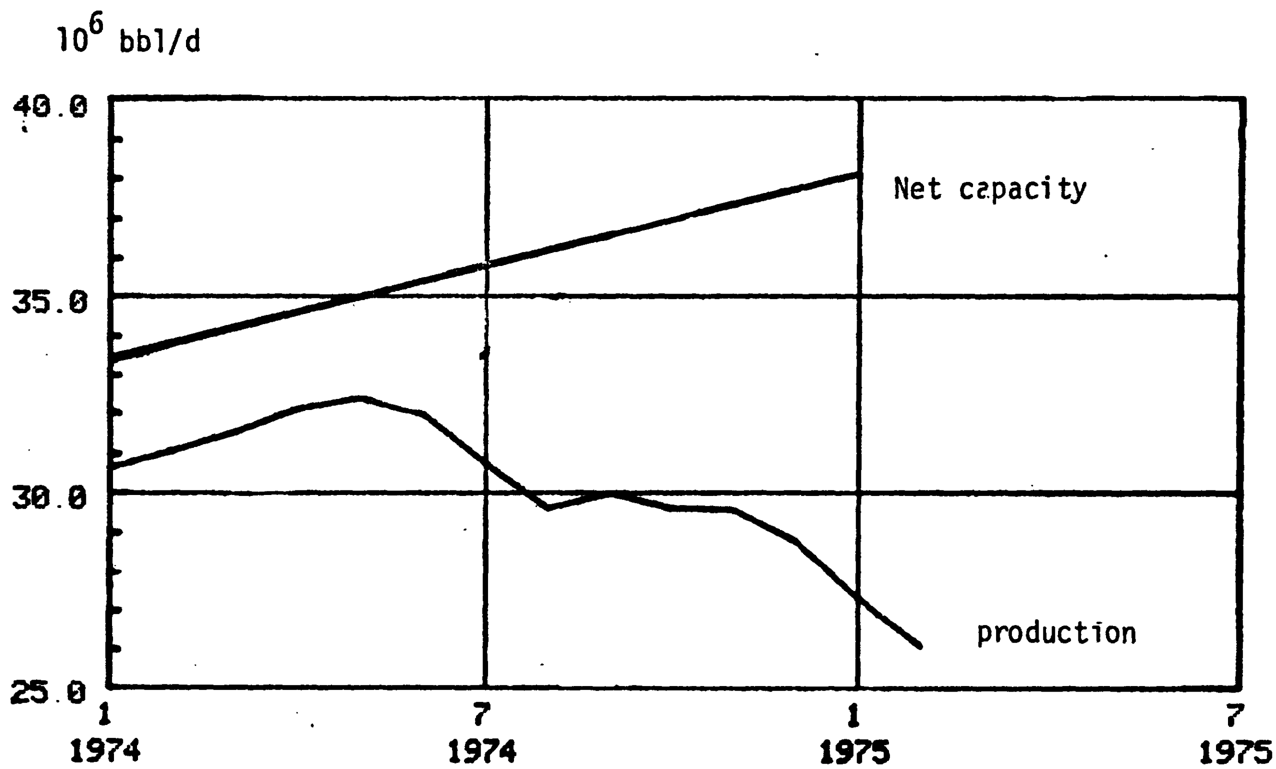
1.	Total well completions (WO)	1972 1973	176 323
2.	Oil well completions	1972 1973	104 177
3.	Percent of completions that represent oil, (2)/(1)		.57
4.	1974 forecast of total completions (WO)		350
5.	1974 estimated oil completions, (3)x(4)		200
6.	1973 estimated new oil wells per month, (2)/12		14.8
7.	1974 estimated new oil wells per month, (5)/12		16.7
8.	Previous peak output, July 1973, in thousands of barrels per day (PIT)		8440
9.	Months since previous peak output		17
10.	Average monthly build-up of new wells, ((7)-(6))/2		.96
11.	New wells per month plus average monthly build-up rate in 1974, (6)+(10)		15.7

12.	New wells since previous peak output, as of December 1974, (11)x(9)	267
13.	Output for first 6 months of 1974, in thousands of barrels per day, (OGJ)	8066
14.	Flowing and pumping wells, first 6 months of 1974 (OGJ)	670
15.	Average well output, first 6 months of 1974, in barrels per day, (13)/(14)	12.040
16.	Capacity added through December 1974, in thousands of barrels per day, (12)x(15)	3214
17.	TOTAL GROSS CAPACITY from sum of previous peak output and capacity added based on well build-up and average productivity per well, in thousands of barrels per day	11,700
18.	1974 annual output estimate in billions of barrels (OGJ)	3.066
19.	Estimated reserves as of 1/1/75 in billions of barrels (OGJ)	164.5
20.	Calculated decline rate, (18)/(19)	.019
21.	Revised decline rate, based on judgment that reserves estimates outside the U.S. are overly generous by API standards, (20)x2. Use of a factor less than two for this correction will produce a net capacity estimate less conservative than the one calculated here.	.038
22.	Capacity loss in billions of barrels per year, (18)x(21)	.117
23.	Change in net capacity in billions of barrels per year, $((16)\times365\times10^{-6})-(22)$	1.056
24.	Change in net capacity on daily basis, in thousands of barrels per day, $((23)/365)\times10^{6}$	2893
25.	TOTAL NET CAPACITY, in thousands of barrels per day, (24)+(8)	11,300

This method is followed for each country in OPEC. 1973 peaks were used for Saudi Arabia (July), Kuwait (January), Libya (October), and Algeria (May), for these countries were contraining output. Note that the all-time peak for Libya was in May 1972, but our estimates are not based on pre-1973 output data.

For Ecuador, Indonesia, and Venezuela certain adjustments were made relating to well productivity. In Ecuador, output from coastal wells was omitted as well as the number of pumping and flowing wells in the coastal area. In Indonesia, we exclude both output and pumping and flowing wells from the districts of Pertamina I, II, IV, and V, Lemigas, PT Stanvac, and Corridor Block. In Venezuela, we do not use the OGJ number of pumping and flowing wells at all. We use for well-productivity in Venezuela the 5-year average (1964-1968) figure from M. A. Adelman, The World Petroleum Market (Johns Hopkins Press, 1972), p. 294. The reason for these adjustments is that the large number of old or small wells would bias downward the productivity per well in these countries.

OPEC Excess Capacity



OPEC PRODUCTION

TRANSP(OP.PN)

	1974	1975
JANUARY	30717.	27373.
FEBRUARY	31120.	26137.
MARCH	31601.	NA
APRIL	32191.	NA
MAY	32435.	NA
JUNE	32050.	NA
JULY	30799.	NA
AUGUST	29678.	NA
SEPTEMBER	30061.	NA
OCTOBER	29665.	NA
NOVEMBER	29653.	NA
DECEMBER	28837.	NA

10^3 bbl/day

OPEC CAPACITY

OP.CAPN

	ANNUAL DATA
Jan. 1974	33313.
Jan. 1975	38021.

10^3 bbl/day

4. DYNAMIC-STOCHASTIC INDICATORS

As stated earlier, the term "leading indicator" generally connotes two separate purposes: the "leading" part suggests some kind of forecast; the "indicator" part connotes clear presentation of relevant data. This section discusses how the two purposes of forecasting and indication relate to explicit mathematical models. The type of models discussed here include interactions among different variables over time (dynamics), and the implications of uncertainty in the variables (stochasics). Whereas Section 3 concentrates primarily on issues of data with some discussion of simple static models, this section emphasizes the interaction between data and explicit dynamic models, with numerical assessments of uncertainty in both models and data.

Our conclusions about "leading" indicators of the snapshot type (analogous to the NBER leading indicators) are pessimistic. The hope for a "cheap forecast" without explicit assumptions about the future seems a false hope. Therefore, the burden of forecasting energy sufficiency must remain with econometric models and simulation. In our framework, these models, coupled with the data used to estimate and verify them, are called "dynamic indicators." When the model also includes quantitative estimates of the uncertainty of data and forecasts, the indicator also qualifies as "stochastic".

In response to FEA priorities, the bulk of our work has been directed not at dynamic/stochastic indicators, but toward the examination and development of "snapshot" indicators. Thus, the purpose of this section is not to propose specific dynamic indicators, but to give a brief overview of some new capabilities for the estimation of dynamic models and the use of such models to make forecasts.

Econometric models and simulation, or course, are by no means new to the FEA. However, it may be appropriate here to describe some relatively new methods and capabilities, which indicate possible future directions for the development of forecasts and dynamic/stochastic indicators. The FEA may want to consider the use of some of these techniques in their work on econometric forecasts and simulation models.

4.1 Background

The purpose of this section is to give an overview of some relatively new capabilities for dealing with models and data. Mathematical details are, for the most part, omitted, with the emphasis placed on practical capabilities and applications. If required, details may be found in Appendix E and in sources referenced there. The methods to be discussed are based on some recent results of the field of modern control theory. The practicality of the methods has been demonstrated for some engineering systems. Applications to social systems are more recent, but have been promising.

The methods are perhaps best collectively described as "full-information maximum likelihood via optimal filtering." The necessary computational tools are implemented in a publicly-available computer program, called General Purpose System Identifier and Evaluator (GPSIE).

The methods of GPSIE are well-suited to problems of energy models and data, since they can be applied to many situations where traditional econometrics methods do not apply. The following capabilities are especially important.

(1) The method works for the general class of nonlinear, time-varying

dynamic models with partial observations. Thus, the models need not be linear in the parameters; the models may contain variables for which there is no data.

(2) The methods of GPSIE do not require that the model have the same periodicity as the data. For example, annual data may be incorporated in a monthly model.

(3) The method can handle errors in all the variables, with error means and variances taken as unknown, and without resort ot instrumental variables. Thus, corrupted data or any variable can be used to help estimate model parameters, without causing those estimates to differ from their maximum-likelihood values. Furthermore, the accuracy of the data can itself be estimated.

(4) In dealing with corrupted data, GPSIE generates optimal estimates of the "true" values of the measured variables. Thus given a good model which relates several "soft" data series, GPSIE can be used to enhance the accuracy of the data. In addition, GPSIE computes confidence bounds on both the original data and the enhanced data.

(5) GPSIE also computes the confidence bounds of simulated variables. Thus, forecasts may be accompanied by numerical estimates of accuracy.

Some of these ideas and capabilities are further discussed and illustrated in the following subsections. Section 4.2 illustrates the computation of confidence bounds for forecasts. Section 4.3 discusses the handling of inaccurate data. Section 4.4 briefly discusses issues of parameter estimation.

4.2 Confidence Bounds on Forecasts

Forecasts are often made using simulation models by hypothesizing future trands for any exogenous inputs, and simulating the model forward in time. The output of the model then represents estimates of future values of endogenous variables of interest.

Like all estimates, the forecast is at best the expected value of a stochastic process. The uncertainty of the forecast has two sources: First, the hypothesized future value of the exogenous variables are themselves uncertain estimates of the actual future values. If the exogenous inputs are subject ot 5 percent error, the forecast-output may be expected to be subject to similar errors, depending on the model. The second source of error is the model itself. Each equation of a model is subject to error, either because of approximate formulation or because of effects omitted altogether.

Thus, it is very unlikely that any forecast will predict exactly, even one month ahead. The question is not whether the forecast will be off, but how much it will be off. As a consequence, it is desirable to compute and display confidence bounds on forecasts wherever possible. When this is done, the uncertainty is often surprisingly large. Not only do the various sources of uncertainty tend to add to each other, but the uncertainty of the forecast may grow rapidly in time.

A simple example illustrates the point. The graph of Figure 4.1 shows a forecast of primary gasoline stocks, based on a simple econometric model.[1]

---

[1] The model is a preliminary one, being only a first step towards a short range model of domestic production of petroleum products (more detailed information on the model and its formulation can be found in a paper by Tepper [12]).
The example model equations are:

$$GSPDD_t = C1*STIX_t + C2*GSF_t *(GSU_{t-1} + GSU_{t-2} + GSU_{t-3})/3$$

$$GSU_t = GSDMD_t/GSF_t$$

$$GSST_t = GSST_{t-1} + (GSPDD_{t-} GSDMD_t-GSIMD_t)*DAYINMO$$

(footnote continued next page)

The forecast trend is bounded above and below by "two-standard-deviation" confidence bounds. Thus, under the assumption of normal (Gaussian) error, the "true" forecast is expected to lie within the confidence bounds with probability of about 95%. (For convenience, in the ensuing discussion all data accuracies will be stated as "± X %", which is taken to mean that two standard deviations of the variable in question are equal to X% of the average value of the variable.)

The model which generated Figure 4.1 has four structural equations, of which two are definitional and have no uncertainty. The remaining two equations are estimated, on the basis of past data, to have uncertainties of ±5% and ±8%. The exogenous inputs (gasoline demand measured as primary stock drawdowns, and gasoline imports) are assumed to have a constant accuracy of ±5%, which is perhaps an optimistic estimate.

Note that these modest uncertainties produce relatively large uncertainty in the forecasts of gasoline stocks. Only a few months out, the confidence bounds have spread to encompass practically any reasonable figure. Thus, any point estimate of gasoline stocks (from models of this type and accuracy), even a few months out, should be taken as very tentative.

---

[1] (continued)

where $GSPDD_t$ = gasoline production, daily average in month t.

$STIX_t$ = index of traditional change in gasoline stocks in month t.

GSF_t = seasonal factor for gasoline demand in month t.

GSU_t = seasonally-adjusted gasoline demand in month t.(daily average)

$GSDMD_t$ = daily average gasoline demand in month t.

$GSST_t$ = gasoline primary stocks at end of month t.

$GSIMD_t$ = daily average imports of gasoline in month t.

$DAYINMO_t$ = number of calendar days in month t.

All data to April, 1974 are from the Bureau of Mines, and thereafter from FEA.

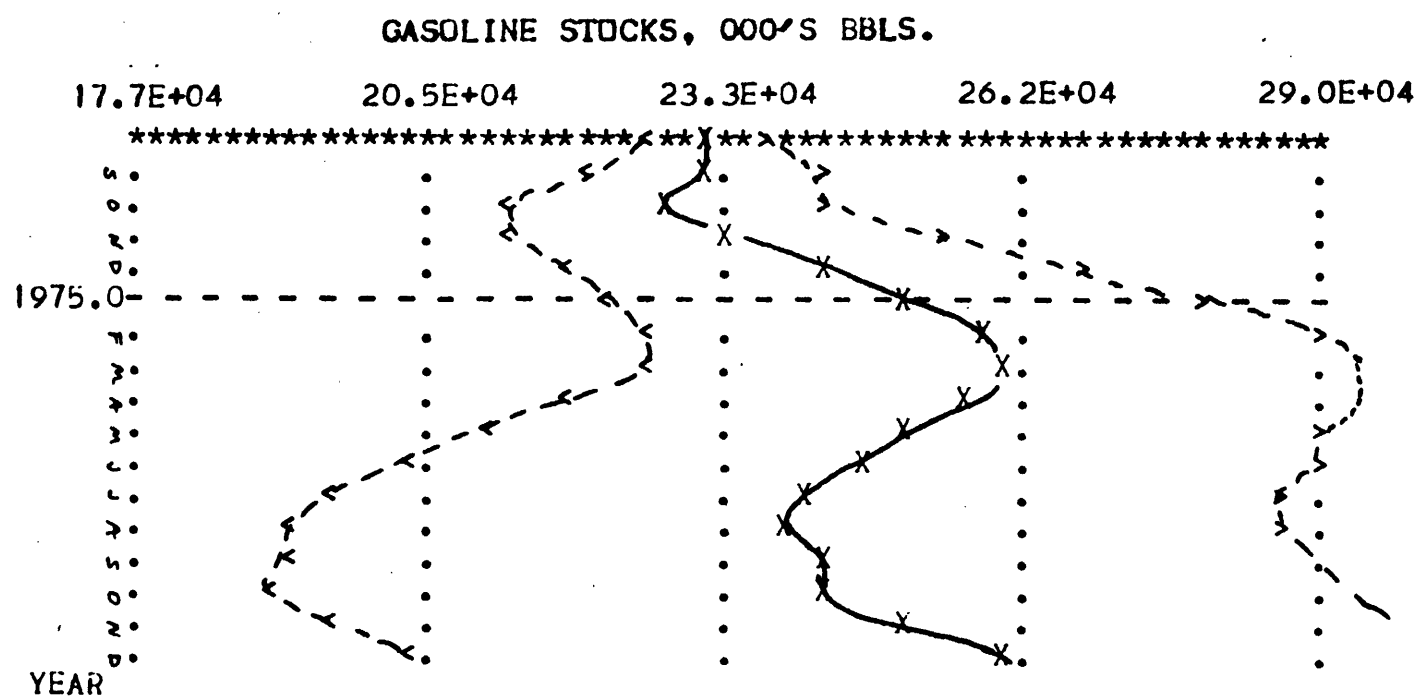


Figure 4.1 Gasoline Stocks, Forecast with confidence bounds; September 1974 through December 1975. (Note: based on tentative model: for illustrative purposes only).

Of course, a more accurate and complete model might yield better (tighter) confidence bounds on the forecasts. Once again, it is important to observe that the model used here is for illustration of techniques only. However, for any particular model and data, the confidence bounds provide a tangible and clear statement of what accuracy to attribute to the forecast.

4.3 Parameter Estimation

Econometric models are usually estimated using ordinary least squares (OLS), or some variant, such as weighted least squares (WLS), or generalized least squares (GLS). This section summarizes some of the advantages of using full-information maximum likelihood (FIML) estimates instead.

The gasoline model used as an example above was intentionally kept simple (at a possible sacrifice in overall accuracy), so that OLS could still be applied, and the results compared with the maximum likelihood values. The results are summarized in Figure 4.2. The estimated parameters in Figure 4.2 are of three kinds. C1 and C2 are structural parameters. In the case of C2, the two methods agree closely; but the maximum likelihood value for C1 differs by 25% from the OLS value. Controlled experiments using simulated data [11] indicate that for OLS and GLS estimates in the presence of moderate error in the measurements (as well as in the equations), such misestimates are not unusual.

The parameters Q1 and Q2 represent the error in two equations of the model. Under OLS, the two equations become a single equation with lagged variables, with the error of the equation represented by an R^2 statistic of .94. The maximum likelihood formulation yields two separate errors of ±5% and ±8%, which continue to yield a single-equation error of about ±10%

Parameter	Ordinary Least Squares Estimate	Full-imformation Maximum likelihood estimate (via optimal filtering)
C1 (magnitude of seasonal stock adjustments)	65	49
C2 (multiplier on demand)	~1	~1
Q1 (error in production equation)	both in one equation under OLS combined error might be taken as 6%, since R^2 for the equation is .94	±5%
Q2 (error in seasonal factors)		±8%
R^1 (error in production data)	no estimate - by OLS to be zero	±4%
R^2 (error in stock data)	no estimate - assumed to be zero	±3%

Figure 4.2 Comparison of estimation results.

(since the squares of the standard deviation are additive). It is difficult to compare the R^2 - .94 directly with the ±10% two-standard-deviation estimate, but the latter is suggestive of a greater equation error than the OLS estimate.

A more significant advantage of the "max-like" method over OLS, however, is illustrated by the ability to estimate R^1 and R^2, the error in the measurement of gasoline production and stocks. Since each of these variables appears in the right-hand side of an equation, OLS must assume the data to be error-free. The FIML method, as implemated is GPSIE, is able to separately estimate these uncertainties in the data, thereby providing an additional test of data validity. The max-like error estimates agree with the a priori estimate of the accuracy of stock and production data - the FEA people we interviewd concerning this data agreed that 3% was a reasonable estimate of the error in gasoline production and stock data. The max-like error estimates are ±3% and ±4%.

4.4 Enhancement of inaccurate data

To evaluate the likelihood of a given parameter value, the methods of GPSIE process the data with an "optimal filter" of which the model is an active component. A pleasant side effect of the filtering process is the production of maximum-likelihood estimates of the "true" values underlying the error-corrupted data. The data error in this case is small, but still serves to illustrate the idea. Figure 4.3 shows data on primary gasoline stocks from 1967 through 1974, with three kinds of data graphed. The X's represent the maximum-likelihood values of the primary gasoline stocks. The Z's (often "hidden" beneath the X's because of the low error in the data) are the actual data, as reported by the BOM and the FEA.

The pointed brackets are the two-standard-deviation confidence bounds on the estimates X. The stimated stocks X are in essence a compromise between one-period forecasts from the model, and the actual data. The compromise is based on the estimated error of the one-period forecast, compared with the estimated measurement-error of the one-period forecast, compared with the estimated measurement-error in the data. If the model equation errors are small compared to the data-measurement errors than the estimate will lie close to the one-step forecast; if the model is relatively inaccurate, the estimate will lie close to the measured data.

Because of GPSIE's ability to deal with monthly and yearly data at the same time, and the capability to "relate" separate data series possibility, using model variables for which there is no data, it is possible to combine several "soft" data series and extract from them "enhanced" versions of the same data.

If this seems to the reader like a cheap way to get good data out of bad, be aware that there is a catch: the filtering process requires a good model of the process from which the data comes. If the model is bad, the "optional" estimates will in fact be as bad as the model. But a good model, estimates correctly, can be used for data enhancement as well as forecasting. With the above caveat of mind, the FEA might consider using models to validate data, in addition to the more common practive of using data to estimate models.

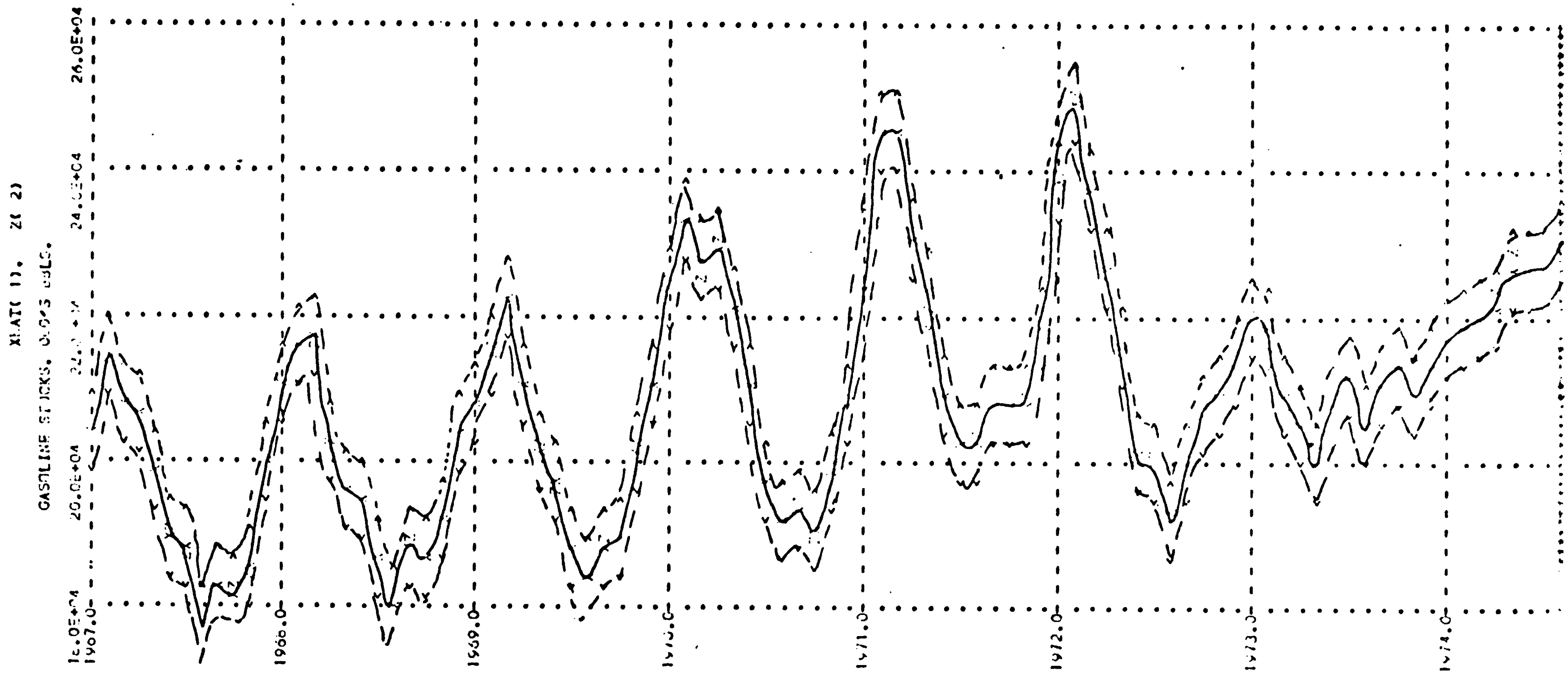


Figure 4.3 Gasoline stocks, January 1967 through August 1975. Estimates of true stocks plotted as X, graphed with solid line. BOM & FEA data plotted as Z (usually in same position as X, due to accuracy of the data). Two-standard-deviation confidence bounds plotted as pointed brackets, and graphed as dashed lines. (Note: estimates based on tentative model; for illustrative purposes only).

SECTION 5

FEA Final Report on Leading Energy Indicators Project

Principal Investigators

Henry D. Jacoby

John J. Donovan

March 1975

5. COMPUTER SYSTEMS FOR PRODUCING ENERGY INDICATORS

Computer systems can be of enormous help in managing the data used in constructing energy indicators, and in computing and displaying the indicators themselves. Many operations that would be prohibitively expensive and time-consuming if done by hand are made possible by modern computing technology. On the other hand, the data management system, if poorly constructed or if no more than a collection of incompatible parts, can be a nearly insurmountable barrier to effective data analysis.

In this section we discuss the desired characteristics of such a system, and present some of the alternative computer packages that are now available. We also cover some of the experimental systems for data management and analysis that are likely to revolutionize work in this area in the near future.

This discussion of computer systems naturally is oriented to the specific problem of producing energy indicators under the definition of this project. But the implications of this discussion are broader than this specific task, and have relevance to a wide range of data management problems within the Office of Data and in other branches of the FEA. For example, as defined in this report (see Section 2), the concept of an "energy indicator" is quite broad: any data series may have value in meeting the objectives set forth for these indicators -- even simple aggregates such as the number of barrels of gasoline imported last month. Indeed, most of the series published in the FEA's Monthly Energy Review this nature. Thus, the whole problem of producing that monthly document -- the data processing, analysis, and plotting -- is one to which this discussion of computer systems is relevant.

By the same token, there are many operations within the Office of Data that do not lead directly to published reports but which involve the similar computation difficulties. If the task involves the management of large blocks of data and the use of these data to compute numbers with some higher order of information content than each of the original data points, then from the standpoint of the management information system, the problems are essentially the same. Therefore, it is hoped that this section may have value not only to the narrow "indicator" problem but as a discussion of wider aspects of data processing within the FEA.

5.1 Desired System Characteristics

If a computer is to be used in the process of preparing energy indicators, there are several characteristics one would desire the system to have. Available commercial software systems have a range of different designs, each coming more or less close to the ideal system, and each having its peculiar quirks and limitations. In order to be able to review critically the characteristics of alternative sytems, it is useful to lay out the desired features. These fall into two classes: the data management system and the modeling and analytical capability.

5.1.1 Data Management System

One wants the system to have the ability to store, process, and retrieve various types of data. Specifically, the following features are desirable:

1. The system must allow the storage of quantities of various types of data. This includes numerical data,

as well as non-numeric data, such as character values for storing names.

2. The user should be given an adequate model of the way data is stored within the system. For example, the user may picture the data as tables, or networks, or trees. Later, we present a relational model for data [6], in which all the data stored by the system are viewed by the user as tables, independent of how the data may physically be stored.

3. The system should permit the user to select and access data according to varying criteria. Since it is impossible to specify in advance in what ways the data will be used, the system should allow for flexibility in query formulation.

4. The system should give the user means for easily viewing, inputting, and updating data. For example, a user should be able to make routine additions and revisions to monthly time series. Facilities should be provided for interactive specification of data to be added to the system, as well as a means for inputting data from files (such as on punched cards, or magnetic tape).

5. One would like the system to provide for <u>validation</u> of data. That is, data input to the system would be checked for constraints (e.g., a zip code must contain five numeric characters)

or <u>semantic</u> constriants (e.g., stocks(t) = stocks(t-1) + production(t) + imports(t) - consumption(t) - exports(t)).

6. One would like the system to provide for storing data about other data, such as levels of confidence in certain data, or sources from which pieces of data were gotten.

7. Users need to be able to impose security policies on the system. (Security <u>policies</u> are the specifications of what types of access to data is permitted. Security mechanisms are the tools for enforcing the security policies). An example of a security policy for a particular group of data may be: give everyone access to read the data; allow anyone with the Office of Energy Statistics to <u>append</u> or add new entries; allow only the person or persons with prime responsibility for the data to revise or alter data currently stored within the data base.

8. The user should have implicit assurance in the data integrity mechanisms and the security mechanisms. While he should not have to be burdened with the details of these systems, he must have confidence that adequate mechanisms exist.

5.1.2 Modeling and Analytical Capability

If the system is to be used to manipulate and analyze data, in addition to simply "managing" them, then there is a set of additional features that are needed.

1. The system should provide for the production of tables with selected raw data or summary data, as for example, the tables now produced in Monthly Energy Review.

2. Facilities for performing arithmetic operations on the data, such as summing or averaging, are needed.

3. The system should give the user the ability to produce plots, graphs, and histograms.

4. The system should have capabilities for more sophisticated statistical methods, such as regression and seasonal adjustment. For example, a data series such as demand for gasoline, which exhibits strong seasonal components, may be more informative if seasonally adjusted.

5. The system should permit the user to specify, construct, and execute mathematical models.

5.2 Commercially Available Systems

5.2.1. TROLL

The modeling system primarily used in this project to produce the leading indicators was TROLL [36]. TROLL was developed at NBER, Technology Square, Cambridge, Massachusetts. TROLL runs under CP or VM on an IBM 360 or 370 computer. It has been effectively used as an analytical tool for energy models [37].

TROLL has good analytical and statistical functions. It does have some shortcomings, however. First, TROLL is not imbedeed in a general purpose high-level language but rather stands alone. As such, it is difficult for a user to do something for which there is not an explicit TROLL command. Second, TROLL has limited data manipulation facilities. It is difficult to store and retrieve different types of data. It is difficult or impossible to protect data in flexible ways. These limitations, as well as TROLL's strengths, are recognized by NBER and they are proposing an advanced system, ACOS/DAZEL, to alleviate these deficiencies.

5.2.2 Other Systems

Numerous other data management and analytic systems are available, each of which satisfies some subset of the characteristics of Section 5.1. Some principal data base systems presently in use and commerically available are: MARK IV, RAMOS, IMS, ENQUIRE, OLIVER, ORACLE, GIS, JANUS, IDMS. Some modeling systems are: TSP, TROLL, EPLAN. The EPLAN/APL analytical component of our GMIS system, which is described in Section 5.3.2.4, is a commercially available product (from IBM).

All the above data management systems lack good analytical capabilities.

All the modeling facilities have deficiencies in data management capabilities. The major deficiency of both classes is lack of flexibility, as we discuss in the next section.

5.3 More Advanced Research-Oriented Systems

Each of the available data management and analysis packages has several limitations, but the major restriction is lack of flexibility in the use, access, and protection of data. This is a particularly damaging limitation in the context of the FEA's needs for several reasons:

(1) Since unforeseen uses and needs for the data inevitably arise, the system must be flexible so that it can adapt to these changing needs. This is particularly true when providing information for policy decisions in so volatile an area as energy.

(2) There are varying constraints imposed by changes in the quality, availability, and protection requirements of data. The system must be able to adjust to such moving constraints.

(3) The system must be able to accommodate changing needs and constraints at reasonable expenditures of cost and effort. Computer systems of a decade or two ago could support most current applications, but in many cases, only at a high cost. A flexible system makes it possible to easily experiment with many uses of the data at modest costs.

Because of these demanding characteristics of the data use patterns within the FEA, it is important to take advantage of the very best software systems that are available commercially now, and to be constantly looking ahead to the more advanced systems that are now being developed. In connection with this project, a system under development at M.I.T. [10,13] has been applied in several areas -- to serve the indicator problem [13,14] and to serve as an information system for New England [10]. It appears that this approach may offer great promise for the future application to the FEA's data management and analysis problems.

The description of the M.I.T. system given below is intended to serve three functions:

(1) To show the importance to FEA of operating systems presently available on the hardware the Agency is now purchasing and to show the flexibility of these operating systems, namely, how VM can be used to allow many incompatible modeling and data base systems to work together.

(2) To indicate a direction that management information systems are likely to take in the near future, so that these developments can be taken into account in the planning of the Office of Energy Data.

(3) To demonstrate a prototype system for producing energy indicators using these new concepts with the possibility that the FEA may want to consider adopting this system even in its current experimental stage.

5.3.1 Introduction to the GMIS System

GMIS(Generalized Management Information System) is a collection of software tools that facilitate the construction of management information systems, models, and user interfaces. The tools are particularly applicable to systems with the following characteristics:

- several classes of users, each of which has a different degree of sophistication
- complex and changing security requirements
- data that exhibits complex and changing inter-relationships
- changing needs to be met by the information system
- need for quick and inexpensive implementation
- complex data validation requirements
- complex models to be built that access data
- rapid and inexpensive construction.

The approach taken here to such a system is a hierarchical approach [8, 9, 10, 24, 26] both in implementation and presentation to the user. The approach is hierarchical in implementation because this technique provides for ease of debugging, independence of hardware, and a basis for investigating properties of completeness, integrity, correctness, and performance. It is hierarchical in its presentation to the user to take cognizance of the fact that levels of user sophistication demand appropriate command environments. As such, the casual system user has powerful, high-level commands at his disposal, while the sophisticated (perhaps also the more mathematically inclined) user has more detailed and basic commands, but with a low tolerance for error.

We have applied this general facility to several energy related areas, including a New England Energy Management Information System (NEEMIS) [10]. The purpose of NEEMIS is to establish a facility (for storing and validating data, retrieving data, interpreting and analyzing data, and constructing and applying models using those data), which will facilitate New England energy policy analysis and decisions.

In this report we will focus on our application of the GMIS facility to building a system for energy indicators. It should be noted that the term GMIS as used here has two distinct but related meanings. On the one hand, GMIS refers to the generalized system containing the facilities for constructing specific information systems; on the other hand, GMIS may refer to any of the particular information systems so constructed, such as the NEEMIS system, or the energy indicator system.

5.3.2 User View of GMIS

Keeping in mind the ultimate purpose of GMIS -- to provide a facility to aid the construction of management information systems, especially for use by public policymakers in making energy decisions -- we recognize several classes of users of the GMIS facility. In this section we shall briefly explain what facilities each class of user will have. The details of the precise syntax of intermediate languages and implementation details are described elsewhere [10].

Figure 5.1 depicts four classes of users, and the facilities available to each. These classes:

1. Non-technical -- e.g., an administrative assistant within the FEA. His objective is to get answers to specific questions and to produce reports.
2. Well-trained -- e.g., a specialist within the FEA who has been trained in the use of the system.
3. Researcher -- e.g., an economist with some computer background who wishes to build a model for a special study.
4. Systems analyst/programmer -- e.g., a computer professional. He may wish to add a new table to the system or change the protection rights on an existing data series.

FIGURE 5.1

GMIS - FUNCTIONS AVAILABLE IN EACH INTERFACE FACILITY.

INTERFACE FACILITY / USER CATEGORY	PREPARED PACKAGES	GMIS INTERACTIVE QUERY FACILITY	MODELING FACILITY	TRANSACT: RELATIONAL QUERY LANGUAGE (DML)	TRANSACT: DATA DEFINITION FACILITY (DDL)	RELATIONAL OPERATORS & PL/1 FACILITY
NON-TECHNICAL (No computer training)	X	X				
WELL TRAINED	X	X	X	X		
RESEARCHER	X	X	X	X	X	
SYSTEMS ANALYST & PROGRAMMERS	X	X	X	X	X	X

Looking across the table, we see the tools available to users of GMIS. Although all levels and facilities of the system are available to all users, it is unlikely that users will venture outside of those tools designated (by "X"). Increased sophistication on the part of any one user will, of course, qualify him/her for a different category.

The tools of the system have been designed in such a way that the interests of the various user groups are met. Let us proceed to briefly describe the facilities at each level.

5.3.2.1 Relational Operator and PL/1 Facility

At this level, the user sees all data as being stored in relations[1]. This includes not only regular user data, but all system data, all access control data, etc. The user at this level has at his command thirteen set-oriented relational operators that are used to perform all operations on all data. It is important to note that user data, system data, access control data, etc., are all accessed in a consistent manner via these thirteen operators, which are based on the relational model of data,[6] and [7], which have their roots in logical systems and predicate calculus [5], [11], [31], [35]. The operators available in GMIS are described in detail in [10]

Since these operators appear as PL/1 subroutine calls within GMIS, the user at this level also enjoys all the power of PL/1.

Notice that both PL/1 and relational operators require precise use and exhibit low tolerance for error.

5.3.2.2 Data Definition Facility

A user at this level has facilities to specify and create relations. We call this facility the Data Definition Language (DDL). The DDL will

---

[1]For our purposes, a relation can be thought of as a table. Each column takes its values from a set called a domain. Each row is an entry. See [6] for more details.

will accept a data specification and will produce an appropriate relational data base, which is then incorporated into the system. The DDL also provides a facility for loading bulk data into the newly constructed relational system from punched cards, magnetic tapes, or magnetic disk files.

In the establishment of a new relation, the system tables are modified to include data about this new relation, as well as provision for specification of access control, etc.

Also available at this level is on-line help with commands, and extensive diagnostics.

With most information management systems, the design of the system -- that is, the design of the data base -- is a vital step in the operation. If done incorrectly, it is often impossible and usually extremely costly in dollars and man-years to restructure the data base to more ably suit the needs.

Not so with GMIS. In fact, it has proved possible to experiment with three different designs in the course of a single month. The DDL permits specification of the data base on-line, and extremely rapidly.

An example of the use of the DDL facility is given in section 5.4.1.1.

5.3.2.3 Query and Update Facility

At this level a user can specify queries of data stored in relations. The user uses a rigid syntax for his queries that we sometimes call "cryptic" English. More specifically, we call this facility a Data Manipulation Language (DML).

An internal document describes a complete DDL and DML that has been specified at M.I.T. [34]. Other attempts at implementing a query facility based on the relational model include: MACAIMS [18], SEQUEL [4], COLARD [1], RIL [15] and M.I.T.'s RDMS.

This facility is available for querying relations established via the DDL (see 5.3.2.2), or possibly the relational operator facility.

The commands that are now operational are listed in the M.I.T. Energy Laboratory documents on TRANSACT. Although conforming to a rigid syntax, they employ English-like keywords and are readable and quite easy to learn. Once again, all data, including system data, are accessed in a consistent manner; and access control specification is an integral part of DML.

Figures 5.7 through 5.11 in section 5.4.1.3 are all examples of the query and update facility.

5.3.2.4 The Modeling Facility

A user of this facility may construct and activate a model interactively via provision of a set of functions called from APL. These functions include regression routines, plotting routines, time series modeling routines, etc., in addition to the standard APL facilities. The language used for modeling is APL/EPLAN, a superset of APL -- i.e., APL with additional facilities. The data that the model uses may be retrieved directly from that stored in the relations.

The APL-oriented modeling facility is the currently implemented facility. Inclusion of additional or different modeling languages, however, poses little problem .

5.3.2.5 GMIS Interactive Query Facility

The user of this facility simply points to a question category he wants answered on a CRT using a "light pen". If the question needs further specification, the system will flash subsequent choices on the scope, and the user will point to the choice that clarifies his query. Also, the graphic facility allows users to "peruse" through a large table using a window facility.

5.3.2.6 Prepared Packages

Users of this facility will request standard reports or invoke common models, for example, a monthly forecasting model. All the user at this level needs to know is the name of the report or model. The system will take care of retrieving the requisite data and invoking the appropriate facility to generate a report or run a model.

5.3.3 Hierarchical Implementation

The purpose of Section 5 is primarily to describe the hierarchy of user facilities in GMIS, as opposed to a discription of the implementation. However, there are two implementation techniques (hierarchical implementation and use of virtual machine) that we feel FEA should become familiar with.

We mention both of the techniques for several reasons.

1. Our hierarchical implementation is an example that we feel FEA should follow in all software development. We advocate this hierarchical approach for all complex software as it allows a straightforward method of design, implementation, and debugging, as well as localizing changes to the software.

2. The specific extension so the relational model of data using the hierarchical approach is one that FEA might adopt.
3. FEA is in the process of choosing an operating system; we wish to demonstrate the flexibility we have gained by using VM.

The GMIS system has been designed and implemented in a hierarchical fashion, in which the system is composed of a series of groups, or layers. Each layer represents a different user environment in which the system can be used. Complete details of hierarchical systems can be found in [10]. Early explanations and applications of the hierarchical implementation approach can be found in [8], [9], and [24].

Just as the user-view of GMIS describes levels of differing power and flexibility, so the actual implementation of the system is carried out. Software developed for the GMIS has been implemented as a multi-level heirarchy in which each level employs only those facilities implemented in the levels below it, as represented in Figure 5.2.

GMIS has paid extensive heed to security of data. Some nineteen types of access have been identified and any owner of data may authorize any user to access that data in any or all of those nineteen ways. The default authority is NO access, rather than the usual approach that allows full access unless otherwise specified. These security specifications are made via facilities in the DML directly.

The relation used to store access control information, as well as all other system relations and descriptors are accessed in an identical manner to regular user data. Thus all data in the system are stored in a consistent fashion, making security checking, as well as access, consistent for any and all data.

Finally, imbedded in the system code are facilities for monitoring program execution for debugging purposes, as well as timing of operations for system tuning. There is also an ability to log all requests made in the DML and the DDL, used mainly for determining optimal data base structure. These facilities may be turned on or off with the DML.

5.4 Detailed Examples of User Capabilities in GMIS

This section describes and illustrates the use of the three very high levels of GMIS:

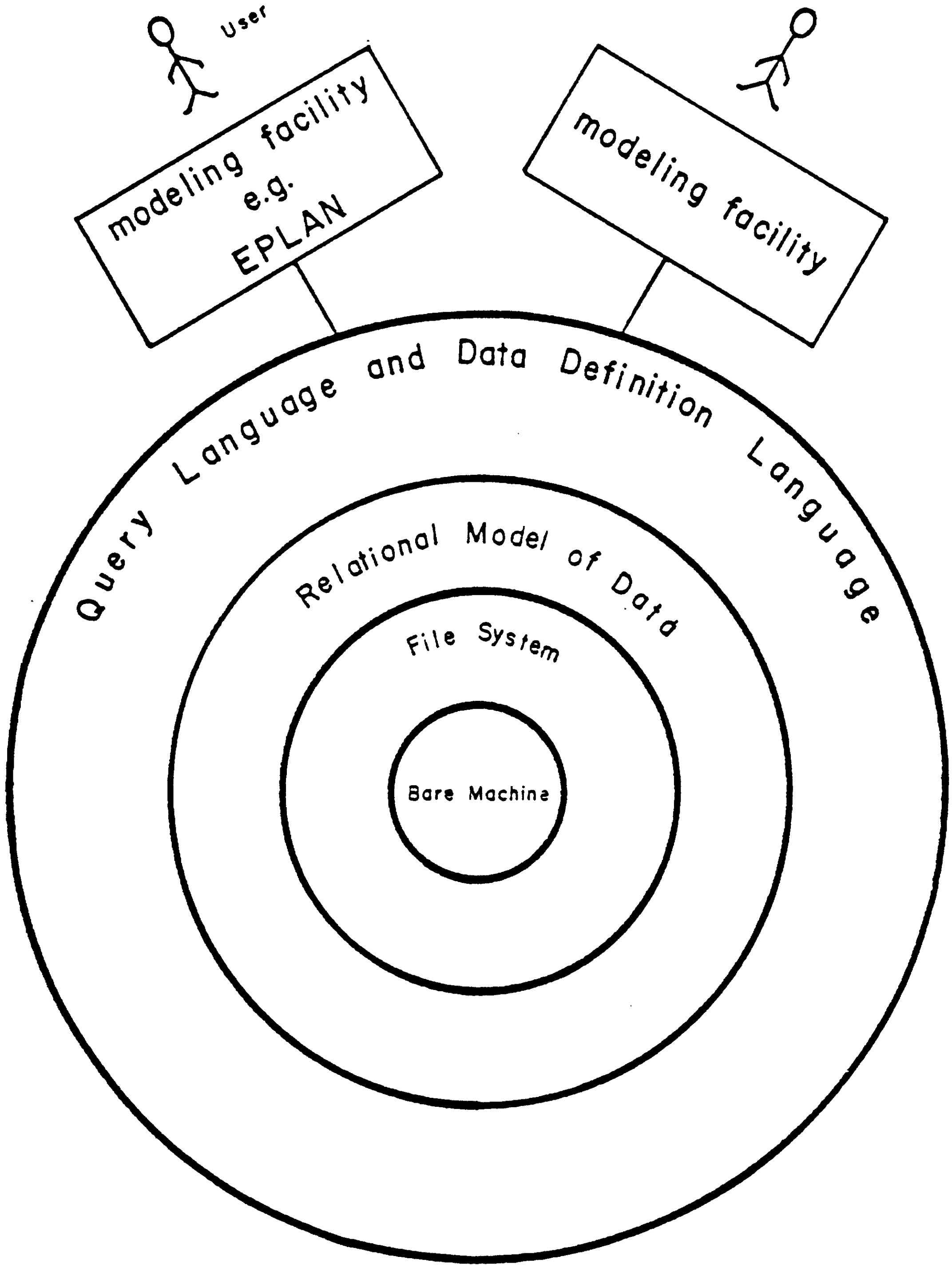


FIGURE 5.2

A HEIRARCHIALLY - STRUCTURED GENERALIZED MANAGEMENT INFORMATION SYSTEM

1. TRANSACT. This is our DDL and DML. It can be used to:
   a) restructure the data;
   b) input the data; and
   c) query data.
2. APL. This facility is a general programming language that is particularly suited to performing mathematical calculations interactively.
3. APL/EPLAN. This is an economic modeling package.

It is important to note that because of the hierarchical approach and our use of VM, we are not limited to these specific facilities. That is, on the same level as the APL/EPLAN facility, we can place a TROLL, TSP, PL/1, FORTRAN, or any user language or package that is now available on an IBM 370.

5.4.1 TRANSACT

TRANSACT encompasses IBM's SEQUEL commands [4] as well as the additional commands presented in the Appendix. TRANSACT also extends SEQUEL's capabilities in several ways to help make it operational for real applications. For example:

1) SEQUEL only interfaced to one terminal -- 3270. We have interfaced TRANSACT to any terminal.
2) We have added a graphic facility.
3) Additional commands have been added.

4) Interface to the modeling facility has been provided.

5) Security has been upgraded.

6) Additional data types are allowed.

7) There is no restriction on number cf entries in tables.

8) Performance measurement facilities are provided.

5.4.1.1 TRANSACT Data Definition

We first (Figure 5.3) give an example of creating a table, inserting data into it by single entries. We view all data as stored in tables. The tables have columns. Entries in columns come from sets of elements. These sets are called domains. In Figure 5.3 we first define two sets or domains. Both sets contain numbers. The first set contains numbers that are interpreted as dates. The second contains numbers, which we will interpret a- numbers. We then define a tab le that consist of five columns. The first column is labelled DATE and entries in this column will be classified as belonging to the set (or domain) DATETYPE. Similarly, four other columns are defined.

Suppose we had a great deal of data to load. Inputting it via the console as in the previous figure would be prohibitively slow and costly.

Thus we have implemented a bulk loading facility as discussed elsewhere [2]. Figure 5.4 demonstrates a series of data "cards" and the appropriate header cards for input into the bulk loader. The bulk loader will accept this data and place it into the appropriate tables.

```
transact
TRANSACT...VERSION 1.0 30JAN75

READY;

define domain datetype (num);
DOMAIN DEFINITION WAS SUCCESSFUL.

READY;

define domain numeric (num);
DOMAIN DEFINITION WAS SUCCESSFUL.

READY;

define table gasoline
   date (datetype),
   demand (numeric),
   produce (numeric),
   import (numeric),
   stocks (numeric)
key is date;
TABLE DEFINITION WAS SUCCESSFUL.

READY;

insert into gasoline (date, demand, produce, import,
   stocks): <720100, 5540, 6151, 51, 239635>;
INSERTION WAS SUCCESSFUL.

READY;

quit;
R;
```

DEFINES 2 DOMAINS

DEFINES A TABLE WITH 5 COLUMNS

INSERTS DATA INTO NEWLY DEFINED TABLE

Figure 5.3

Example of Table Creation and Data Entry

```
type gasoline data

$LOADTAB GASOLINE
         DATE        1        1        1        6
         DEMAND      1        8        1        11
         IMPORT      1        13       1        15
         PRODUCE     1        17       1        20
         STOCKS      1        22       1        27
$ENDCOL
   721200 6378 069 6424 212770
   730100 6118 059 6341 221823
   730200 6437 095 6141 216367
   730300 6513 071 6150 207581
   730400 6541 063 6377 204708
   730500 6907 102 6714 202081
   730600 6364 174 6993 208374
   730700 7023 133 6986 211488
   730800 7249 157 6830 205122
   730900 6581 127 6820 210278
   731000 6677 194 6621 214525
   731100 6823 210 6375 207343
   731200 6223 182 6099 209395
   740100 5804 163 5900 217463
   740200 6100 184 5969 219058
   740300 6162 225 5982 220307
   740400 6457 260 6311 223752
   740500 6406 228 6301 229878
   740600 6895 145 6642 220652
   740700 6941 122 6335 227195
   740800 6849 192 6776 231015
   740900 6656 140 6431 229972
   $ENDLOAD
   $ENDINP

   R;
```

FORMAT CARDS DESCRIBING THE DATA

DATA

COMMAND TO LOAD DATA

```
load gasoline
STAND-ALONE PHASE/2 DATA BASE LOADER
BEGIN PROCESSING LOAD DATA FOR TABLE GASOLINE
INPUT ROW         1 ALREADY PRESENT IN TABLE
PROCESSING CONTINUES WITH NEXT ROW
        32 ROWS HAVE BEEN ADDED TO TABLE
END-OF-INPUT ENCOUNTERED
PROCESSING COMPLETED SUCCESSFULLY
R;
```

Figure 5.4

Example of Bulk Loader Use

5.4.1.2 TRANSACT System Commands

This level has a number of "system commands" for inquiring about tables as opposed to their contents. For example,

- list all tables that have been created
- list all domains
- list information about one table.

Fiugres 5.5 and 5.6 illustrate these commands.

5.4.1.3 TRANSACT Query and Update Facility

Figures 5.7 and 5.8 exhibit queries to tables that have been created -- all queries start with the word SELECT.

Figure 5.9 illustrates the use of the insert, update, and delete commands.

Figures 5.10 and 5.11 show how the SELECT command can be used for specifying queries that require data from more than one relation.

```
list tables;
```

LIST ALL TABLES WE HAVE CREATED

```
                         LIST OF TABLES
                         ---- -- ------

TERMINAL  SUPRVISR  COMPANY   RECEIVE   SHIPPING  COMMENT
CAPACITY  INVNTORY  TRANSACT  ALLOCATE  VESSEL    VESDESC
VESOWNER  VESPORT   GASOLINE  CARS

READY;

describe table gasoline;

DESCRIPTION OF TABLE GASOLINE
    NAME       DOMAIN     TYPE  C  KEY  INV
    ----       ------     ----  -  ---  ---

    DATE       DATETYPE   NUM   0  YES  NO
    DEMAND     NUMERIC    NUM   0  NO   NO
    PRODUCE    NUMERIC    NUM   0  NO   NO
    IMPORT     NUMERIC    NUM   0  NO   NO
    STOCKS     NUMERIC    NUM   0  NO   NO

READY;
```

LIST SYSTEM INFORMATION ABOUT THE TABLE GASOLINE

Figure 5.5
Examples of Inquiries of Tables

```
list domains;  <------  LIST ALL DOMAINS WE HAVE CREATED

LIST OF DOMAINS:  NAME        TYPE   USE   LLF
                  ----        ----   ---   ---

                  ENTRY       CHAR    4      1
                  FUELNUM     NUM     4   ----
                  CONF        NUM     2   ----
                  AMOUNT      NUM    11   ----
                  FRACTION    CHAR    1      4
                  AGG         CHAR    3     15
                  DATE        NUM     4   ----
                  TYPE        CHAR    1     15
                  CONSTRCT    CHAR    1      4
                  COUNTRY     CHAR    2      7
                  TONNAGE     NUM     2   ----
                  LENGTH      NUM     4   ----
                  HPR         NUM     1   ----
                  HFP         CHAR    1      0
                  HFPTYPE     CHAR    1      0
                  YEAR        NUM     1   ----
                  PASSNGRS    NUM     1   ----
                  EQUIP       CHAR    1      0
                  ENGNO       NUM     1   ----
                  ENGNAME     CHAR    1     13
                  PORTDESC    CHAR    1     48
                  DATETYPE    NUM     1   ----
                  NUMERIC     NUM     4   ----
                  ALPHA       CHAR    0     15
                  CODE        CHAR    0      7
                  TIME        NUM     0   ----
                  C           CHAR    0     33
                  CODES       NUM     0   ----
                  UNITS       CHAR    0      0
                  CARDTYPE    CHAR    0      0
                  FOREAMNT    NUM     0   ----
                  FOREUNIT    CHAR    0      0
                  FOREFUEL    CHAR    0      0
                  FOREDATE    NUM     0   ----
                  FCDTYPE     CHAR    0      0

READY;
```

Figure 5.6

Examples of Inquiries of Tables (cont'd)

```
select * from gasoline;
```

QUERY

DATE	DEMAND	PRODUCE	IMPORT	STOCKS
720100	5549	6151	51	239633
720200	5710	5989	66	249927
720300	6412	5913	67	236831
720400	6283	5833	52	225153
720500	6445	6023	74	214736
720600	6822	6244	75	200143
720700	6673	6612	69	200719
720800	6938	6588	81	192706
720900	6453	6605	70	199690
721000	6350	6532	71	207776
.21100	6479	6436	69	208930
721200	6378	6424	69	212770
730100	6118	6341	59	221823
730200	6437	6141	95	216367
730300	6513	6150	71	207581
730400	6541	6377	63	204708
730500	6907	6714	102	202081
730600	6964	6993	174	208374
730700	7023	6986	133	211488
730800	7249	6880	157	205122
730900	6581	6620	127	210278
731000	6677	6621	194	214525
731100	6823	6375	216	207343
731200	6223	6099	188	209395
740100	5804	5900	163	217463
740200	6100	5969	184	219058
740300	6162	5982	225	220307
740400	6457	6311	260	223752
740500	6406	6301	228	229878
740600	6895	6642	145	226652
740700	6941	6835	122	227195
740800	6849	6776	192	231015
740900	6656	6481	140	229972

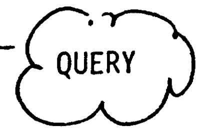
ANSWER

EADY:

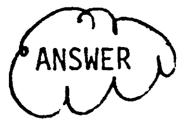
Figure 5.7

Sample Query of a Table

```
select date,demand from gasoline where date >= 740100;
```

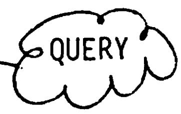


DATE	DEMAND
740100	5804
740200	6100
740300	6162
740400	6457
740500	6406
740600	6835
740700	6941
740800	6349
740900	6656



READY;

```
select date,demand,produce from gasoline
where demand < produce;
```



DATE	DEMAND	PRODUCE
720100	5549	6151
720200	5710	5989
720900	6453	6605
721000	6350	6532
721200	6378	6424
730100	6118	6341
730600	6964	6993
730900	6581	6620
740100	5804	5900

READY;

Figure 5.8

Sample Table Queries

```
select * from gasoline
where demand < produce
and stocks > 220000;

      DATE     DEMAND    PRODUCE     IMPORT     STOCKS
      ----     ------    -------     ------     ------

    720100       5549       6151         51     239633
    720200       5710       5983         66     249927
    730100       6118       6341         59     221823

READY;

select avg(demand) from gasoline where date >= 740100;
THE RESULT OF YOUR QUERY IS:
          6474

READY;
```

Figure 5.8 (cont'd)

Sample Table Queries

```
insert into gasoline (date, demand, produce, import,
   stocks): <741000, 6700, 6500, 150, 228422>;
INSERTION WAS SUCCESSFUL.

READY;

update gasoline set demand=6859 where date=740800;
UPDATE WAS SUCCESSFUL.

READY;

select * from gasoline where date >= 740700;

      DATE     DEMAND    PRODUCE     IMPORT     STOCKS
      ----     ------    -------     ------     ------

    740700      6941       6835        122     227195
    740800      6859       6776        192     231015
    740900      6656       6481        140     229972
    741000      6700       6500        150     228422

READY;

update gasoline set demand=6849 where date=740800;
UPDATE WAS SUCCESSFUL.

READY;

delete gasoline where date=741000;
DELETION WAS SUCCESSFUL.

READY;

select * from gasoline where date >= 740700;

      DATE     DEMAND    PRODUCE     IMPORT     STOCKS
      ----     ------    -------     ------     ------

    740700      6941       6835        122     227195
    740800      6849       6776        192     231015
    740900      6656       6481        140     229972

READY;
```

Figure 5.9

Illustration of Insert, Update, and Delete Commands

```
select date, sales from cars where date in
   select date from gasoline where produce > demand;;

      DATE      SALES
      ----      -----

    720100        610
    720200        698
    720900        741
    721000        932
    721200        719
    730100        736
    730600        909
    730900        754
    740100        551

READY;
```

Figure 5.10

Sample Use of Select Command

```
select date, demand, produce, import, carsales
from row in gasoline
compute carsales =
        select value(sales)
        from cars
        where date = row.date;;
```

DATE	DEMAND	PRODUCE	IMPORT	CARSALES
720100	5549	6151	51	610
720200	5710	5939	66	698
720300	6412	5913	67	772
720400	6283	5833	52	774
720500	6445	6023	74	888
720600	6822	6244	75	877
720700	6673	6612	69	769
720800	6938	6582	81	656
720900	6453	6605	70	741
721000	6350	6532	71	932
721100	6479	6436	69	391
721200	6378	6424	69	719
730100	6118	6341	59	736
730200	6437	6141	95	775
730300	6513	6150	71	964
730400	6541	6377	63	863
730500	6907	6714	102	972
730600	6964	6993	174	909
730700	7023	6986	133	808
730800	7249	6880	157	686
730900	6531	6620	127	754
731000	6677	6621	194	858
731100	6823	6375	216	778
731200	6223	6093	138	574
740100	5804	5900	163	551
740200	6100	5969	184	568
740300	6162	5982	225	654
740400	6457	6311	260	703
740500	6406	6301	228	767
740600	6895	6642	145	698
740700	6941	6835	122	691
740800	6849	6776	192	668
740900	6656	6481	140	591

READY;

Figure 5.11

Sample Use of Select Demand (cont'd)

```
select date, demand, carsales from row in gasoline
where produce > demand and date >= 721100
compute carsales =
        select value(sales) from cars
        where date = row.date;;
```

DATE	DEMAND	CARSALES
721200	6378	719
730100	6118	736
730600	6964	909
730900	6531	754
740100	5804	551

READY;

Figure 5.11 (cont'd)

Sample Use of Select Demand

5.4.2 APL -- Stand-Alone

APL [21], [23], and [28] is an extremely powerful and concise tool for performing mathematical calculations interactively. It is especially suited for working with vectors and matrices and has a number of operators for manipulating such data types. We have structured the GMIS System (Figure 5.2) such that APL runs in a separate virtual machine (one for each user of APL). The APL facility can thus be used independently from the data base facility, TRANSACT, or an APL user may request data from TRANSACT. In this section we shall give examples of using APL as a stand-alone facility, i.e., using APL witout envoking any levels below it.

The references listed above present a thorough description of APL. On the top line of Figure 5.12 we present an example of how APL can be used in a mode similar to that of a desk calculator. When

2 + 3

is typed by the user, the system responds by typing the resulting value. (Lines entered by the user are indented.) In the following statement

A ← 2 + 3

the resulting value is assigned to the variable A rather than printed out. The command

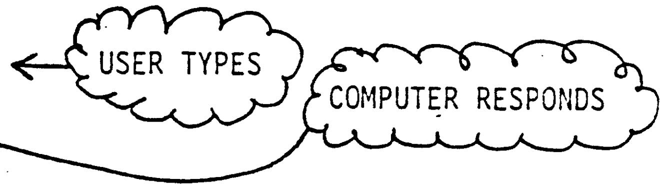
B ← ⍳12

demonstrates how a variable can be set equal to an entire vector, a list or numbers, in this case the first 12 integers. The statement

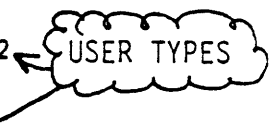
+/B

computes the sum of the elements of B.

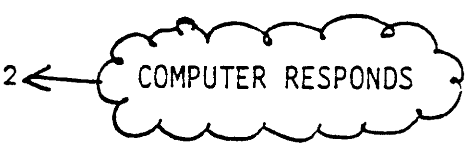
```
      2
2
      2+3
5
```




```
      A←2+3

      A
5

      B←ι12
```




```
      B
1 2 3 4 5 6 7 8 9 10 11 12
```




```
      +/B
78

      !B
1 2 6 24 120 720 5040 40320 362880 3628800 39916800
479001600

      ⎕←C←12?100
87 25 55 47 95 79 33 94 7 62 49 17

      +/C
650

      C[⍋C]
7 17 25 33 47 49 55 62 79 87 94 95

      C[⍒C]
95 94 87 79 62 55 49 47 33 25 17 7
```

Figure 5.12

APL Examples

```
      10+C
97 35 65 57 105 89 43 104 17 72 59 27

      B+C
88 27 58 51 100 85 40 102 16 72 60 29

      B×C
87 50 165 188 475 474 231 752 63 620 539 204

      MATRIX← 3 4 ρ B+C

      MATRIX
 88   27   58   51
100   85   40  102
 16   72   60   29

      SQUARE_MATRIX← 1 0 1 1/MATRIX

      SQUARE_MATRIX
 88   58   51
100   40  102
 16   60   29

      ⌹SQUARE_MATRIX
 0.02095798263     ¯0.005822600818    ¯0.01637764933
¯0.005357806849    ¯0.007335293919     0.01637764933
¯0.02264814251      0.01838893959      0.009633911368
```

Figure 5.12 (cont'd)

APL Examples

!B

computes the factorial function at each element of B. (The factorial of 6 is defined as 6 • 5 • 4 • 3 • 2 • 1 = 720.)

In the statement

⎕ ← C ← 12?100

C is assigned a vector of 12 numbers between 1 and 100, chosen randomly. (The ⎕ ← specifies that the result is also to be printed.)

In succeeding lines of Figures 5.12 we show how we can: sort the vector C, in ascending or descending order; add 10 to each number in C; compute the element-by-element sum and product of vectors B and C; convert the vector B + C into a matrix (named MATRIX); produce SQUARE-MATRIX by deleting a column from MATRIX; and invert SQUARE-MATRIX.

The programming capability of APL is illustrated in Figure 5.13, where we envoke a previously defined (not shown) function, PRIMES-UP-TO, which computes the prime numbers up to any specified argument.

```
      PRIMES_UP_TO 10
2 3 5 7

      PRIMES_UP_TO 50
2 3 5 7 11 13 17 19 23 29 31 37 41 43 47

      PRIMES_UP_TO 100
2 3 5 7 11 13 17 19 23 29 31 37 41 43 47 53 59 61 67 71 73
79 83 89 97
```

Figure 5.13

Example of an APL Function

5.4.3 Modeling Level

We have developed an interface between a modeling facility (EPLAN) and the TRANSACT data base facility.[1] The modeling facility appears to the user as a level (Figure 5.2). At this modeling level is the full power of APL and EPLAN, as well as all the functions we have added for conversion, interface to the data base facility, and report generation. These added functions consist mainly of programs for transmitting TRANSACT commands to and receiving data from the data base management component of the system, and programs for converting the data into APL or EPLAN format.

EPLAN [33] is a set of APL functions that provide one with econometric modeling and forecasting tools. It includes functions for analyzing and manipulating data; tabulating or plotting data series; defining models; estimating parameters via regression; and calculating solutions to models.

An advantage of EPLAN over some other modeling systems is the ease with which the user can define and use analytical functions not provided for within EPLAN. Because EPLAN is used within the normal APL environment, an EPLAN user who is familiar with APL can enter APL statements or APL function definitions that cause the execution or definition of any analytical methods not contained within EPLAN; furthermore, any such user-defined APL functions may quite simply be added to the user's version of EPLAN so that the new functions become a regular part of that user's EPLAN system. (Alternately, they may be saved in a user's library).

---

[1] We could have and are implementing interfaces with other modeling facilities, e.g., TROLL, TSP, PL/1, etc.

5.4.3.1 Stand-Alone EPLAN

Of course,one may use the EPLAN/APL facility as a stand-alone level. That is, one need not call for data at the TRANSACT level. The following are examples of this use. It should be noted that words containing underscored characters are names of EPLAN functions.

The first two input lines in Figure 5.14

SERIES1 ← 12 1973 1 D F . . .

SERIES2 ← 12 1973 1 D F . . .

define SERIES1 and SERIES2 as monthly time series of data, beginning with January 1973. The next two lines of Figure 5.14 that the user inputted are:

SERIES1

SERIES2

which caused APL to print out the values of SERIES1 and SERIES2. (The first element of each vector is a special coding used by EPLAN.)

The line in Figure 5.14

□ ← SERIES3 ← SERIES1 D SERIES2

defines a now time series, SERIES3, as SERIES1 divided by SERIES2. The □ causes that series to be printed.

The statement

7 6 9 2 TABULATE 'SERIES1, SERIES2, SERIES3'

envokes the EPLAN tabulate function, which prints a table of values of SERIES1, SERIES2, and SERIES3. (The 7, 6, 9, 2 specify the width of the printout, number of positions, and number of decimal positions for each entry.)

In Figure 5.15 we plot SERIES1 and SERIES2 by envoking the EPLAN plot function. The 35, 54, 4 to the left of PLOT specifies the size of the plot (vertical and horizontal), and the horizontal indent.

```
      SERIES1← 12 1973 1 DF 5000 5100 5300 5200 5100
         4900 4700 4600 4500 4600 4700 4900

      SERIES2← 12 1973 1 DF 2500 2600 2700 2700 2600
         2500 2400 2400 2300 2300 2200 2100
```

Define a timer series

```
      SERIES1
2367712 5000 5100 5300 5200 5100 4900 4700 4600 4500 4600
4700 4900
```

APL - print series

```
      SERIES2
2367712 2500 2600 2700 2700 2600 2500 2400 2400 2300 2300
2200 2100

      ⎕←SERIES3←SERIES1 D SERIES2
2367712 2 1.961538462 1.962962963 1.925925926 1.961538462
1.96 1.958333333 1.916666667 1.956521739 2 2.136363636
2.333333333

      7 6 9 2 TABULATE 'SERIES1,SERIES2,SERIES3'
```

EPLAN function to print the table below

\# of rows

decimal point

	1973 1	1973 2	1973 3	1973 4	1973 5	1973 6
*SERIES*1	5000.00	5100.00	5300.00	5200.00	5100.00	4900.00
*SERIES*2	2500.00	2600.00	2700.00	2700.00	2600.00	2500.00
*SERIES*3	2.00	1.96	1.96	1.93	1.96	1.96

	1973 7	1973 8	1973 9	1973 10	1973 11	1973 12
*SERIES*1	4700.00	4600.00	4500.00	4600.00	4700.00	4900.00
*SERIES*2	2400.00	2400.00	2300.00	2300.00	2200.00	2100.00
*SERIES*3	1.96	1.92	1.96	2.00	2.14	2.33

Figure 5.14

Example of EPLAN Stand-Alone

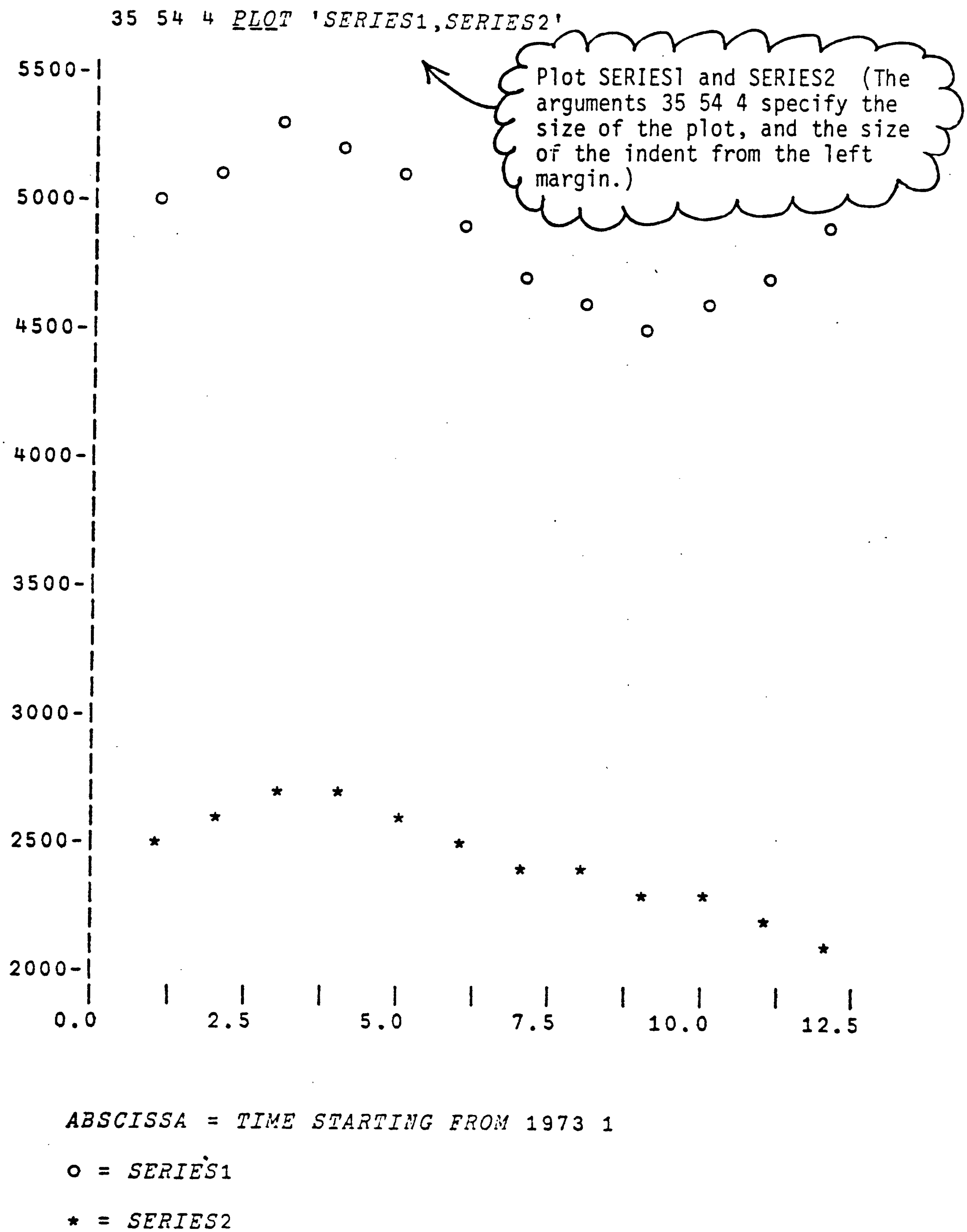


Figure 5.15

EPLAN Ploting Function

EPLAN also has functions for performing correlations and regressions on time series. For example, Figure 5.16 demonstrates an EPLAN correlation function applied to the two previously defined series. Following the correlation request we have used one of EPLAN's regression functions to find out how well the data fits the following equation:

$$\text{SERIES2}(t) = a_1 + a_2 \cdot \text{SERIES1}(t) + a_3 \cdot \text{SEIRES2}(t-1)$$

and what the best coefficient values are.

```
    COR 'SERIES1,SERIES2'
   1                  0.7518306849
   0.7518306849       1

    'SERIES2' REGRESS '1,SERIES1,(1 LAG SERIES2 )'
WITH:

COEF/VALUE/ST ERR/T-STAT.....

 1   ¯705.79531  448.78291    ¯1.57269
 2      0.18619    0.11965     1.55610
 3      0.90382    0.19413     4.65587

NO OF VARIABLES.........          2.00000
NO OF OBSERVATIONS......         11.00000
SS DUE TO REGRESSION....     357513.73164
SS DUE TO RESIDUALS.....      47940.81381
F-STATISTIC.............         29.82959
STANDARD ERROR..........         77.41190
R*2 -STATISTIC..........          0.88176
R*2 CORRECTED...........          0.86862
DURBIN WATSON STATISTIC.          1.12600
SERIES2← ( ¯705.79531442664 T 1 ) P ( 0.1861898890259 T SE
RIES1 ) P ( 0.90382244143033 T (1 LAG SERIES2))
```

Calculate simple correlation coefficients between SERIES1, SERIES 2

regress SERIES2 (dependent variable) against independent variables 1, SERIES1, and SERIES2 lagged 1 time unit, i.e., in

$$SERIES2(t) = a_1 \cdot 1 + a_2 \cdot SERIES1(t) + a_3 \cdot SERIES2(t-1)$$

use regression to estimate a_1 a_2 a_3.

blank response by user to WITH: specifies use of ordinary least squares method

Figure 5.16

EPLAN Correlation and Regression

5.4.3.2 Interaction between APL/EPLAN Level and TRANSACT

Let us begin with an example using the modeling facility to perform some analytical functions on a data series. However, that data series is stored not in the APL/EPLAN facility but in the general data base system, TRANSACT. Thus from the modeling facility we must request TRANSACT for the desired data. The reader may keep in mind Figure 5.2.

Figure 5.17 depicts a user interaction using the modeling facility to request data from TRANSACT and to print that data. Note that a user of TRANSACT views data as stored in relations (tables), specifically, the user command (Figure 5.17) at the modeling level.

SEQPUT 'SELECT * FROM GASOLINE;'

results in the modleing facility forwarding the SELECT command down to the TRANSACT facility. The TRANSACT facility returns the desired data in a file.

The user command (Figure 5.17)

⎕ ← Q1 ← SELECT 'GASOLINE'

reads that data from the file and assigns the data to variables of the same name as the columns of the relation. Q1 is a vector of those variables, and ⎕ prints that list.

The user commands (Figure 5.17)

DATE

DEMAND

STOCKS

result in the printing of the values of those vectors.

The user command DEMO (Figure 5.17) envokes an APL function we have written to print the table GASOLINE.

Get data from GASOLINE relation in data base

```
      SEQPUT 'SELECT * FROM GASOLINE ;'
```

Put data into APL variables

```
      ⎕←Q1←SEQGET 'GASOLINE'
DATE
DEMAND
PRODUCE
IMPORT
STOCKS
```

Display value of APL variable DATE

```
      DATE
720100 720200 720300 720400 720500 720600 720700 720800
720900 721000 721100 721200 730100 730200 730300 730400
730500 730600 730700 730800 730900 731000 731100 731200
740100 740200 740300 740400 740500 740600 740700 740800
740900
```

Display value of APL variable DEMAND

```
      DEMAND
5549 5710 6412 6283 6445 6822 6673 6938 6453 6350 6479 6378
6118 6437 6513 6541 6907 6964 7023 7249 6581 6677 6823 6223
5804 6100 6162 6457 6406 6895 6941 6849 6656
```

Display value of APL variable STOCKS

```
      STOCKS
239633 249927 236831 225153 214736 200143 200710 192706
199690 207776 208930 212770 221823 216367 207581 204708
202081 208374 211488 205122 210278 214525 207343 209395
217463 219058 220307 223752 229878 225652 227195 231015
229972
```

Figure 5.17

Example of EPLAN and TRANSACT Interface

List the data

DEMO

DATE	*DEMAND*	*PRODUCE*	*IMPORT*	*STOCKS*
01/72	5549	6151	51	239633
02/72	5710	5989	66	249927
03/72	6412	5913	67	236831
04/72	6283	5833	52	225153
05/72	6445	6023	74	214736
06/72	6822	6244	75	200143
07/72	6673	6612	69	200710
08/72	6938	6588	81	192706
09/72	6453	6605	70	199690
10/72	6350	6532	71	207776
11/72	6479	6436	69	208930
12/72	6378	6424	69	212770
01/73	6118	6341	59	221823
02/73	6437	6141	95	216367
03/73	6513	6150	71	207581
04/73	6541	6377	63	204708
05/73	6907	6714	102	202081
06/73	6964	6993	174	208374
07/73	7023	6986	133	211488
08/73	7249	6880	157	205122
09/73	6581	6620	127	210278
10/73	6677	6621	194	214525
11/73	6823	6375	216	207343
12/73	6223	6099	188	209395
01/74	5804	5900	163	217463
02/74	6100	5969	184	219058
03/74	6162	5982	225	220307
04/74	6457	6311	260	223752
05/74	6406	6301	228	229878
06/74	6895	6642	145	226652
07/74	6941	6835	122	227195
08/74	6849	6776	192	231015
09/74	6656	6481	140	229972

TSERIES Q1

Put the data into EPLAN time series

DATE
DEMAND
PRODUCE
IMPORT
STOCKS

Figure 5.17 (cont'd)

Example of EPLAN and TRANSACT Interface

5.4.3.3 PLOTTING

Let us now plot the data requested from TRANSACT. Before we envoke the EPLAN plot function, we must change the format of the data from what TRANSACT gave to EPLAN to a format the EPLAN functions deal with.

The user command 'TSERIES Q1' of Figure 5.18 envokes an APL function we have written to convert the data series Q1 from a table format to a time series format. The user command

20 40 10 PLOT 'STOCKS'

envokes the EPLAN function, which takes the time series STOCKS and produces the plot of Figure 5.18.

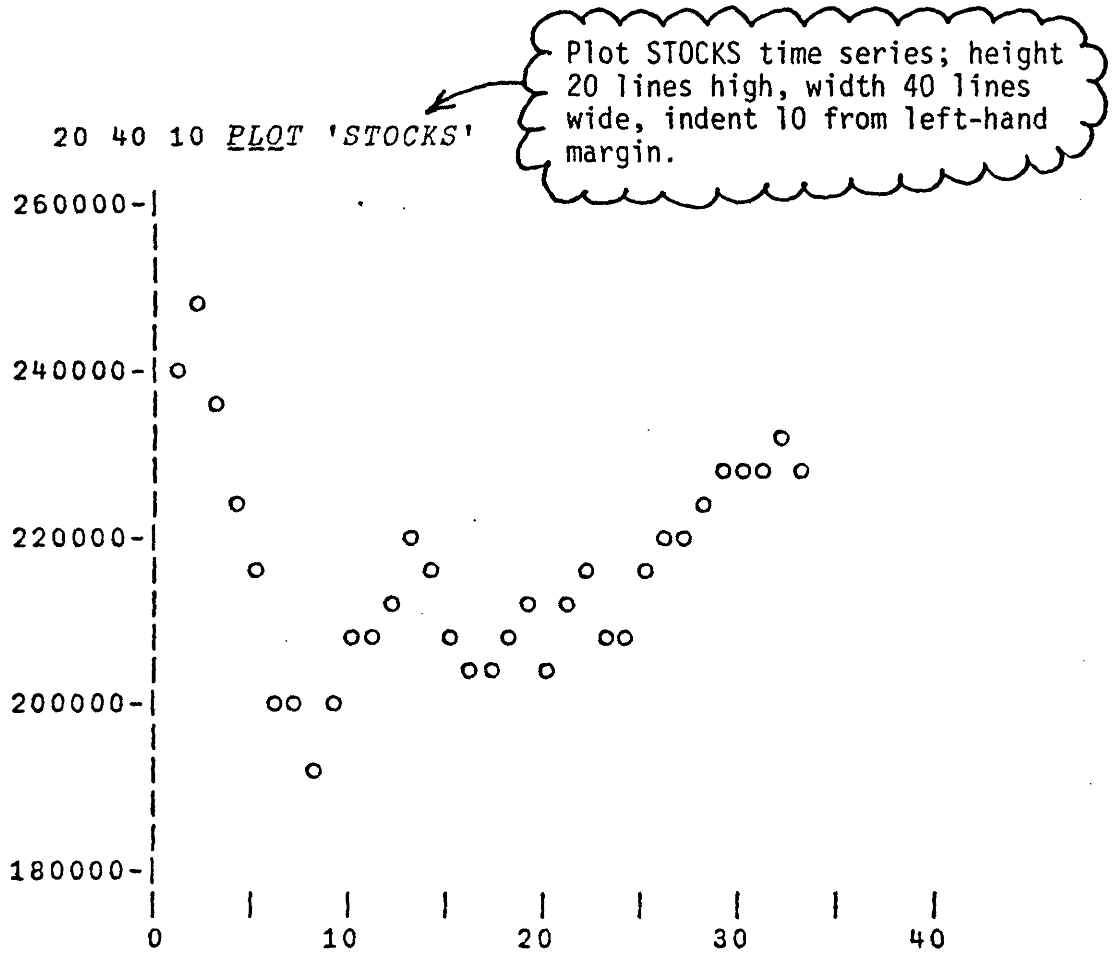
Suppose we wished to plot some function of the data. We could, for example, write a function such as in Figure 5.19, where the user command.

STOCK_FLOW_RATIO STOCKS D DEMAND

defines a new time series named 'STOCK_FLOW_RATIO' to be equal to the series STOCKS divided by the series DEMAND. The next line plots the new series.

We can also perform operations in the plot function. The first statement of Fgiure 5.20 divides all the numbers in the SALES domain by corresponding numbers of DAYS, and assigns the list of numbers to the vector named SALES_RATE. The second command uses this vector and plots a function as indicated.

The use of these facilities to construct energy indicators is demonstrated with the indicator of gasoline consumption shown in Section 3.3.2.



ABSCISSA = TIME STARTING FROM 1972 1

o = *STOCKS*

Figure 5.18

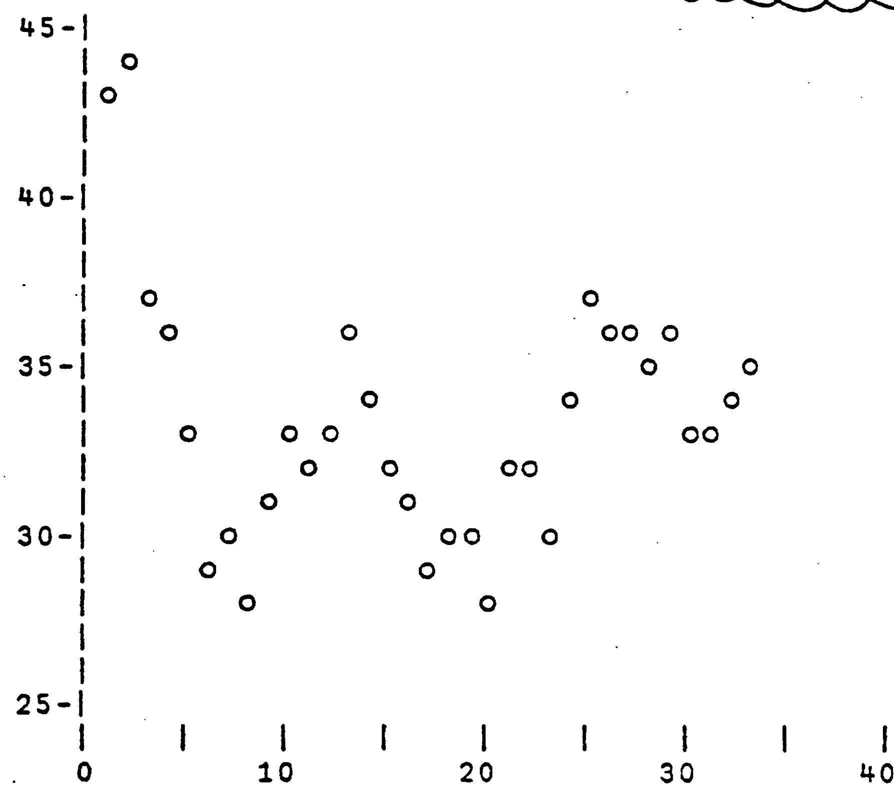
Example of EPLAN Plotting

```
STOCK_FLOW_RATIO←STOCKS D DEMAND
```

Define time series STOCK_FLOW_RATIO as STOCKS divided by DEMAND.

```
20 40 10 PLOT 'STOCK_FLOW_RATIO'
```

Plot STOCK_FLOW_RATIO.

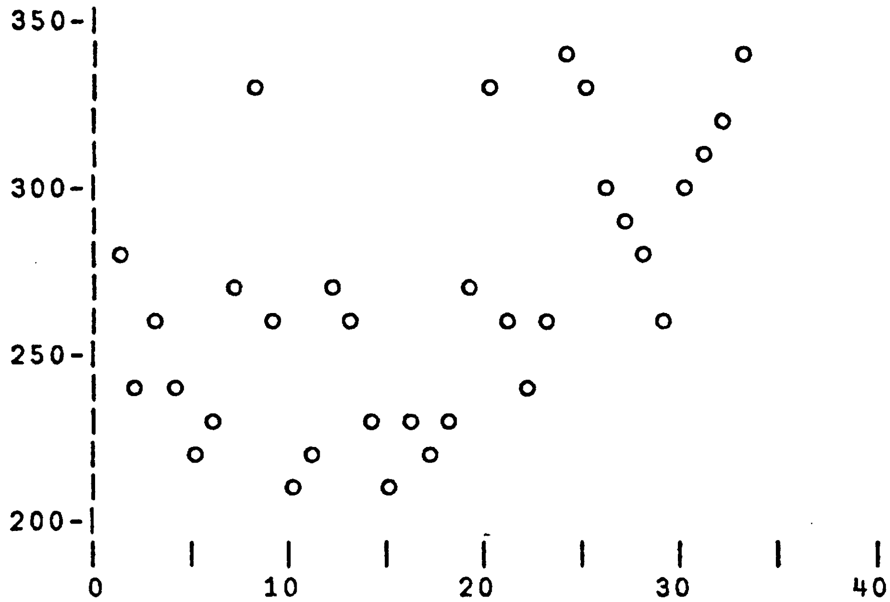



```
ABSCISSA = TIME STARTING FROM 1972 1

o = STOCK_FLOW_RATIO
```

Figure 5.19

PLOT of a Function

```
SALES_RATE←SALES D DAYS

20 40 10 PLOT 'DEMAND D SALES_RATE'
```




```
ABSCISSA = TIME STARTING FROM 1972 1

o = DEMAND D SALES_RATE
```

Figure 5.20

Plot of a Function

5.4.3.4 More Advanced Examples

In this section we present an example of a session that a researcher might have in creating a model. Here we perform several regressions in order to ascertain to what extent the data we have fits some possible forecasting equations for forecasting gasoline demand. In Figure 5.21 we first test an equation of the form:

$$DEMAND(t) = a_1 + a_2 \quad DEMAND(t-1).$$

The system responds by outputting information indicating the goodness of fit.

Next, in Figure 5.22 we regress on the equation:

$$DEMAND(t) = a_1 + a_2 \cdot DEMAND(t-1) + a_3 \cdot DEMAND(t-2)$$

This equation might be proposed on the basis that there is a seasonal component that can be predicted from demand of a year ago.

Finally, we suppose that the following forecasting equation, which suggests that recent gasoline stocks and the recent rate of sales of new cars are useful predictions of gasoline demand, has been suggested:

$$DEMAND(t) \; a_1 + a_2 \cdot DEMAND(t-1) + a_3 \cdot DEMAND(t-12)$$
$$SALES + a_4 \cdot STOCKS(t-1) + a_5$$

The resulting regression coefficients and statistics are shown in Figure 5.23.

```
      'DEMAND' REGRESS '1,(1 LAG DEMAND)'
WITH:

COEF/VALUE/ST ERR/T-STAT.....

 1    2626.45866    792.41839       3.31448
 2       0.60156      0.12160       4.94700

NO OF VARIABLES.........            1.00000
NO OF OBSERVATIONS......           32.00000
SS DUE TO REGRESSION....     1729353.64778
SS DUE TO RESIDUALS.....     2119925.57097
F-STATISTIC.............           24.47284
STANDARD ERROR..........          265.82736
R*2 -STATISTIC..........            0.44927
R*2 CORRECTED...........            0.44927
DURBIN WATSON STATISTIC.            1.99660
DEMAND← ( 2626.4586562221 T 1 ) P ( 0.60156187489019 T (1
LAG DEMAND))
```

Regress DEMAND(t) = $a_1 + a_2 \cdot$ DEMAND(t-1)

values of a_1, a_2

the best fit equation

Figure 5.21

Possible Forecasting Demand

```
      'DEMAND' REGRESS '1,(1 LAG DEMAND),(12 LAG DEMAND)'
WITH:

COEF/VALUE/ST ERR/T-STAT.....

 1    1123.99636  1137.72892       0.98793
 2       0.47307     0.17426       2.71476
 3       0.36116     0.15796       2.28641

NO OF VARIABLES.........           2.00000
NO OF OBSERVATIONS......          21.00000
SS DUE TO REGRESSION....   1514950.49299
SS DUE TO RESIDUALS.....   1178428.45940
F-STATISTIC.............          11.57012
STANDARD ERROR..........         255.86764
R*2 -STATISTIC..........           0.56247
R*2 CORRECTED...........           0.53944
DURBIN WATSON STATISTIC.           1.73723
DEMAND← ( 1123.9963631427 T 1 ) P ( 0.47306541255257 T (1
LAG DEMAND) ) P ( 0.36116457008549 T (12 LAG DEMAND))
```

Regress

$$DEMAND(t) = a_1 + a_2 \cdot DEMAND(t-1) + a_3 \cdot DEMAND(t-12)$$

Figure 5.22

Seasonal Component of Demand

```
      SEQPUT 'SELECT * FROM CARS;'
```

Select all data from relation CARS

```
      ⎕←Q2←SEQGET 'CARS'
DATE
SALES
```

Put data into APL variables DATE, SALES.

```
      TSERIES Q2
DATE
SALES
```

Put data into EPLAN time series format

```
      SALES_RATE←SALES D DAYS
```

Define SALES_RATE as SALES in months divided by number of DAYS in month

```
      'DEMAND' REGRESS '1,(1 LAG DEMAND),(12 LAG DEMAND),
         (1 LAG STOCKS),(1 LAG SALES_RATE )'
WITH:
```

Regress

$$DEMAND(t) = a_1 \cdot 1 + a_2 \cdot DEMAND(t-1) + a_3 \cdot DEMAND(t-12) + a_4 \cdot STOCKS(t-1) + a_5 \cdot SALES\_RATE(t-1)$$

```
COEF/VALUE/ST ERR/T-STAT.....

 1   ¯967.33254  1689.13269   ¯0.57268
 2      0.31263     0.15970    1.95756
 3      0.30573     0.14441    2.11712
 4      0.01034     0.00678    1.52650
 5     51.98043    14.59529    3.56145

NO OF VARIABLES.........          4.00000
NO OF OBSERVATIONS......         21.00000
SS DUE TO REGRESSION....   2036084.95146
SS DUE TO RESIDUALS.....    657294.00092
F-STATISTIC.............         12.39071
STANDARD ERROR..........        202.68418
R*2 -STATISTIC..........          0.75596
R*2 CORRECTED...........          0.71289
DURBIN WATSON STATISTIC.          2.21316
DEMAND← (¯967.3325379656 T 1 ) P ( 0.3126292938511 T (1 LA
G DEMAND) ) P ( 0.30573483619364 T (12 LAG DEMAND) ) P ( 0
.010343502450714 T (1 LAG STOCKS) ) P ( 51.980430625339 T
(1 LAG SALES_RATE))
```

Figure 5.23

Rate of Sales Factor

5.5 Use of VM

We are using and advocate the use of one mechanism (the virtual machine concept) for the following:

1) A mechanism whereby an environment can be developed that greatly enhances the effectiveness of a public policy researcher.
2) A mechanism for multi-user coordination of access and update to a central data base
3) A mechanism for creating an environment where several different modeling facilities can access the same data base
4) A mechanism for creating an environment where several different and potentially incompatible data management systems can all be accessed by the same user models or facilities
5) A mechanism for increased security and reliability.

Appendix C describes the VM concept as applied to solving the above. In Appendix C we also investigate and formalize the performance implications of this scheme, specifically addressing the response time degradation.

5.5.1 VM Concept

The advent of large scale digital computer systems has been accompanied by development of a multiplicity of operating systems permitting users to access hardware in an efficient manner. The first operating systems were oriented toward processing one job at a time in sequence, dedicating the entire resources of the system to executing that job. The next generation of operating systems permitted true sharing of the system resources between jobs being executed, thus permitting a more optimal utilization of the system. This gain in utilization efficiency was incurred at the cost of some degradation in the time required to execute any one job.

One of the external costs of changing computer systems and/or operating systems involves the conversion of application programming systems. IBM and others are dealing with this problem through the concept of "virtual machines" [3, 16, 25, 29].

The emerging generation of VM operating systems may be viewed as extending timesharing capabilities to include more than one operating system on one physical machine. This concept of "virtual machines" has been developed by IBM to the point of a production system release, VM/370. A virtual machine may be defined as a replica of a real computer system simulated by a combination of a Virtual Machine Monitor (VMM) software program and appropriate hardware support. For example, the VM/370 system enables a single IBM System/370 to appear functionally as if it were multiple independent System/370's (ie.., multiple "virtual machines"). Thus, a VMM can make one computer system function as if it were multiple physically isolated systems. A VMM accomplishes this feat by controlling the multiplexing of the physical hardware resources in a manner analogous to the way that the telephone company multiplexes communications enabling separate and, hopefully, isolated conversations over the same physical communications link.

By restricting itself to the task of multiplexing and allocating the physical hardware, the VMM presents an interface that appears identical to a "bare machine". In fact, it is usually desirable to laod a user-oriented operating system into each virtual machine to provide the functions expected to modern operating systems, such as Job Control language, command processors, data management services, and language processors. Thus, each virtual machine is controlled by a separate, and possibly different, operating system. The feasibility of this solution has been demonstrated on the VM/370 system and the earlier CP-67 and CP-40 systems.

The advantages of the VMM may be summarized as follows:

1. The cost of software conversion is eliminated when changing operating systems if the new system is VM.
2. The application systems programmer may choose the operating system providing the most capability for his particular application problem.
3. New programming systems developed at other centers can be assimilated very rapidly.
4. When software is being upgraded to a new operating system, the applications can be run in parallel until the new system is developed.
5. Use of VM permits communication of data between programs using different operating systems on a VM machine.
6. As an independent operating system, VM offers capabilities not available on other IBM operating systems, especially for uses that require data management and modeling.
7. In Donovan and Madnick [12] it is shown that a hierarchically structured operating system, such as produced by a Virtual Machine System, which combines a virtual machine monitor with several independent operating systems (VMM/OS), provides substantially

better software security than a conventional two-level multiprogramming operating system approach.

The advantages of VMM essentially involved transparency of the VMM to the user and thus permitting timesharing of operating systems on one physical machine. The potential disadvantages are similar to those encountered when shifting from a batch to a timesharing operating system, namely, degradation of performance due to the incurred overhead and synchronization costs of the VMM system.

Estimating the costs of using VMM will depend upon a number of factors, including:

- Number of operating systems to be used
- Type of job mix being run, especially the I/O characteristics,
  - between operating systems
  - within operating system
- Scheduling and priority procedures implemented
  - between operating systems
  - within an operating system

Any computer center considering conversion to a VMM system must measure and evaluate these costs and benefits and, assuming that VMM is chosen, must use the information to develop scheduling and priority schemes which provide minimum system degradation while achieving the objectives of the center. Appendix C attempts to formalize performance questions for a potential FEA use.

5.5.2 Architecture of Ultimate GMIS System

Figure 5.24 depicts the interaction of multiple virtual machines operating on a single real computer. It is this configuration that will constitute the ultimate GMIS system. The virtual machines depicted across the top of the page are running modeling or analytic systems. Note that each of these systems may be running under different operating systems, e.g., TROLL runs under CP/CMS, where TSP may run under the operating system MVT.

All the modeling or analytic virtual machines may request data form the general data base system TRANSACT. Communication between virtual machines is accomplished via virtual card readers.

At the bottom of the page are depicted different and perhaps incompatible data base systems, all of which may be accessed by any of the modeling facilities.

5.5.3 Present GMIS VM's

The present operational GMIS system (February 1975) is a single user system consisting of two virtual machines as depicted in Figure 5.25. The exact configuration of each VM being simulated can be seen very concretely in Figure 5.26, where the configurations of two different VM's running on the same (physical) computer are described. It should be noted that the unit-record devices (reader, punch, printer) are also virtual. The DASD disk packs are physical, but it can be seen that the number of cylinders on each pack is smaller than would be found normally; a disk pack can be divided up among several users, so that each of them has a "mini-disk" whose storage capacity is a fraction of an entire pack.

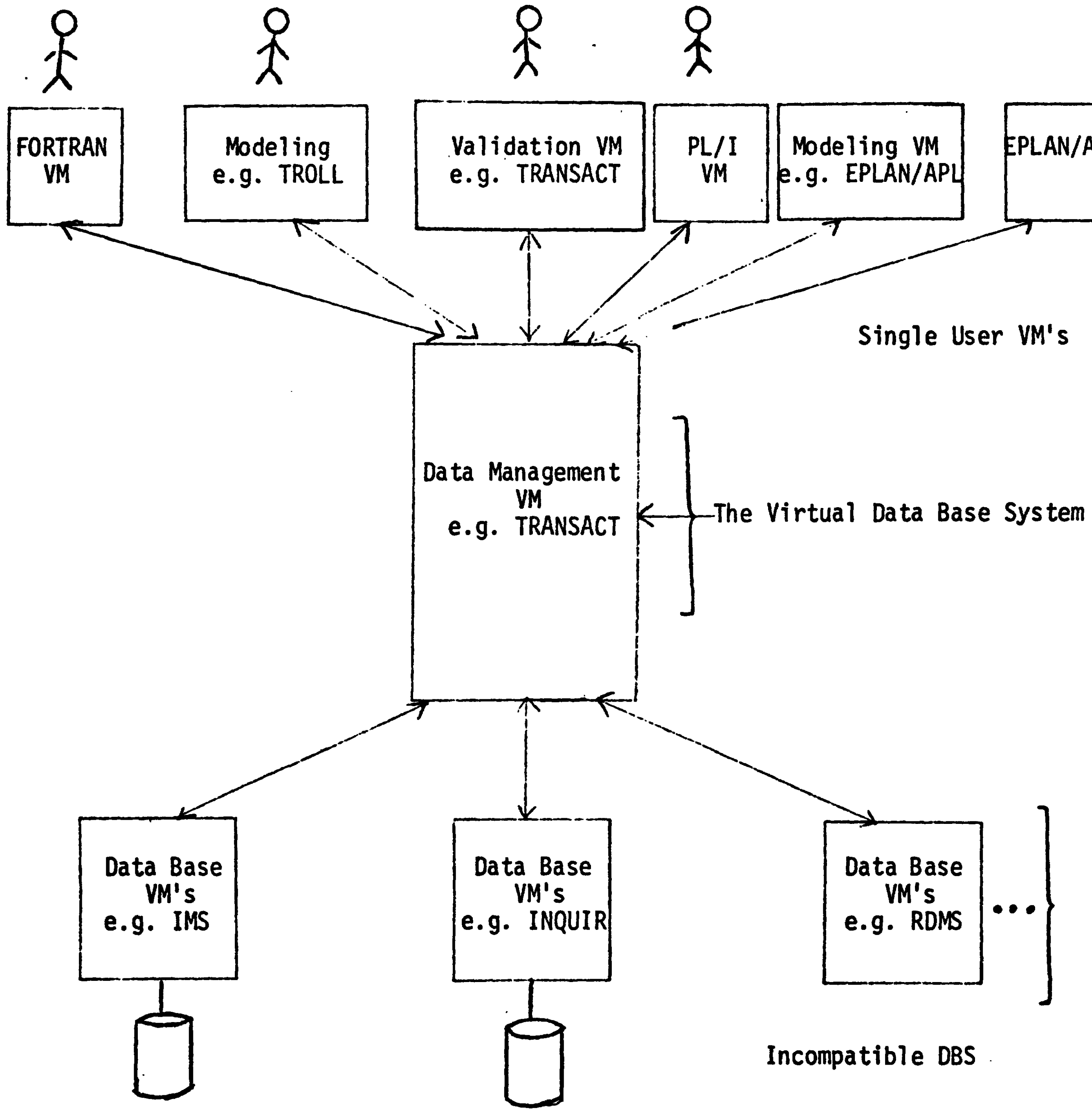

FORTRAN VM
Modeling e.g. TROLL
Validation VM e.g. TRANSACT
PL/I VM
Modeling VM e.g. EPLAN/APL
EPLAN/APL
Single User VM's
Data Management VM e.g. TRANSACT
The Virtual Data Base System
Data Base VM's e.g. IMS
Data Base VM's e.g. INQUIR
Data Base VM's e.g. RDMS
Incompatible DBS

Figure 5.24

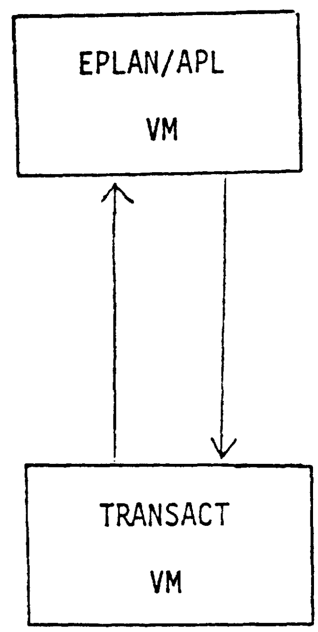


Figure 5.25

Architecture of Present GMIS

TRANSACT Virtual Machine Configuration

```
query virtual
STORAGE = 01024K
CHANNELS = SEL
CONS 009 ON LINE 05F    TERM STOP
     009 CL T  NOCONT NOHOLD COPY 01    READY
     009 FOR NEEMIS4  DIST NEEMIS4
RDR  00C CL *  NOCONT NOHOLD   EOF      READY
PUN  00D CL A  NOCONT NOHOLD COPY 01    READY
     00D TO  NEEMIS2  DIST NEEMIS4
PRT  00E CL S  NOCONT NOHOLD COPY 01    READY
     00E FOR NEEMIS4  DIST NEEMIS4
DEV  OFF PSEUDO TIMER
DASD 102 3330 VUSR02 R/O 050 CYL
DASD 10E 3330 VUSR04 R/O 100 CYL
DASD 10F 3330 VUSR04 R/O 100 CYL
DASD 190 2314 CMS370 R/O 060 CYL
DASD 191 3330 VUSR01 R/W 006 CYL
DASD 194 3330 VUSR03 R/O 025 CYL
DASD 19A 2314 19ASYS R/O 055 CYL
DASD 19B 3330 VUSR01 R/O 060 CYL
DASD 19E 3330 VUSR03 R/O 040 CYL
DASD 235 2314 CPCMS2 R/W 010 CYL
DASD 236 2314 CPCMS2 R/W 001 CYL
DASD 240 3330 VUSR02 R/W 003 CYL
DASD 245 2314 CPCMS1 R/W 020 CYL
DASD 323 3330 VUSR02 R/W 001 CYL
DASD 324 3330 VUSR02 R/W 001 CYL
DASD 325 3330 VUSR04 R/W 001 CYL
DASD 326 3330 VUSR04 R/W 001 CYL
DASD 335 3330 VUSR05 R/W 010 CYL
DASD 336 3330 VUSR06 R/W 001 CYL
DASD 340 3330 VUSR05 R/W 002 CYL
```

Figure 5.26

Configuration of Present GMIS VM's

APL/EPLAN Virtual Machine Configuration

```
query virtual
STORAGE = 01024K
CHANNELS = SEL
CONS 009 ON LINE 05F    TERM STOP
     009 CL T  NOCONT NOHOLD COPY 01    READY
     009 FOR NEEMIS2  DIST NEEMIS2
RDR  00C CL *  NOCONT NOHOLD    EOF     READY
PUN  00D CL A  NOCONT NOHOLD COPY 01    READY
     00D TO  NEEMIS4  DIST NEEMIS2
PRT  00E CL A  NOCONT NOHOLD COPY 01    READY
     00E FOR NEEMIS2  DIST NEEMIS2
DEV  0FF PSEUDO TIMER
DASD 102 3330 VUSR02 R/O 050 CYL
DASD 10E 3330 VUSR04 R/O 100 CYL
DASD 10F 3330 VUSR04 R/O 100 CYL
DASD 190 2314 CMS370 R/O 060 CYL
DASD 191 3330 VUSR02 R/W 006 CYL
DASD 19A 2314 19ASYS R/O 055 CYL
DASD 19C 2314 MEYER1 R/O 030 CYL
DASD 19E 3330 VUSR03 R/O 040 CYL
DASD 240 3330 VUSR02 R/W 003 CYL
DASD 340 3330 VUSR03 R/W 002 CYL
```

Figure 5.26 (cont'd)

Configuration of Present GMIS VM's

5.6 How FEA Might Take Advantage of the Topics of Section 5

We can identify five immediate ways that the work reported in this section can be used by FEA.

5.6.1 Virtual Machine Discussion

The FEA is in the process of choosing an operating and data management system. In Appendix C, we present a scheme using VM for analytical and Data Management uses in a way that heretofore has not been exploited. The scheme to using VM in this way and the advantages thereby derived should be weighed in FEA's present considerations of their applications.

5.6.2 Awareness of Data Base Technology of the Future

The GMIS system, though not commercially available, has demonstrated the feasibility and practicality of ideas that will be in data management systems of the future. FEA can prepare itself to take advantage of these developments as soon as they become commercially available.

5.6.3 Use of the GMIS System

It is possible for FEA to use our system. Let us outline some of the issues and procedures that FEA would follow if they wish to use our system.

5.6.3.1 Status of GMIS

Actually there are two independent but nearly functionally equivalent GMIS facilities.

One facility (the Gutentag facility) uses some experimental software from IBM in some of the lower levels. Some software was developed jointly by M.I.T. and IBM (middle levels); the highest levels use software developed solely by M.I.T. except for standard operating system. The other facility (the Smith facility) is being developed solely by M.I.T. [34].

The approaches are functionally equivalent in the sense that they both use a hierarchical implementation and there is a nearly one-to-one correspondence between levels. Why two approaches? The IBM software was not intended for this hierarchical approach and as such is often clumsy to work with. Further, the IBM software was not intended for use in an operational system and as such lacks some important practical features (e.g., limits on storage, limits on data types, etc.). Also the IBM software does not span all levels, thus lacks user features. Yet there exists IBM software that does work for some levels. The Smith approach, while elegant and without the deficiencies (e.g., protection), has as of November only a prototype operational and a design of a complete facility, thus the development of an operational system will take longer.

Both facilities run on IBM machines under VM. M.I.T. would release the Smith GMIS facility for use of FEA. M.I.T. would release the M.I.T. software of the Gutentag system but FEA would want to get a similar release or access to the IBM software; M.I.T., NERCOM[1], the State Energy Offices, City of New York all have such releases.

The software once released could either be used on FEA's 370's or any commercial timeshare service, e.g., NCSS timeshare, etc. Or if a joint study arrangement were made similar to M.I.T.'s or NERCOM, then IBM's Cambridge Scientific Centers facility might be used via remote terminals.

5.6.3.2 IBM/M.I.T. Joint Study

In October, M.I.T. and IBM signed an agreement dated October 21, 1974 relating to a joint study to explore the feasibility of developing an energy related data base using relational data models. The agreement sets forth IBM's and M.I.T.'s contribution of proprietary software, computer time, and manpower, as well as defining tasks and responsibilities of each group.

This joint study allowed M.I.T. access to certain proprietary software which we used in some of the lower levels of the Gutentag GMIS system.

In December 1973, M.I.T. started work on an energy information system for New England (NEEMIS). The sponsor and initiator of this work was New England Regional Commission (NERCOM). The GMIS facility has been applied to produce NEEMIS. Access to the NEEMIS application programs would require the approval of NERCOM.

---

[1] To allow NERCOM, State Energy Offices and FEA Region 1 to use the Gutentag GMIS facility which in turn uses some IBM proprietary software, IBM is considering a similar joint study agreement with NERCOM.

5.6.3.3 FEA Involvement in NEEMIS

Since December 1974, M.I.T.'s development effort of NEEMIS has not changed its goal of developing such a facility, even though the greatly heightened interest on the part of FEA Region 1 has resulted in some revision of work schedules. Potentially, however, the FEA Region 1 involvement may have a positive impact on the NEEMIS effort. With the FEA assistance we are now more hopeful of obtaining terminal inventory information as well as flow information. The number of fuels that might be tracked by NEEMIS may increase, and the information on fuel flows is now expected to be obtained from the supplier/importers, instead of from the distributors.

5.6.4 FEA Implement Own GMIS

FEA could use this report and Appendixes to develop on a commercially supported system an equivalent facility. This alternative would entail substituting commercial software for various levels of the NEEMIS software.

The following table gives examples of the types of substitutions that seem appropriate.

Our Level	Section Described	Possible Substitute Software
Prepared Packages	5.3.7	PL/1 programs that would have to be written
NEEMIS Interactive Query Facility	5.3.6	No such commercial facility is available but could be written using PL/1 language
NEEMIS High-Level Query Facility	5.3.5	No such facility exists. We would suggest that this level be eliminated as it is difficult using standard software
Modeling Facility	5.3.4	TROLL (NBER in Cambridge) or XSIM (distributed by Dynamics of Cambridge) could be used
DML and DDL Facility	5.3.2 & 5.3.3	JANUS
Relational Operator	5.3.1	Could be constructed using IMS

5.6.5 Future Experiments FEA May Wish to Try

M.I.T. will continue its major effort in the development of the GMIS facility and its application to energy related problems, analysis and policy making. FEA might want to join us in some of these endeavors or M.I.T. might want to join FEA in some of its work. Specifically we see the following interesting areas:

- Performance of VM for FEA's uses.
- Other modeling interfaces to GMIS e.g.,TROLL, TSP.
- Other data management interfaces to GMIS.

5.6.5.4 Expand GMIS

The major deficiencies of the NEEMIS software in its current state of development (February 1975) are:

- limited documentation
- implementation of only one modeling facility
- no automatic facilities for multi-user coordination and access to the data base
- limited secondary storage capacity (limited to one disk)
- no report generator facility
- software limitations as to table sizes, types of entries, number of entries, number of domains (e.g., only 15 domains, 2,000,000 table entries, entries must be only character strings or integers, entries cannot be floating point)

- no automatic mechanisms to facilitate data verification
- limited arithmetic facilities on data stored in tables (e.g., columns cannot be added together)
- no facilities for the completely unsophisticated user (e.g., no English language commands, no scope interface)
- data security limited to that provided by the operating system

Most of these deficiencies could be corrected with a few months of effort.

5.6.5.5 Evaluation of Other Systems for FEA Use

Evaluate present commercially available systems for their suitability to FEA's needs. Evaluate advanced research systems (e.g., ACOS/DASEL, GMIS) for their suitability.

6. REFERENCES*

1. American Petroleum Institute, Annual Statistical Review.

2. Evans, M.K., Macroeconomic Activity: Theory, Forecasting, and Control (Harper and Row, 1969).

3. "'Leading' Indicators Discover Recession," New York Times, November 3, 1974.

4. M.I.T. Energy Laboratory, "Survey of Basic Data Series," draft working paper for the Office of Data Policy, Federal Energy Administration, August 14, 1974.

5. M.I.T. Energy Laboratory, "Problems of Data Transfer and Management," draft working paper for the Office of Data Policy, Federal Energy Administration, August 15, 1974.

6. M.I.T. Energy Laboratory, "Conceptual Framework for Energy Indicators," draft working paper submitted to the Office of Data Policy, Federal Energy Administration, September 19, 1974.

7. M.I.T. Energy Laboratory, "Preliminary Results on Selected Sufficiency Indicators," draft working paper submitted to the Office of Data Policy, Federal Energy Administration, November 7, 1974.

8. M.I.T. Energy Laboratory, "'Snapshot' Indicators of Energy Sufficiency," draft working paper submitted to the Office of Data Policy, Federal Energy Administration, December 23, 1974.

9. Wesley C. Mitchell, Business Cycles: The Problem and Its Setting (NBER, 1927).

10. TROLL Reference Manual, National Bureau of Economic Research, 1972.

11. Senge, P.M., "An Experimental Evaluation of Least-Squares Estimation," Working Paper, System Dynamics Group, Sloan School of Management, M.I.T., November 14, 1974.

12. Tepper, Jonathan, "A Short-Range Model of the Domestic Production of Refined Petroleum Products," Masters Thesis Progress Report, Sloan School of Management, M.I.T., December, 1974.

13. U.S. Federal Energy Administration, National Energy Information Center, Monthly Energy Review.

14. U.S. Bureau of Mines, Minerals Yearbook.

15. U.S. Bureau of Labor Statistics, BLS Handbook of Methods (Bulletin 1711), 1971.

10 a. Peterson, D.W. and F.C. Schweppe, "Code for a General Purpose System Identifier and Evaluation (GPSIE)" IEE Transactions on Automatic Control, December, 1974.

----

*References for Section 5 are listed separately on the following pages.

REFERENCES (SECTION 5)

1. Bracchi, G. et al.: "A Language for a Relational Data Base Management System," PROCEEDINGS, 5th PRINCETON CONFERENCE ON INFORMATION SCIENCE, 1972.

2. Bradley, "The Bulk Loader", M.I.T. Internal Document, November 1974.

3. Buzen, J. P., Peter P. Chen, and Robert P. Goldberg, "Virtual Machine Techniques for Improving System Reliability", Proceedings of the ACM Workshop on Virtual Computer Systems, (March 26-27, 1973).

4. Chamberlin, D. D. and R. F. Boyce: "SEQUEL: A Structured English Query Language," PROCEEDINGS 1974 ACM/SIGFIDET Workshop.

5. Church, A.: "The Calculi of Lambda-Conversion," ANNALS OF MATHEMATICS STUDIES, No. 6, Princeton University Press, 1941.

6. Codd, E. F.: "A Relational Model of Data for Large Shared Data Banks," CACM, Vol. 13, No. 6, June 1970, p. 377-387.

7. Codd, E. F.: "A Data Base Sublanguage Founded on the Relational Calculus," PROCEEDINGS 1971 ACM/SIGFIDET Workshop.

8. Dijkstra, E.: "T.H.E. Multiprogramming System," CACM, May 1968.

9. Donovan, John J.: SYSTEMS PROGRAMMING, McGraw-Hill, New York, 1972.

10. Donovan, John J. and Henry D. Jacoby: "A Hierarchical Approach to Information System Design," Report CISR-5, M.I.T. Sloan School Working Paper 762-75, January 1975.

11. Donovan, John and Henry Ledgard: "A Formal System for the Specification of the Syntax and Translation of Computer Language," PROCEEDINGS 1967 FJCC.

12. Donovan, John J. and Stuart E. Madnick, "Application and Analysis of the Virtual Machine Approach to Computer System Security and Reliability," M.I.T. Sloan School Report CISR-2, May 1974 (to be published in the IBM SYSTEMS JOURNAL, May 1975).

13. Donovan, John, J., Louis M. Gutentag, Stuart E. Madnick, and Grant H. Smith . "An Application of a Generalized Management Information System to Energy Policy and Decision Making: The User's View", AFIPS Conference, 1975 NCC.

14. Essrig, Marvin J., Applications of a Generalized Management Information System to Energy (proposed title), M.S. thesis, Sloan School of Management, M.I.T., (May 1975)

15. Fehder, P.C.: "The Representation Independent Language," IBM Technical Report RJ1121, November 1972.

16. Goldberg, R. P., "Virtual Machines: Semantics and Examples", Proceedings of IEEE Computer Society Conference, (September 1971), 141-142.

16a. Goldberg, R. P., Architectural Principles for Virtual Computer Systems, PHD dissertation, Harvard University, (November 1972).

16b. Goldberg, R. P., "Architecture of Virtual Machines", Proceedings AFIPS, National Computer Conference, vol. 42, pp. 309-318, 1973.

17. Goldberg, R. P., "Survey of Virtual Machine Research", Computer, Volume 7, Number 6, (June 1974), pp. 34-45.

18. Goldstein, I. & Alois Strnad: "The MACAIMS Data Management System", M.I.T. Project MAC TM-24, April 1971.

19. Hatfield, D. J., "Experiments on Page Size Program Access Patterns, and Virtual Memory Performance", IBM Journal of Research and Development, vol. 16, no. 1, pp. 58-66, January 1972.

20. Hogg, J. and P. Madderom, "The Virtual Machine Facility -- How to Fake a 360", Internal Note, University of British Columbia and University of Michigan Computer Center, (1973).

21. IBM, APL/360 User's Manual, Form Number GH20-0906 (January 1973).

22. IBM, "IBM Virtual Machine Facility/370: Introduction", Form Number GC20-1800, (July 1972).

23. Iverson, Kenneth E., A Programming Language, John Wiley and Sons, 1962.

24. Madnick, S. E.: "Design Strategies for File Systems," M.I.T. Project MAC TR-78, October 1970.

25. Madnick, S. E., "Time-Sharing Systems: Virtual Machine Concept vs. Conventional Approach", Modern Data 2, 3 (March 1969), 34-36.

26. Madnick, Stuart E. and John J. Donovan, Operating Systems McGraw-Hill, New York, 1974.

27. Morrison, J. E., "User Program Performance in Virtual Storage Systems, IBM Systems Journal, vol. 12, no. 3, pp. 216-237,1973.

28. Pakin, Sandra, APL/360 Reference Manual, Science Research Associates, 1968.

29. Parmelee, R. P., T. I. Peterson, C. C. Sullivan, and D. S. Hatfield, "Virtual Storage and Virtual Machine Concepts", IBM Systems Journal, vol. 11, no. 2, pp. 99-130, 1972.

30. Popek, G. J. and C. Kline, "Veritable Secure Operating Systems Software", AFIPS Conference, 1974 NCC.

31. Post, E. L.: "Formal Reductions of the General Combinatorial Decision Problem," AMERICAN JOURNAL OF MATHEMATICS, Vol. 65, 1943, p. 197-215.

32. PRIME Computer, Inc., "DOS/VM Reference Manual", (1974).

33. Schober, F., EPLAN- An APL-based Language for Econometric Modeling and Forecasting, IBM Philadelphia Scientific Center, 1974.

34. Smith, Grant N., "Internal Intermediate Language, Version 2," M. I. T. Sloan School Report CISR-6, November 1974.

35. Smullyan, R.: "Theory of Formal Systems," Study 47, Princeton University Press, 1961.

36. TROLL REFERENCE MANUAL, National Bureau of Economic Research, 1972.

37. MacAvoy, P. L., "Alternative regulatory policies for dealing with the natural gas shortage," The Bell Journal of Economics and Management Science, Autumn 1973, Vol. 4, No. 2.

APPENDIX A: POSSIBLE AREAS FOR CONSTRUCTION OF "SNAPSHOT" INDICATORS

At one stage in the project a complete list was made of all areas where suggestions had been made for the construction of indicators. Sources for this list indluded:

1) Research by the project team.
2) Suggestions by personnel in the office of Energy Data.
3) Interviews with FEA personnel outside the office of Energy Data.
4) Interviews with personnel in industry and the financial community.
5) Review of places in the Project Independence Report where there is an indication of a need for information or of a quantity which is important to the process of attaining energy self-sufficiency.

This list is presented here without significant editing or correction; no attempt has been made to cull out poor suggestions or duplicate ideas. A reduced list is presented in Section 3.2.

The list is broken down according to the six categories of indicators presented in Section 2.2 where the source of the suggestion was the Project Independence Review, it is noted PIR with a page reference.

1. LONG RUN DOMESTIC SUPPLY

1.1 Amount of electricity generated by nuclear sources.
1.2 Coal as a percent of total fuels used to produce electricity.
1.3 Drilling rigs in operation; well completions.
1.4 Capital expenditures in petroleum and coal industry.

1.5 Distribution of sources of electrical energy, and projected distribution of planned generating capacity.

1.6 Back orders of drilling rigs, movable platforms, and drag lines.

1.7 Rates of inflation and hourly wage increases (PIB, p. 323).

1.8 Estimates of potential production levels for each fuel and each major production region, as a function of price, taking into account institutional factors and production lead times (PIB, p.63).

1.9 Leasing rate of federal lands (PIB p. 64).

1.10 Leasing location and timing (PIB, p. 73).

1.11 Leasing bonuses; average bonus bid per acre (PIB, p. 73).

1.12 Reserves of oil, and gas (PIB, p. 65).

1.13 New discoveries of oil, coal, and gas (PIB, p. 65).

1.14 Price elasticities of supply for oil, gas, coal, nuclear and synthetic fuels (PIB, p. 68).

1.15 Bottlenecks and construction lead times for new facilities (PIB, p. 68-69).

1.16 Orders for new plants (PIB, p. 70).

1.17 Natural gas wells drilled (PIB, p. 74).

1.18 Natural gas additions to reserves (PIB, p. 74).

1.19 Natural gas excess demand and/or curtailments (PIB, p. 74).

1.20 Natural gas regulated prices (PIB, p. 74).

1.21 Drilling success rates (PIB, p. 76).

1.22 Discoveries as a function of footage drilled and finding rates (PIB, p. 78)

1.23 Depletion factor or decline rate for exhaustible resources (PIB, p.78).

1.24 Drilling and exploration costs; cost per foot of a gas well (PIB, p. 88).

1.25 Share of production of particular mining methods (PIB, p. 101).

1.26 Productivity of different mining methods (PIB, p. 101).

1.27 Estimates of uranium ore reserves and prices at which minable (PIB, p. 113).

1.28 Uranium mining and milling capacity (PIB, p. 114).

1.29 Uranium enrichment capacity (PIB, p.114).

1.30 Spent fuel reprocessing capacity (PIB, p. 114).

1.31 Uncertainties about schedule for on-line nuclear generating capacity (PIB, p. 114).

1.32 Construction plans and level of investment in nuclear power industry (PIB, p. 114).

1.33 Ability of utilities to raise funds on capital markets (PIB, p. 115).

1.34 Fractions of electric power generation from various sources: Hydro, fossil fuels, and nuclear.

1.35 Capital requirements for planned expansion in electric utility sector.

1.36 Site availability for electric power plants (PIB, p. 128).

1.37 Capital cost of new electric plants (cost per kw of installed capacity): Base loaded, intermediate loaded, peak loaded (PIB, p. 124-126).

1.38 Expected rate of return of shale oil investment projects (PIB, p. 130-134).

1.39 Sulfur emission limits, especially in Colorado, as they relate to shale oil development (PIB, p. 134).

1.40 Average percent utilization and/or reserve ratio of electric generating capacity (PIB, p. 173).

1.41 Average percent efficiency of generation and distribution (PIB, p. 174).

1.42 Number of new electric plants by type (PIB, p. 185).

1.43 Use of renewable vs. non-renewable energy resources (PIB, p. 202).

1.44 Regional breakdown of percent of nation's coal production, and percent of nation's total coal reserves (PIB, p. 203).

1.45 Availability (present units plus new units minus units replaced) of: Drilling rigs, fixed drilling platforms, mobile drilling platforms, oil country tubular goods, steel products, steel pipe and tubing, walking draglines, and steam turbine generators, and domestic capacity to produce them, and net exports (PIB, p. 232-245).

1.46 Availability of materials required in production of drilling rigs: Steel products, steel castings, forgings, tubular goods, and other equipment (PIB, p. 257).

1.47 Cumulative dollar investment for expansion of transportation network by 1985 in oil, gas, and coal industries (PIB, p. 266).

1.48 Cumulative steel requirements for expansion of transportation network by 1985 (PIB, p. 267).

1.49 Transportation cost of coal; absolute and percent of total cost (PIB, p. 268).

1.50 Transportation costs as percent of total delivered cost for: Crude to refinery, natural gas to processing plant, and all products from refineries or plants (PIB, p. 268).

1.51 A need/availability indicator for steel pipe used in transportation (PIB, p. 274).

1.52 Structural steel, angular steel, sheet steel, and plate steel, as required for rolling stock replacements and additions (PIB, p. 277).

1.53 Lock and dam capacities in waterway system versus future usage estimates. (The capacity of a waterway is controlled by the lock with the smallest throughput) (PIB, p. 278).

1.54 Energy investment as a percent of total business fixed investment (PIB, p. 279-280).

1.55 Asset expansion of electric utilities (PIB, p. 285).

1.56 Revenues of utilities against expected revenues given rate increases, consumption expectations (PIB, p. 288).

1.57 Peak load growth rate versus total power output growth rate for utilities, a proxy for differential capital requirements (PIB, p. 288).

1.58 Energy employment trends for utilities, oil & gas extraction, oil refining, coal, pipelines (PIB, p. 296).

1.59 Energy sector employment of the following, (tracking the trend against updated projections and needs): Nuclear engineers, metallurgical engineers, mechanical engineers, health technicians, millwrights, physicists, pipefitters, plumbers, welders, electricians, boilermakers, carpenters, electrical engineers, geologists, draftsmen, other technicians (PIB, p. 298-301).

1.60 Corporate profits before taxes for natural gas, crude petroleum, petroleum refining, coal, and electric utilities, relative to all industries.

1.61 Reserve/production ratios for crude oil and natural gas.

2. LONG-RUN CONSUMPTION

2.1 Total energy consumed (BTU's per GNP dollar, current and constant dollars) and energy consumed, deleting energy used for heating.

2.2 Airline passenger load factors.

2.3 Energy consumed for all modes of transportation (BTU's per ton-mile of freight moved, weighted by amount of freight carried).

2.4 Gasoline consumption of new autos sold (ton-miles per gallon, monthly).

2.5 Industrial energy usage efficiency indicator.

2.6 Total gross consumption of energy (PIB, p. 318).

2.7 Growth of gross national product, personal consumption, gross private domestic investment, employment, and productivity (PIB, p. 320).

2.8 Annual housing starts by type (PIB, p. 322).

2.9 Rates of inflation (PIB, p. 323).

2.10 Rates of increase in hourly wages (PIB, p. 323).

2.11 Total personal income, and shares derived from labor income, transfer payments, and business & property income. (PIB, p. 334-335).

2.12 Percent of after tax income spent on gasoline by urban households (PIB, p. 338).

2.13 Percent of after tax income spent on electricity by urban households (PIB, p. 339).

2.14 Percent of after tax income spent on all energy purchases by urban households (PIB, p. 340).

2.15 Natural gas excess demand (PIB, p. 74).

2.16 Natural gas curtailments (PIB, p. 74).

2.17 Natural gas share of intrastate sales (PIB, p. 74).

2.18 Growth rate of crude oil consumption (PIB, p. 75-76).

2.19 Growth rate of natural gas consumption (PIB, p. 86).

2.20 Consumption uses of different fuels (use categories include industrial, residential, commercial, transportation, ...)(PIB, p. 86).

2.21 Growth rate of coal consumption, by consuming sector (PIB, p. 98).

2.22 Percentage share of energy consumption by all fuels (PIB, p. 98).

2.23 Percentage of total energy consumed as electricity (PIB, p. 118).

2.24 Growth rate of electric power consumption (PIB, p. 118).

2.25 Fractions of electric power generation from various sources: Hydroelectric, fossil (coal, natural gas, oil), and nuclear.

2.26 Energy demand of electric utilities, by type of fuel (PIB, p. 157).

2.27 Miles per gallon rating of automobiles (PIB, p. 162).

2.28 Average number of commuters per automobile (PIB, p. 157).

2.29 Total U.S. vehicle miles (PIB, p. 164).

2.30 Total motor vehicles in operation, and total sold per year (PIB, p. 164).

2.31 Parking capacity in urban areas (PIB, p. 164).

2.32 Home insulation consumption (PIB, p. 164).

2.33 Major appliances in use, and total sold per year, by type (PIB, p. 170).

2.34 Lighting sales (in lumens), (PIB, p. 169).

2.35 R & D dollars expended for production efficiency programs (PIB, p. 160).

2.36 Percentage efficiency of generation & distribution for electric utilities (PIB, p. 174).

2.37 Number of electric utility plants switched from oil & gas to coal (PIB, p. 185).

2.38 Number of new electric power plants, by type (PIB, p. 185).

2.39 Number of current & new residential & commercial heating systems, by type of fuel used (PIB, p. 186).

2.40 Number of heat pumps installed, commercially & in residences.

3. SHORT-RUN DOMESTIC SUPPLY ADEQUACY

3.1 Days of supply remaining of petroleum and coal.

3.2 Capacity to produce unleaded gasolines.

3.3 Total domestic production of crude (PIB, p. 318).

3.4 Total petroleum imports (PIB, p. 318).

3.5 Estimates of potential production levels for each fuel, and major production regions, as a function of price (taking into account institutional factors and production lead times) (PIB, p. 63).

3.6 Natural gas curtailments (PIB, p. 74).

3.7 Degree of dependence on foreign supplies of crude oil and refined products (PIB, p. 76).

3.8 Electricity interruptions (PIB, p. 118-125).

3.9 Reliability of nuclear plants (Number of shutdowns; downtime) (PIB, p. 128).

4. PRICES

4.1 Price of energy per BTU for all fuels (current & constant dollars).

4.2 Average national BTU price (PIB, p. 113).

4.3 Rates of inflation (PIB, p. 323).

4.4 Rates of increase in hourly wages (PIB, p. 323).

4.5 Regional energy prices (PIB, p. 326, 328-329).

4.6 Absolute differences between maximum & minimum average regional BTU prices (PIB, p. 331).

4.7 Price elasticities of supply for oil, coal, gas, nuclear, and synthetic fuels (PIB, p. 68).

4.8 Natural Gas regulated prices (PIB, p. 74).

4.9 Prices for long term coal contracts & spot market (PIB, p. 106).

4.10 Transportation cost of coal (absolute and percent of total cost) (PIB, p. 268).

4.11 Transportation costs as percent of total delivered cost for crude to refinery, natural gas to processing plant, and all products from refineries or plants (PIB, p. 268).

4.12 Consumer and wholesale price indices for all fuels and electricity.

4.13 World price of energy per BTU for all fuels.

5. INTERNATIONAL MARKET

5.1 Total petroleum imports (PIB, p. 318).

5.2 Monthly dollar outflows attributable to petroleum imports (PIB, p. 324).

5.3 Degree of dependence on foreign crude supplies (PIB, p. 76).

5.4 Percent usage of domestic versus imported fuels (PIB, p. 201).

5.5 Net exports of the following items: oil & gas drilling rigs, fixed and mobil drilling platforms, oil country tubular goods (million tons), steel products (million tons), steel pipe & tubing, walking draglines, and steam turbine generators (large) (PIB, p. 232, 239, 242, 245).

5.6 Excess production capacity in OPEC

5.7 International production and consumption of all fuels.

5.8 Imports/stocks for crude oil and refined products.

5.9 Imports/total domestic consumption of crude oil and refined products.

6. ENVIRONMENTAL AND SOCIAL IMPACTS

6.1 Coal as a percent of total fuels used to produce electricity.

6.2 Capacity to produce unleaded gasolines.

6.3 Environmental quality index for selected areas of the U.S.

6.4 Current and projected installations of stack gas cleaning equipment.

6.5 Sulfur emission limits (for shale oil) (PIB, p. 134).

6.6 Energy resource production from new areas of production compared with existing areas of production (PIB, p. 200).

6.7 Amount and percentage of western versus eastern coal used (PIB, p. 200).

6.8 Amount and percentage of energy from nuclear fuels versus fossil fuels (PIB, p. 200).

6.9 Indicator of proximity of energy facilities to demand centers and of proximity of energy facilities to production centers (PIB, p. 201).

6.10 Percentages of total energy usage of renewable and non-renewable resources (PIB, p. 201).

6.11 Number of injuries, disabling injuries, and fatalities per unit of coal mined (PIB, p. 203).

6.12 Percentage of strip mining land which undergoes reclamation (PIB, p. 203).

6.13 Regional breakdown of the following:

-percentage of U.S. coal production in that region

-percentage of U.S. coal reserves in that region (PIB, p. 203).

6.14 Uranium content of coal being mined, by region (PIB, p. 204).

6.15 Sulfur content of coal being mined, by region (PIB, p. 204).

6.16 For oil & gas production, amount of waste materials generated (PIB, p. 204).

6.17 Number and volume for each of oil spills, classified as follows:

-spills at terminals

-spills from ships offshore

-spills from offshore production facilities

-onshore pipeline accidents

-spills from ships weighted by barrel-miles of oil shipped (PIB,p.205-7).

6.18 Animal deaths attributable to oil spills (birds, fish) (PIB, p. 206).

6.19 For fossil fuel power plants, the number and percentage of the total of facilities with given types of pollution control devices (PIB, p. 210).

6.20 Amount of radioactive particles emitted by region, and amount of radioactivity by region (PIB, p. 211).

6.21 Number and severity of accidents and "near accidents" at nuclear power plants (PIB, p. 211).

6.22 Amount of solid radioactive wastes generated which must be safely disposed of (PIB, p. 211).

6.23 For the following items, the amount of environmental residual generated both by energy sector and type of facility, and the amount of residual produced per amount of energy generated:

-acids (equivalent tons/day).

-bases (equivalent tons/day)

-total dissolved solids (tons/day)

-suspended solids (tons/day)

-organics or oil spills

-thermal water pollution (BTU's/day)

-particulates (tons/day)

-nitrogen oxides (tons/day)

-sulfur oxides (tons/day)

-hydrocarbons (tons/day)

-carbon monoxide (tons/day)

-aldehydes (tons/day)

-solids (tons/day)

-fixed land (acres/year) (alternative uses precluded for some time)

-incremental land (acres/year) (maximum excluded from alternative uses)

(PIB. pp. 214ff.)

6.24 Number of power plants switched from oil & gas to coal (PIB, p. 185).

APPENDIX B

DATA DOCUMENTATION

The attached sheets provide documentation of the basic data series that are in the indicator data bank as of February 28, 1975. They are presented here in alphabetical order. Series taken from the F.E.A. Monthly Energy Review are not documented here.

DOCUMENTATION FOR: CPI.DS

NAME: Consumer Price Index for #2 Fuel Oil.

DEFINITION: Indicates current retail rates for #2 heating fuel (primarily for residential rise) relative to some base period.

UNITS: Index numbers

FREQUENCY: Monthly

SOURCE: Survey of Current Business, Bureau of Labor Statistics.

COMMENTS: Data available from February, 1968 to November, 1974 (most current update). Beginning with January, 1971, base period is 1967 = 100. Previous to 1971, base period is 1957-59 = 100.

DOCUMENTATION FOR: CPI.EL

NAME: Consumer Price Index for Electric Power.

DEFINITION: Indicates current retail rates for residential electric power relative to some base period.

UNITS: Index numbers

FREQUENCY: Monthly

SOURCE: Survey of Current Business, Bureau of Labor Statistics.

COMMENTS: Data available from February, 1968 to November, 1974 (most current update). Beginning with January, 1971, base period is 1967 = 100. Previous to 1971, base period is 1957-59 = 100.

DOCUMENTATION FOR: CPI.MG

NAME: Consumer Price Index for Motor Gasoline.

DEFINITION: Shows current level of retail prices for gasoline relative to some base period.

UNITS: Index Numbers

FREQUENCY: Monthly

SOURCE: Survey of Current Business, Bureau of Labor Statistics.

COMMENTS: After January 1971, the base year is 1967 = 100.

DOCUMENTATION FOR: GAB

NAME:	Corporate Profits Before Taxes - All Industries - Total
DEFINITION:	Indicates total level of before - tax profit for all U.S. industry.
UNITS:	Millions of (current) dollars.
FREQUENCY:	Annual
SOURCE:	Survey of Current Business.
COMMENTS:	Data available from 1948.

DOCUMENTATION FOR: GABM12

NAME: Corporate Profits Before Tax - Coal Mining.

DEFINITION: Indicates total level of corporate profit for coal mining industry.

UNITS: Millions of (current) dollars.

FREQUENCY: Annual

SOURCE: Survey of Current Business.

COMMENTS: Data available beginning in 1948.

DOCUMENTATION FOR: GABM13

NAME:	Total Corporate Profits Before Tax - Crude Petroleum and Natural Gas Mining.
DEFINITION:	Indicates total corporate profits for petroleum and natural gas extracting industries.
UNITS:	Millions of (current) dollars.
FREQUENCY:	Annual
SOURCE:	Survey of Current Business
COMMENTS:	Data available beginning with 1948.

DOCUMENTATION FOR: GABN29

NAME: Corporate Profits Before Tax - Petroleum Refining and Related Industries.

DEFINITION: Indicates total level of corporate profit for petroleum refining and related industries.

UNITS: Millions of (current) dollars

FREQUENCY: Annual

SOURCE: Survey of Current Business

COMMENTS: Data available beginning in 1948.

DOCUMENTATION FOR: GABUT

NAME:	Corporate Profits Before Tax - Electric, Gas and Sanitary Services.
DEFINITION:	Indicates total level of corporate profit for electric, gas and sanitary services sectors.
UNITS:	Millions of (current) dollars.
FREQUENCY:	Annual
SOURCE:	Survey of Current Business
COMMENTS:	Data available beginning in 1948.

DOCUMENTATION FOR: GAS.PROD

NAME: Total Natural Gas Production

DEFINITION: GAS.PROD--Gas production is the total volume of natural gas withdrawn from producing reservoirs less the volume returned to such reservoirs in cycling, repressuring of oil reservoirs and conservation operations. Corrections for shrinkage are also made. The net change in underground storage volumes is not included in production. Thus net production relates specifically to the depletion of the proved gas reserves of the natural reservoirs.

UNITS: Billions of cubic feet.

FREQUENCY: Annual

SOURCE: American Gas Association and American Petroleum Institute.

COMMENTS: Data available from 1960 through 1973. Marketed gas volumes are not comparable to net gas production as a portion of the gas withdrawn from reservoirs is consumed in field operations as lease and plant fuel.

DOCUMENTATION FOR: GAS.RSVS

NAME: Proved Natural Gas Reserves

DEFINITION: Proved reserves of natural gas (estimated as of December 31 of any given year) are the estimated quantities of all natural gases and natural gas liquids statistically defined as such, which geological and engineering data demonstrate with reasonable certainty to be recoverable in future years from known reservoirs under existing economic and operating conditions.

UNITS: Billions of cubic feet.

FREQUENCY: Annual

SOURCE: American Gas Association and American Petroleum Institute.

COMMENTS: Data available from 1960 through 1973.

DOCUMENTATION FOR: LPMI

NAME:	Total Number of Employees - Mining Industries
DEFINITION:	Indicates total number of employees engaged in mining industries (seasonally adjusted)
UNITS:	Thousands of workers
SOURCE:	U.S. Department of Labor, Bureau of Labor Statistics, "The Employment Situation".
COMMENTS:	Data available beginning in 1947.

DOCUMENTATION FOR: LPN29

NAME:	Total Number of Employees - Petroleum and Coal Products Industry.
DEFINITION:	Indicates total number of employees engaged in the petroleum and coal products industries.
UNITS:	Thousands of workers
SOURCE:	U.S. Department of Labor Bureau of Labor Statistics, "The Employment Situation".
COMMENTS:	Data available beginning in 1947.

<u>DOCUMENTATION FOR: PC</u>

NAME: The Consumer Price Index (CPI), All Items.

DEFINITION: The consumer price index is a statistical measure of changes in prices of goods and services bought by urban wage earners and clerical workers, including families and single persons. The CPI is a weighted aggregative index number with fixed or constant annual weights, or it often is referred to as a market basket index.

INTERPRETATION: The CPI is used extensively to measure changes in the purchasing power of the consumer dollar. Also it is a broad measure of the degree of inflationary pressure in the economy.

UNITS: Index numbers, 1967 = 100 after January 1971.

FREQUENCY: Monthly

FORMULA:

$$I_{i:o} = \frac{\sum\left[(p_o q_o)\frac{p_i}{p_o}\right]}{\sum(p_o q_o)} \times 100$$

p_o = average retail price of selected commodities in a base period

p_i = average retail price in current period

q_o = composite of quantities sold at retail of same selected commodities in a base period

q_i quantities sold at retail in current period

SOURCE: U.S. Department of Labor, Bureau of Labor Statistics, "The Consumer Price Index," (published monthly)

CONTACT: John Curtis, FEA, Office of Data Policy, (202) 961-7986.

COMMENTS: Series available in different forms and using various base years since early 1900's. Computerized data base summarizes data from 1947 on. There are a number of limitations on the use of the CPI which is currently being overhauled so that the typical market basket is more representative of the typical consumer's expenditures. Also see PW documentation.

DOCUMENTATION FOR: PET.DMD

NAME:	Total Petroleum Consumption During Year
UNITS:	Millions of barrels/day.
SOURCE:	Federal Energy Administration. PIB report.
COMMENTS:	Data available from 1950 through 1972.

DOCUMENTATION FOR: PET.PROD

NAME:	Oil Production During Year
DEFINITION:	Crude oil production is the volume of liquids statistically defined as crude oil, which is produced from oil reservoirs during a year. The amount of such production is generally established by measurement of volumes delivered from least storage tanks (i.e., the point of custody transfer) to pipelines, trucks, or other media for transport to refineries or terminals.
UNITS:	Millions of barrels per year
FREQUENCY:	Annual
SOURCE:	American Gas Association, American Petroleum Institute, Bluebook.
COMMENTS:	Data available from 1960 through 1973.

DOCUMENTATION FOR: PET. RSVS

NAME:	Proved Reserves of Oil at End of Year
DEFINITION:	Proved reserves of crude oil (estimated as of December 31 of any given year) are the estimated quantities of all liquids statistically defined as crude oil, which geological and engineering data demonstrate with reasonable certainty to be recoverable in future years from known reservoirs under existing economic and operating conditions.
UNITS:	Millions of barrels
FREQUENCY:	Annual
SOURCE:	American Gas Association, American Petroleum Institute, Bluebook.
COMMENTS:	Data available from 1960 through 1973.

DOCUMENTATION FOR: PW

NAME:	Wholesale Price Index
DEFINITION:	Indicates the overall level of wholesale prices relative to some base period.
UNITS:	Index numbers
FREQUENCY:	Monthly
SOURCE:	U.S. Department of Labor, Bureau of Labor Statistics, "Wholesale Prices and Price Indexes", (published monthly).
COMMENTS:	Series available in different forms and using different base years since early 1900's. Computerized data base summarizes data from 1947 on.

DOCUMENTATION FOR: PX

NAME:	Expenditures for New Plant and Equipment - all Industries.
DEFINITION:	Indicates total level of expenditure (seasonally adjusted) for new plant and equipment for all industries.
UNITS:	Billions of (current) dollars.
FREQUENCY:	Quarterly
SOURCE:	U.S. Department of Commerce, Social and Economic Statistics Administration, Bureau of Economic Analysis, and The Securities and Exchange Commission "Joint Statistical Report".
COMMENTS:	Data available from 1947. Contains two quarter projection based on expectations. Seasonal adjustments are made at annual rates.

DOCUMENTATION FOR: PXBM

NAME:	Expenditures for New Plant and Equipment - Mining Industry.
DEFINITION:	Indicates total level of expenditure (seasonally adjusted) for new plant and equipment - mining industry.
UNITS:	Billions of (current) dollars
FREQUENCY:	Quarterly
SOURCE:	U.S. Department of Commerce, Social and Economic Statistics Administration, Bureau of Economic Analysis and the Securities and Exchange Commission "Joint Statistical Report ".
COMMENTS:	Data available beginning with 1947. Contains two quarter projection based on expectations.

DOCUMENTATION FOR: PXMNG

NAME:	Expenditures for New Plant and Equipment - Petroleum Industry
DEFINITION:	Indicates total level of expenditure (seasonally adjusted) for new plant and equipment - petroleum industry.
UNITS:	Billions of (current) dollars
FREQUENCY:	Quarterly
SOURCE:	U.S. Department of Commerce Social and Economic Statistics Administration, Bureau of Economic Analysis, and the Securities and Exchange Commission "Joint Statistical Report".
COMMENTS:	Data available from 1947. Contains two quarter projection based on expectations.

DOCUMENTATION FOR: RCAR6D

NAME: Number of New Domestic Passenger Cars Sold at Retail

UNITS: 000's of autos

SOURCE: U.S. Department of Commerce, Social and Economic Statistics Administration, Bureau of Economic Analysis, "Survey of Current Business"

COMMENTS: Data available from 1958.

DOCUMENTATION FOR: RIGS60

NAME:	Total Number of Rotary Dulling Rigs Running
DEFINITION:	Indicates total number of rotary drilling rigs for U.S., excluding: cable tools, stacked rigs and rigs moving to new locations.
INTERPRETATION:	Domestic discoveries of oil and gas and subsequent production cannot take place without the required drilling equipment represented by RIGS6Q.
UNITS:	Numbers of rigs
FREQUENCY:	Monthly
SOURCE:	Hughes Tool Co.
COMMENTS:	Data available from January, 1960 to May, 1975 (estimated).

APPENDIX C

USE OF VIRTUAL MACHINES IN INFORMATION SYSTEMS

John J. Donovan

May 1975

Energy Lab in Association
with Sloan School

Report No. MIT-EL-75-010

Abstract

This paper presents a scheme using the virtual machine concept for creating:

1) An environment for increasing the effectiveness of researchers who must use analytical, modeling systems and have complex data management needs.

2) A mechanism for multi-user coordination of access and update to a central data base.

3) A mechanism for creating an environment where several different modeling facilities can access the same data base.

4) A mechanism for creating an environment where several different and potentially incompatible data management systems can all be accessed by the same user models or facilities.

The paper investigates and formalizes the performance implications of this scheme specifically directed at the question of response time degradation as a function of number of virtual machines, of locked time of the data base machine, and of query rate of the modeling machine.

1. Introduction

Many applications demand both a very good analytical and modeling capability, as well as a flexible data base management capability. They demand that the capability be on-line and interactive. These demands are particularly acute in information systems for assisting public policy decisions and in particular we have found in the area of energy [Donovan: 1975; MacAvoy: 1974]. Such systems have a spectrum of users ranging from the non-technical to the researcher to the computer professional. Each grouping demands a different level of detail in capabilities. Further such systems have:

- a need to build models quickly.
- a need to place complex protection rights on data.
- a need to validate data.
- a need to access data according to any number of criterion.
- a need for mechanisms for changing the system to meet new demands and different data series and needs.
- a need to handle all types of data.

Modeling systems like TROLL [TROLL: 1972], EPLAN [Schober: 1974], and TSP [Hall: 1975] provide flexible analytical capabilities such as sophisticated statistical methods, arithematic operation, plots, graphs, histograms and facilities for constructing and executing mathematical models. All of these have some short comings but the most serious shortcoming is in their limited data management capabilities. There are very limited facilities for protecting data, storing different types of data, changing the structure of data or tables in the system, validating data, quering data by specifying different conditions. Some of these facilities are single user non interactive systems. None allow multiple users accessing the same data base.

Corresponding there exist data management systems like IMS, DBDG, ENQUIRE, TOTAL which provide some degree of data manipulation capabilities but are seriously lacking in analytical or modeling capabilities. They also lack the flexibility in use, access, and protection of data demanded by some applications [Jacoby: 1975]. They do however have considerably more data capability than the modeling systems previously mentioned. This lack of flexibility is a particularly damaging limitation in the context of the certain applications for several reasons:

1. Since unforseen uses and needs for the data inevitably arise, the system must be flexible so that it can adapt to these changing needs. This is particularly true when providing information for policy decisions in so volatile an area as energy.
2. There are varying constraints imposed by changes in the quality, availability, and protection requirements of data. The system must be able to adjust to such moving constraints.
3. The system must be able to accommodate changing needs and constraints at reasonable expenditures of cost and effort. Computer systems of a decade or two ago could support most current applications, but in many cases, only at a high cost.

A flexible system makes it possible to easily experiment with many uses of the data at modest costs.

We have developed a very flexible data management system called TRANSAC [Donovan & Jacoby: 1974] that meets these criteria. The purpose of this paper is however not to promote any one modeling system or data management system but rather to present a scheme whereby the good features of any system can be best utilized.

2. Interfacing modeling facilities and data base facilities

Let us explain a scheme whereby we could interface a modeling system e.g. TROLL, to a data base system.

For conceptual purposes, let us just speak of two separate machines, one at Yale which is running TROLL under Yale's operating system and one at M.I.T. which is running the data base system under M.I.T.'s operating system.

The interface scheme would be whenever the Yale machine needs data, it would request a courier to run to M.I.T. and get the data out of the data management machine. The courier would then bring the data back for the modeling machine.

3. Problems with interfacing

Starting with the scheme of using two independent computer systems, let us evolve into a proposed viable scheme which we advocate.

1. Many modeling facilities are single user non-interactive batch oriented (e.g. TROLL is single user, IBM's TSP is batch oriented). A multiuser interactive facility is desirable.

   - Solution: place each modeling facility on a separate machine.

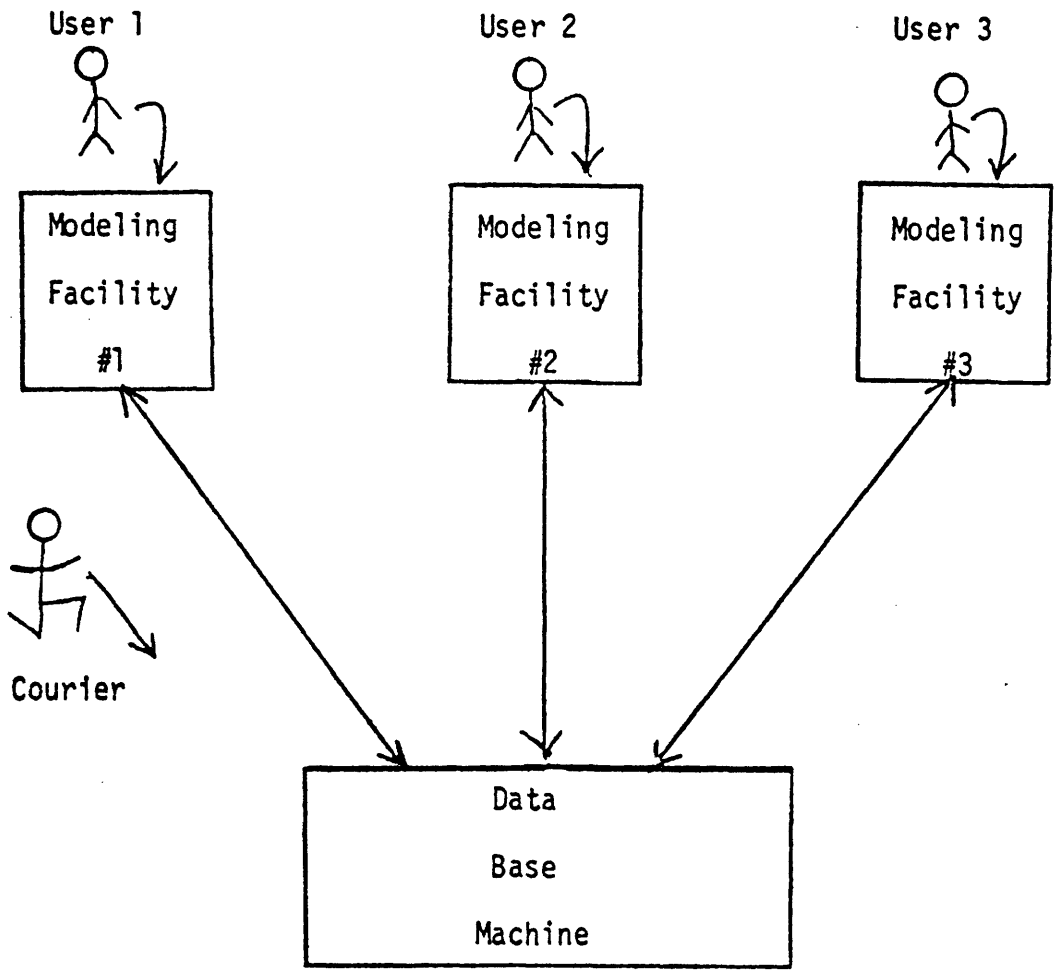


Figure 1
Multiple Users of the Same Data Base

2. We would like more than one use (modeler) to be able to access the data base at one time.
   - Solution: Allow many machines to communicate with the same data base machine as in Figure 1.

3. The solution to 2 creates the problem of coordination of updating the single data base.
   - Solution: Only one modeling system will be serviced by the data base machine at one time.

4. Not every user will want the same modeling facility; some will want TSP; others, TROLL, etc.
   - Solution: One solution is to require all users to convert and all existing models be redone in one modeling language. Another solution is to run a courier between machines that have different modeling capability on them and the single data base machine as in Figure 1.

5. Data series may already exist in several and incompatible data base management systems. How can a user access these data series.
   - Solution: Interface machines that have different data base systems as in Figure 2.

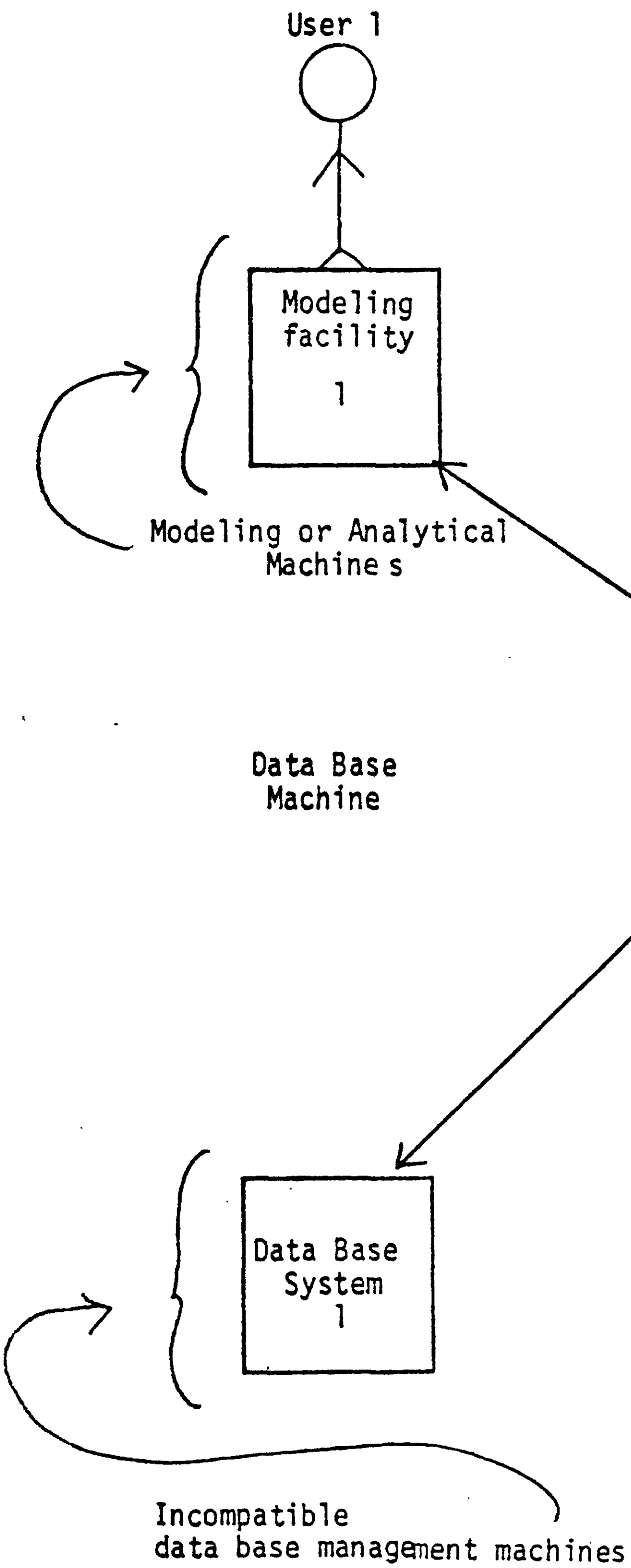

User 1
User 2
User 3
Modeling facility 1
Modeling facility 2
Modeling facility 3
Modeling or Analytical Machine s
General Data Base Machine
Data Base Machine
Data Base System 1
Data Base System 2
Data Base System 3
Incompatible data base management machines

Figure 2

6. The cost of many separate machines is high.

   Couriers between all these machines are slow and not practical.

   - Solution: Have all these machines run on the same machine, that is, have one machine simulate several machines (virtual machines). On some of these virtual machines, run the modeling facilities; on others run the data base facilities; on one run the general data base facility. What about communication? This will be discussed in Section 5.

7. What about performance?

   We discuss this in Section 6.

4. Description of Virtual Machine Concepts

A virtual machine may be defined as a replica of a real computer system simulated by a combination of a Virtual Machine Monitor (VMM) software program and appropriate hardware support. (See [Goldberg: 1973] for a more precise definition.) For example, the VM/370 (IBM 72) system enables a single IBM

System/370 to appear functionally as if it were multiple independent System/370's (i.e., multiple "virtual machines"). Thus, a VMM can make one computer system function as if it were multiple physically isolated systems. A VMM accomplishes this feat by controlling the multiplexing of the physical hardware resources in a manner analogous to the way that the telephone company multiplexes communications enabling separate and, hopefully, isolated conversations over the same physical communications link.

By restricting itself to the task of multiplexing and allocating the physical hardware, the VMM presents an interface that appears identical to a "bare machine". In fact, it is usually desirable to load a user-oriented operating system into each virtual machine to provide the functions expected of modern operating systems, such as Job Control Language, command processors, data management services, and language processors. Thus, each virtual machine is controlled by a separate, and possibly different, operating system. The feasibility of this solution has been demonstrated on the VM/370 system and the earlier CP-67 and CP-40 systems

In addition to VM/370 and its predecessors, several other operational virtual machine systems have been developed, such as the DOS/VM of PRIME Computer, Inc. [PRIME: 1974], the virtual machine capability provided under the Michigan Terminal System (MTS) [Morrison: 1973], and a virtual machine system for a modified PDP-11/45 used by UCLA for data security studies [Popek & Kline: 1974].

The VMM concept, once understood, is quite simple and logical. Unfortuntely, it is sufficiently different from most conventional operating systems that many people have difficulty in understanding the concept. The papers [Buzen: 1973, Goldberg: 1973, Hogg: 1973, Madnick: 1969, Parmellee: 1972, and Madnick & Donovan: 1974] give additional insight.

At first the idea of replicating the bare machine interface may seem foolish since you end up back where you started. The key difference is VM/370 produces the effect of multiple bare machines. In this way each user appears to have his own 370 computer. Thus, each user can select the operating system (e.g. OS/360 DOS, etc.) of his choice to run on his "private" computer.

How does VM/370 produce this feat? How do the users of VM/370 communicate with it? Programs running under VM/370, usually operating systems physically execute in problem state but can behave as if they were in supervisor state. When they issue a privileged instruction, such as START I/O or SET STORAGE KEY, an interrupt occurs and control transfers to VM/370. The interrupt is handled in such a way that the program thinks that the privileged instruction was actually executed. Thus, these privileged instruction interrupts are the subtle interfaces between users and VM/370.

Additional advantages of VM are outlined in [Buzen: 1973] and [Madnick & Donovan: 1974, 1975].

5. Use of VM in Information Systems

As was discussed in Section 4, having multiple machines gives the effect of having multiuser modeling facilities which can access data stored in several different data bases. Proposed communication between all these was via courier. Another possibility and the scheme we advocate is to simulate several different machines on one machine using the VM concept. This section discusses the implications and mechanics of this possibility.

Combining the solutions of the previous section, we could, for example, create a configuration of VM's whose architecture could be depicted as in Figure 2, where each box denotes a virtual machine.

5.1 Communications between VM's

Configuring several VM's on one real machine as in Figure 2 allows several modeling systems to access data from a single data base management system. When a modeling facility issues a request for data, that request is output on a virtual card punch and sent to the data management machine's virtual card reader. The data management machine reads the request, selects the data, and transfers the data back to the modeling facility via the transfer of data from the data management virtual punch to the modeling facility's virtual reader.

Note that no (physical) cards are involved in this process. The "card files" which are punched and read, are in fact stored on (physical) disks for the transfer.

The amount of reprogramming and design involved in modifying the data base management system DBMS to accept requests and output data to its virtual card devices is relatively small, compared to the amount of work and/complexity that would be involved in rewriting the modeling system to include a facility for data handling, for multiusers, for interactive editing, for synchronization of data base access and updating.

Since all modeling facilities have mechanisms to store data in files and facilities to operate on this data, the modification to a modeling system under the VM scheme consists of adding three commands:

- adding a command to convert the data outputted from the DBMS into the format that the modeling facility uses.

By adding two more commands, a modeling system which has very poor data management capabilities can appear to a user as if he had a very powerful facility for storing, quering, updating, and manipulating data.

- adding a command that has as possible arguments the commands of the data base system. The modeling system "passes" the command on to the data base machine via virtual cards.
- adding a command which prints data passed back to the modeling facility.

This scheme will also work with most data base systems, as most of them have (or it is easy to add)

a mechanism for reading request in from files or cards and outputting results to cards or files.

5.2 Multiuser Coordination

The basic problem with having multiple users of the same data base is how to prevent race conditions and uncertainties resulting from several users accessing and updating the same data base. A mechanism we advocate is to have the data base virtual machine only allow one user to access or update its VM at one time. Thus, whenever the data base virtual machine is processing a request, it queues all other requests. The queue is serviced on a FIFO basis.

The performance implications of this approach have not been experimentally tested. A mathematical analysis of the performance is presented in Section 6.

5.3 Multiple Modeling Interfaces

Adding the commands outlined in the previous section to other modeling facilities and running each of these different modeling facilities in a separate VM allows several different modeling facilities to communicate with the same data base. Thus, incompatible systems, such as TROLL and EPLAN, can work from the same data base.

5.4 Incompatible Data Management System

Let us suppose that there is a need to create a DBMS that uses data from several data bases, each of which is on an incompatible data base system. We reject copying all data bases into one data base system because, for example, the existing DB systems may be specialized to keep the data up to date. Thus, how can we treat these four physically separate data bases as one logical unit?

A solution to this problem is also shown in Figure 2, where we could configure three virtual machines to allow the mutually incompatible data base management systems to run on the same physical computer. We then implement another VM to act as an interface, analyzing the data query and funneling it to the appropriate DBMS (via virtual card files). All of these mechanisms can be made invisible to the user, who can use the system as though he had all the data in one "virtual" data base.

Note the "user" in this sense can be a modeling facility or a person, i.e., a user here is anything that makes a data request.

5.5. A Practical Example

We have configured a cluster of VM as in Figure 2 to produce a total system for research in energy policy analysis. We call the system GMIS (General Management Information System). Figure 3 depicts the ultimate GMIS system [Donovan et al: 1975], where across the top several modeling or analytical systems are depicted as running on separate virtual machines. Note that each of these analytical systems may be running under a different operating system, e.g., TSP running under MVT, TROLL running under CP/CMS, EPLAN running under VSZ. TRANSACT [Donovan and Jacoby: 1975] is a data base system based on the relational model of data [Codd: 1970] and uses some IBM software [Chamberlain: 1974]. TRANSACT is implemented in a hierarchical fashion [Dijkstra: 1968, Madnick: 1970, Donovan: 1972], and as such it is a very flexible and powerful data management system. Across the bottom of Figure 3 are depicted several data base systems, each of which may be incompatible and running under different operating systems.

Note that in this paper, independently of any one data base system, we are advocating the use of VM to produce an environment where multiple analytical machines can be used on the same facility and these analytical systems have access to data base systems.

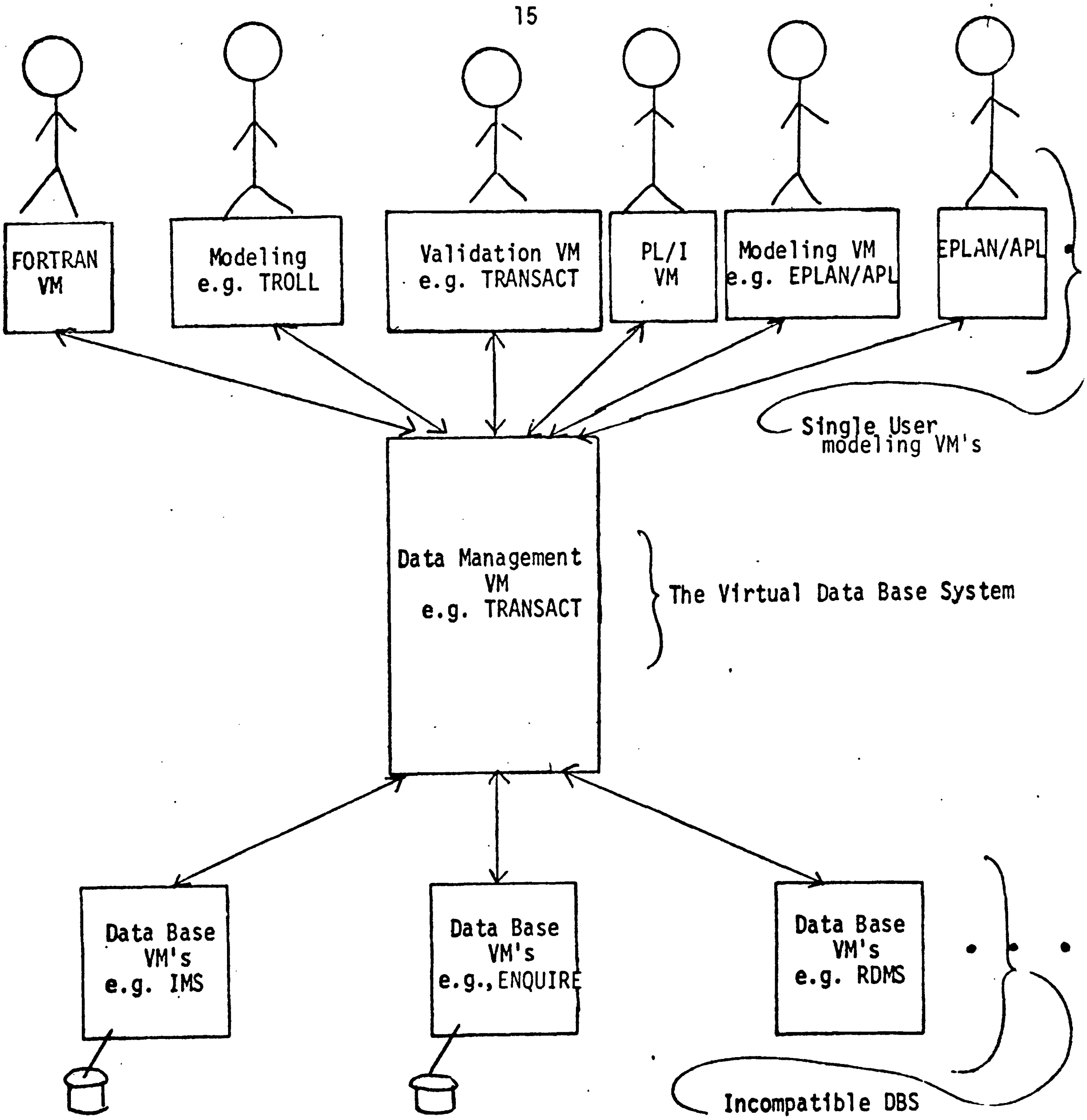

FORTRAN
VM
Modeling
e.g. TROLL
Validation VM
e.g. TRANSACT
PL/I
VM
Modeling VM
e.g. EPLAN/APL
EPLAN/APL
Single User
modeling VM's
Data Management
VM
e.g. TRANSACT
The Virtual Data Base System
Data Base
VM's
e.g. IMS
Data Base
VM's
e.g., ENQUIRE
Data Base
VM's
e.g. RDMS
Incompatible DBS

Figure 3

6. Performance

Not only does the VM approach solve all the problems of Section 3 but it also has the following cost benefits:

- no conversion cost in bringing up existing models as long as they run on an IBM machine (independent of lanuguage or operating system).
- no retraining cost involved as programmer's may use whatever system they are familiar with.
- little cost involved in implementing the simple interfaces.

What is the possible disadvantage - performance , which is reflected in additional overhead costs: For example, the following questions arise:

- How many users (modeling machines) can use the same data base machine? That is, what is the degradation of response time as a function of the number of modelers?
- What is the degradation cost due to the synchronization mechanism?
- What is the degradation cost due to VM?

We have separated the two performance costs:

(1) due to lock out synchronization mechanisms and

(2) due to VM overhead.

The approach to answering these questions we take here is an analytical one. We will first analyze the performance issues of lock out by configuring a system of separate real machines. We then analyze the cost of VM by configuring the separate real machines or as virtual machines on one real machine. Other approaches to gain other factors of performance in VM are discussed in [Hatfield: 1972, Goldberg: 1974].

7.1 Analysis as Separate Machines (performance degradation due to lock out)

Assuming a configuration as in figure 1, where several modeling facilities each running on a separate real machine are accessing and updating a data base which is managed by a data base management system running on its separate real machine. What is the degradation of performance with each additional user? What is it as a function of the length of time the DB machine takes to process a request?

An access or update to the DB machine may be initiated either by a query from a person which would be passed on by the modeling machine or by a model executing on the modeling machine.

In either case, the DB machine while processing a request locks out (queues) all other requests. Let us write a function that specifies total response time of a model.

$$T_{total} = T_{overhead} + T_{model} + T_{request\ and\ wait}$$

where

T_{total} = total response time of a task, (e.g., a model) that is, total time from the start of execution of a model to the answer.

$T_{overhead}$ = amount of CPU time spent executing instructions in the operating system of the modeling machine.

T_{model} = amount of CPU time executing the instruction associated with the model.

$T_{request\ and\ wait}$ = time modeling machine waits for request to be processed plus time spent waiting for request to be serviced by the DB machine.

What one would want to know is what happens to T_{total} as a function of the number of users. That is, how many users can we tolerate on the system.

Assume that:

(1) a configuration of separate real machines as in figure 1.

(2) the time spent in executing the model in a modeling machine before issuing a request for data to the DB machine is negative exponentially distributed with mean $1/\lambda$

(3) the time for the DB machine to serve a request is negative exponentially distributed with mean $1/\mu$

(4) the order of service at the DB machine is FIFO

(5) the number of modeling machines is m

We can formualte the probelm as a machine-repairman model [Satty: 1961] as shown in Figure 4. The steady state equations are:

$$\left.\begin{aligned} m\lambda P_0 &= \mu P_1 \\ [(m-i)\lambda+\mu]\, P_i &= (m-i+1)\lambda P_{i-1} + \mu P_{i+1} \end{aligned}\right\} \text{ for } 0 < i < m$$

$$\mu P_m = \lambda P_{m-1}$$

Where P_i is the steady-state probability that there are i modeling machines waiting and being served.

The solution is:

$$P_i = P_o \left(\frac{\lambda}{\mu}\right)^i \quad \binom{m}{i} i\,!$$

$$= P_o \left(\frac{\lambda}{\mu}\right)^i \frac{m!}{(m-i)!}$$

where $i = 1, \cdots, m$

where,

$$P_o = \left(\sum_{i=o}^{m} \frac{P_i}{P_o}\right)^{-1} =$$

$$= \left(\sum_{i=o}^{m} (\lambda/\mu)^i \binom{m}{i} i!\right)^{-1}$$

$$= \left(\sum_{i=o}^{m} (\lambda/\mu)^i \frac{m!}{(m-i)!}\right)^{-1}$$

The average response time for a request to DB machine as derived by [Little: 1961] is:

$$R = \sum_{i=i}^{m} \frac{i\,P_i}{\mu\left(1-P_o\right)}$$

Figure 5 illustrates the wait and process time for a single request as a fundtion of modeling machines. For example, with $\frac{\lambda}{\mu} = 1$, five users on the system degrades the response time of each user by a factor of four. With a $\frac{\lambda}{\mu}$ ratio of less than .1 there is almost no degradation of response until a large number of users are using the system

Note: In all the remaining graphs μ is set at a constant value of 1.0, and N (number of data requests) is a constant 10. The values of $T_{overhead}$ is a constant equal to 1.0.

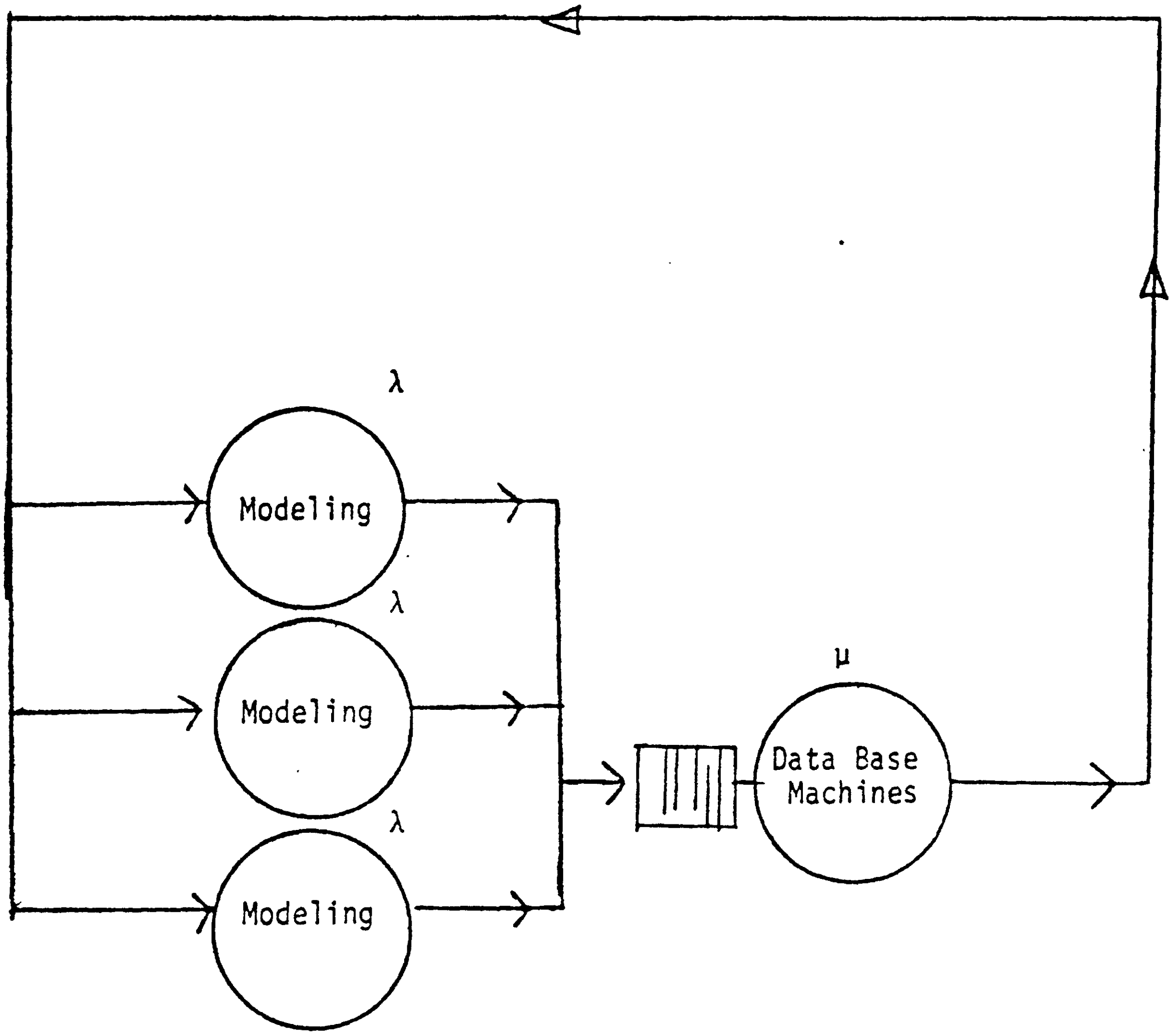


Figure 4

Model of Figure 1

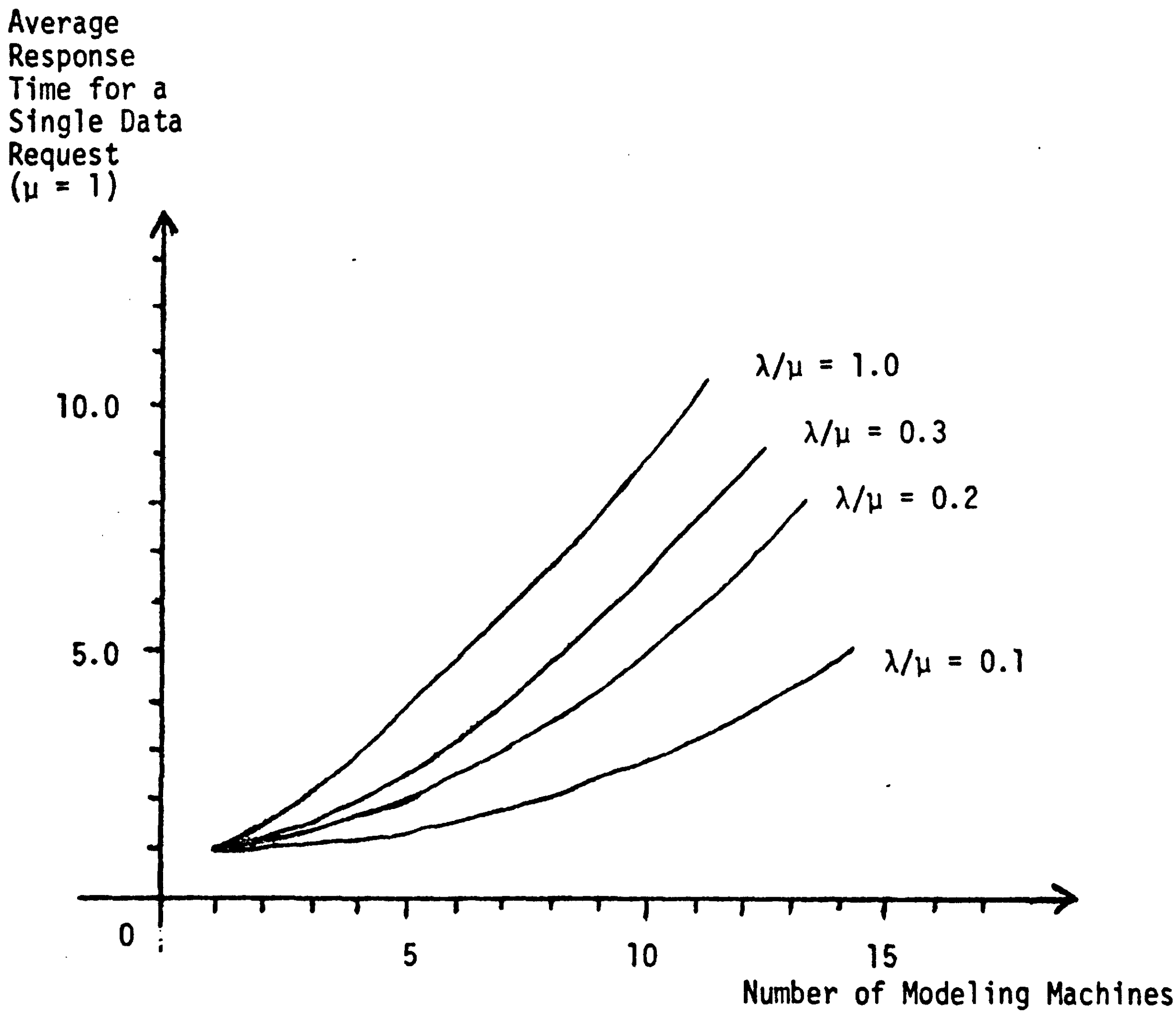


Figure 5

Response Time for a Single Request

Assume that the average number of data requests to DB machine in running a model in a modeling machine is N. The data base is locked only while a request is being processed. We are assuming there is no reason to lock the data base for the whole period while a model is running. The situation where a data base must be locked for the entire period of execution (e.g., a possible danger that other modeling machines will change sensitive data in between requests) requires another anlaysis.

The total time waiting for data from the Data Base machine is:

$$T_{\text{wait for data}} = N \cdot R$$

The average time spent in executing a model in a modeling machine is a constant:

$$T_{\text{model}} = N \cdot (1/\lambda)$$

The overhead of the operating system of one modeling machine is fixed and is equal to a constant T_{overhead}. The total time to execute a model in the modeling machine is:

$$T_{\text{total}} = T_{\text{overhead}} + T_{\text{model}} + T_{\text{wait-for-data}}$$

and is plotted in Figure 6.

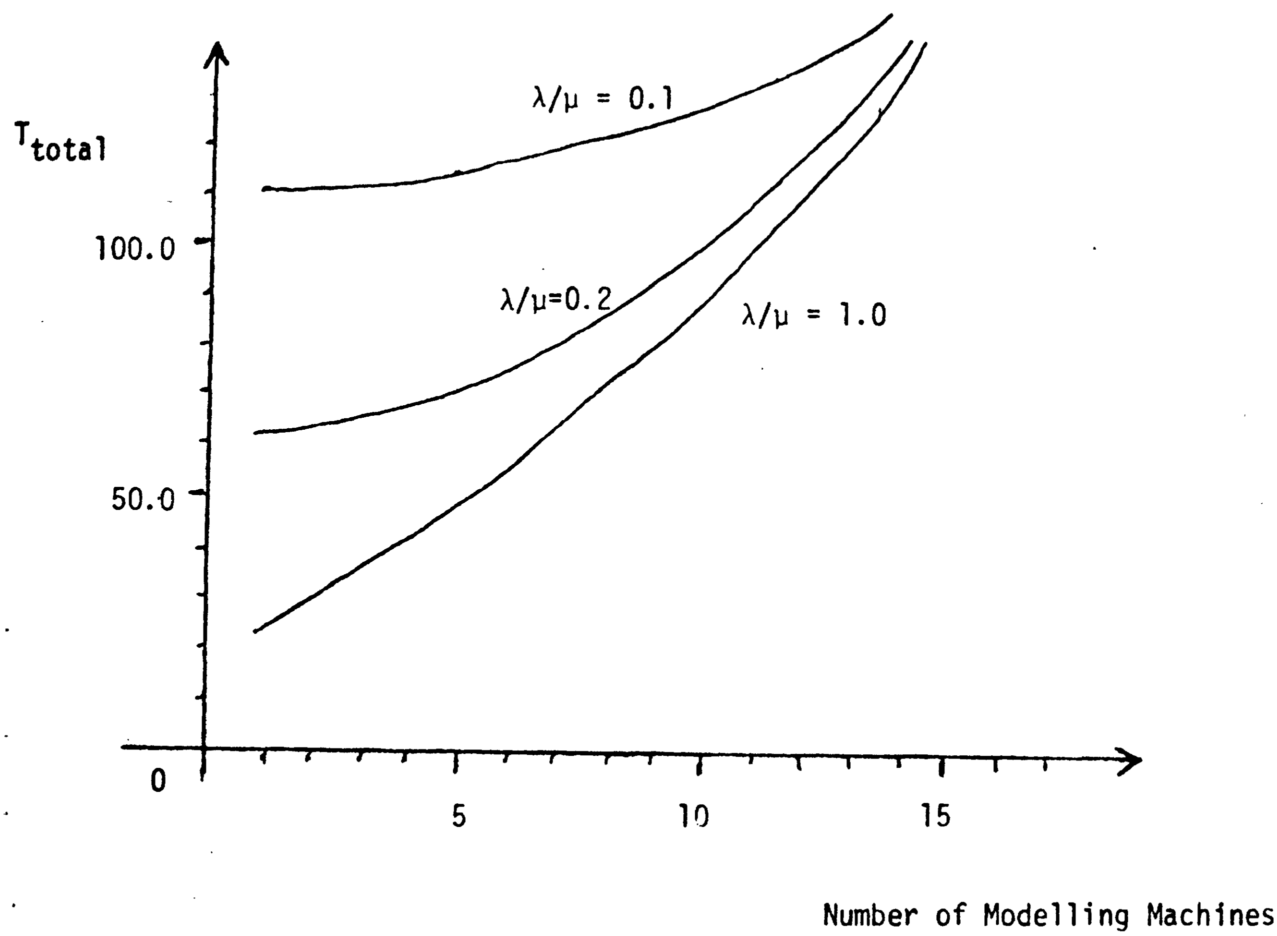


Figure 6

Total Response Time

6.2 Analysis if All Machines are VM's on One Real Machine

If all machines are run as virtual machines on one real machine, what is the additional degradation of response time?

In the VM configuration actually the real machine spends a small portion of its time on each VM. As the number of VM's increase, then each VM will get less of the real CPU's time thus further increasing the elapse time between the start of a model and the production of the answer.

The analysis is further complicated by the fact that as some VM's become locked then others get more of the real CPU's time, therefore, they generate requests faster. However, the DB VM gets more of the CPU's time thereby processing requests faster. For example, if there are ten virtual machines, each one receives one-tenth of the real CPU. However, if seven of the ten are in a locked state, then the remaining three receive one-third of the CPU. Thus, these three run (in real time) faster than they did when ten were running. The following is an analysis of VM's performance for the use outlined in this paper.

We have assumed that the virtual speeds of VM's are constant and equal. However when some VM's are blocked (i.e., waiting for data from the DB VM), the remaining VM's (including DB VM) are allocated a larger share of CPU processing power and became faster in real time. We assume that each unblocked VM receives the same amount of CPU processing power and at the initial state m machines are running (i.e., the data base machine is stopped if no modeling machines are making requests). 'λ'is request rate of each modeling VM when there are m VM's running. 'μ' is the service rate at which the data base virtual machine is running when there are m-1 modeling VM and one data base VM running. Thus, we may write the relations:

$$\mu_i = \frac{m}{m-i+1}\,\mu \qquad (i = 1, 2, \ldots, m)$$

$$\lambda_o = \lambda$$

$$\lambda_i = \frac{m}{m-i+1}\,\lambda \qquad (i = 1, 2, \ldots, m)$$

where i (i = 0.1,...,m) is the number of modeling VM's being blocked. Using a birth/death process model [Drake: 1967], the state transition diagram is shown in Figure 7.

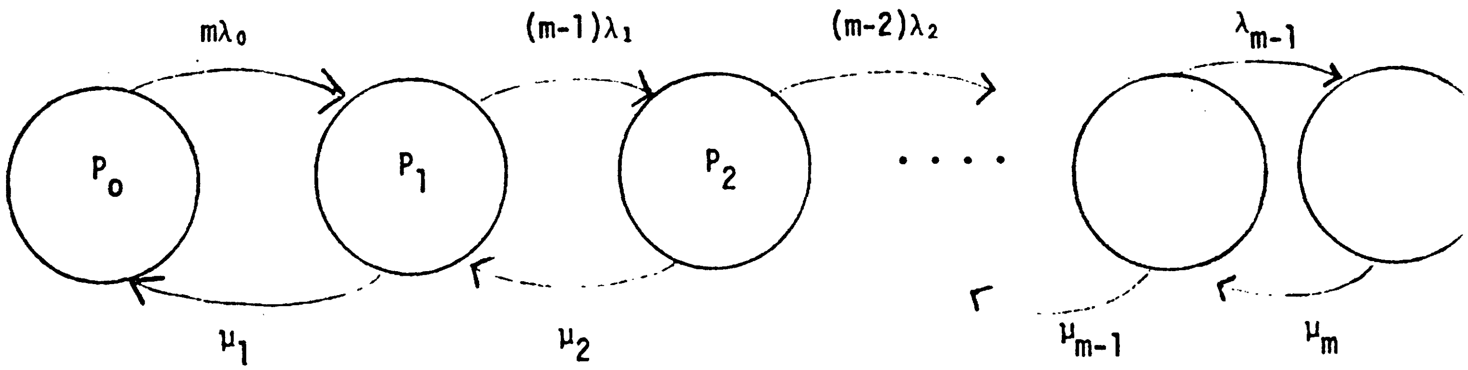


Figure 7

State Transition of Multi-VM Model

From this model, the steady state equations are [Drake 1967]

$$m\,\lambda_0\,P_0 = \mu_1\,P_1$$

$$[\,(m-i)\,\lambda_i + \mu_i\,]\,P_i = (m-i+1)\,\lambda_{i-1} + \mu_{i+1}\,P_{i+1} \qquad i<m$$

$$\mu_m\,P_m = \lambda_{m-1}\,P_{m-1}$$

The solution of the above set of equations is:

$$P_i = \left(\frac{\lambda}{\mu}\right) \frac{(m-1)!}{(m-i)!} \quad (m-i+1) \quad P_o \qquad (i = 1, 2, 3, \ldots, m)$$

where

$$P_o = \left[1 + \sum_{i=1}^{m} \left(\frac{\lambda}{\mu}\right)^i \frac{(m-1)!}{(m-i)!} \quad (m-i+1)\right]^{-1}$$

The average response time for a request to the DB VM in this VM configuration is obtained by generalizing the analysis [Little: 1961] to this situation where there is queue dependency.

$$R' = \frac{\sum_{i=1}^{m} iP_i}{\sum_{i=1}^{m} \mu_i P_i}$$

Figure 8 illustrates the response time of a single request as a function of the number of modeling VM's.

Similar to equation of section 6.1,

$T'_{overhead} = T_{overhead}$

$T'_{wait\text{-}for\text{-}data} = N.R.$

T'_{model} is calculated similarly to the way T_{model} was calculated in section 6.1. That is, $T'_{model} = N \cdot \frac{1}{\lambda}$. However, the λ's vary. Thus we take a weighted sum and get the following. (Note that if λ_i are constant, this reduces to the T_{model} of section 6.1.)

$$T'_{model} = N.\left(\frac{\sum_{i=o}^{m-1} P_i \left(\frac{m-i}{m}\right)}{\sum_{i=o}^{m-1} P_i \left(\frac{m-i}{m}\right) \lambda_i}\right)$$

$$T'_{total} = T'_{overhead} + T'_{model} + T'_{wait\text{-}for\text{-}data}$$

Figure 9 illustrates the total time to execute a model as a function of the number of modeling VM's.

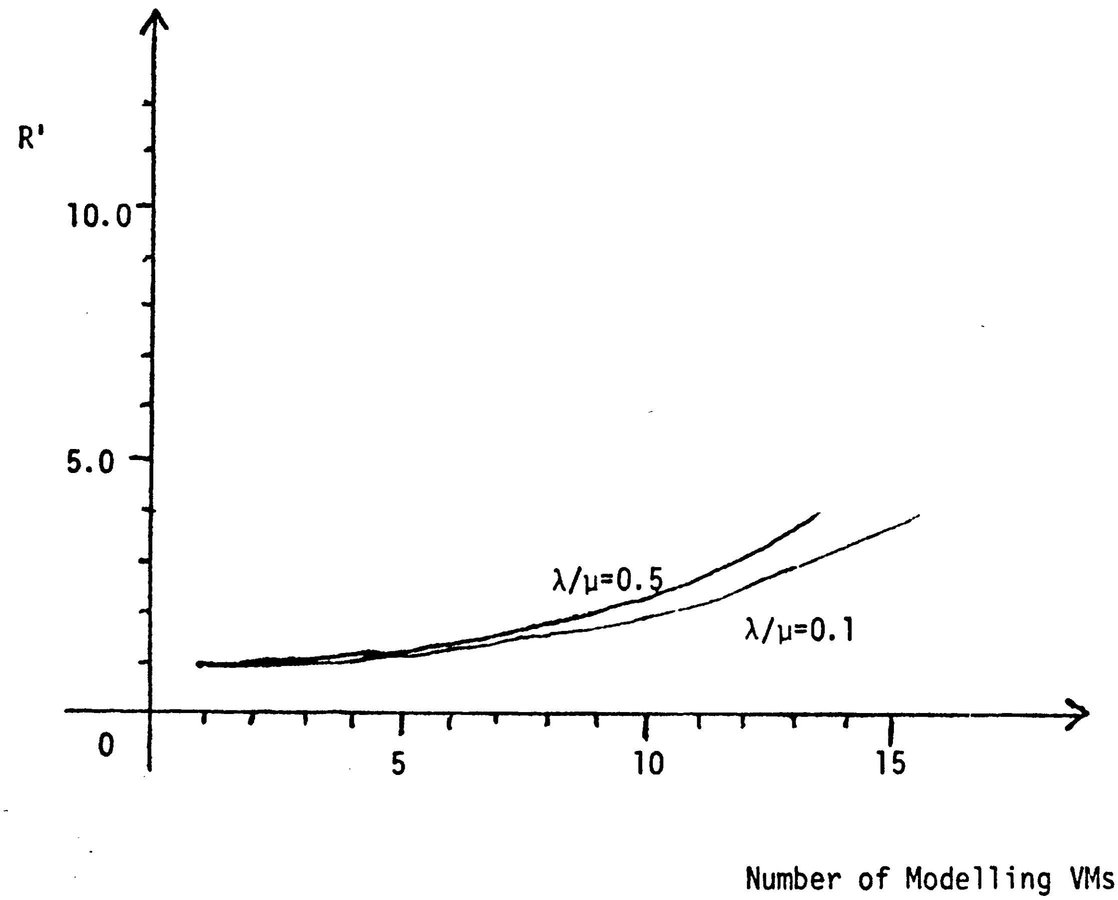


Figure 8

Response Time of a Single Request in a VM Configuration

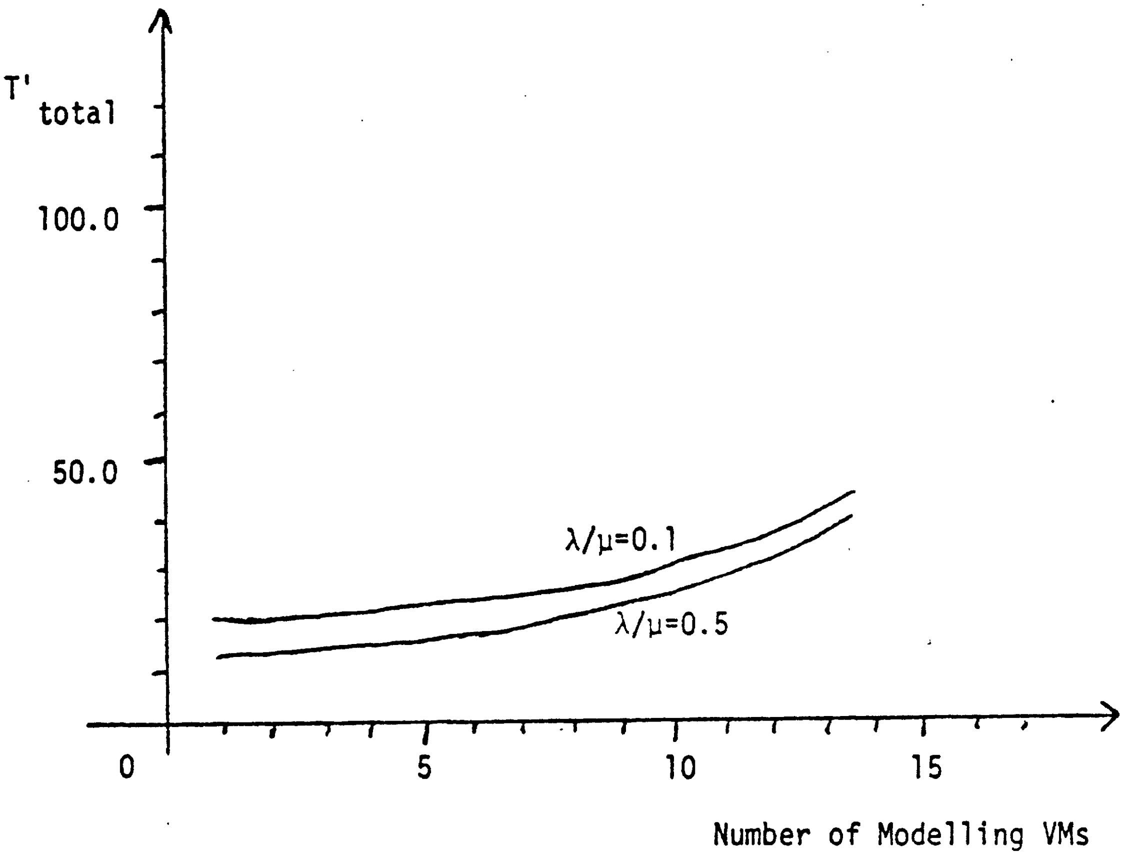


Figure 9

Total Elapsed Time in a VM Configuration

7. Techniques for Reducing Synchronization Overhead

The synchronization of the access and updates to the DB Virtual Machine is accomplished by what we call a spin lock. That is, if the modeling VM encounters a "locked" DB VM, then it must wait (in a queue); until the DB VM is unlocked, the modeling machine cannot do anything else.

The use of locks where the VM's must wait if encountering a lock, as we have seen, has an adverse effect upon system performance. Several techniques may be used to reduce this synchronization overhead, and the relative merits of each must be weighed.

One approach is to use a single lock (as we have done) to cover all shared data bases in the single DB VM. The alternative is to identify all separate data bases carefully and associate a separate lock with each.

There are many factors to be considered in choosing between a precise lock approach (i.e., a large number of separate locks) and an overall lock approach (i.e., one lock for all data bases). In the precise approach, considerable overhead is incurred in setting and resetting locks, even though the particular data base is not needed by any other VM. This multitude of locks also greatly complicates debugging.

In the overall lock approach (also called brute force), the lock may be on for long periods of time (up to 50 percent or more). This greatly increases the likelihood of software lock-out and the resulting slow response time.

8. Techniques for Reducing Effect of VM on Response Time

The basic reason for the degradation of performance due to VM is the fact that one real machine is being used to simulate several VM's. That is, one real CPU spends a little time on VM #1, then on VM #2, then on VM #3 and so forth. Thus, each VM only gets a fraction of real CPU time.

One method of increasing the amount of real CPU each VM gets is to increase the number of real CPU's. That is, use a multiprocessor configuration. Note all processors are executing instructions in the same memory.

The trade off is, the cost of the extra processors and their real effect. That is, each additional processor incurs some overhead and introduces a lower level set of locking problems. The lower level locking problem arises from having to lock "system" data bases whose access and updating must be synchronized (e.g., the system table which keeps track of what process the processor should be assigned to).

Treating each VM as corresponding to a separate process, we may perform a similar analysis [Madnick and Donovan: 1975] to determine the effectiveness of additional CPU's.

9. Summary

Running individual modeling facilities on separate machines all interfaced to a single database machine creates a total facility that is multiuser, interactive, suited to individual tastes and provides access to a single common data base. Simulating all these machines as virtual machines on one real machine provides a mechanism for fast and inexpensive communication between machines.

Multiple use of a single database creates the problem of synchronization of access and updates to that database. The spin lock provides a synchronization mechanism, however, at a performance cost in increased delays in response time. Figure 10 dotted curves give these times assuming separate real machines.

The performance implications of the use of VM can be seen in Figure 10, that is, the degradation because of VM becomes significant with large numbers of VM's.

Response time degradation due to a lock can be improved by partitioning the data base and using more than one lock. Degradation due to overhead associated with VM (one real processor simulating many) may be improved by adding more processors.

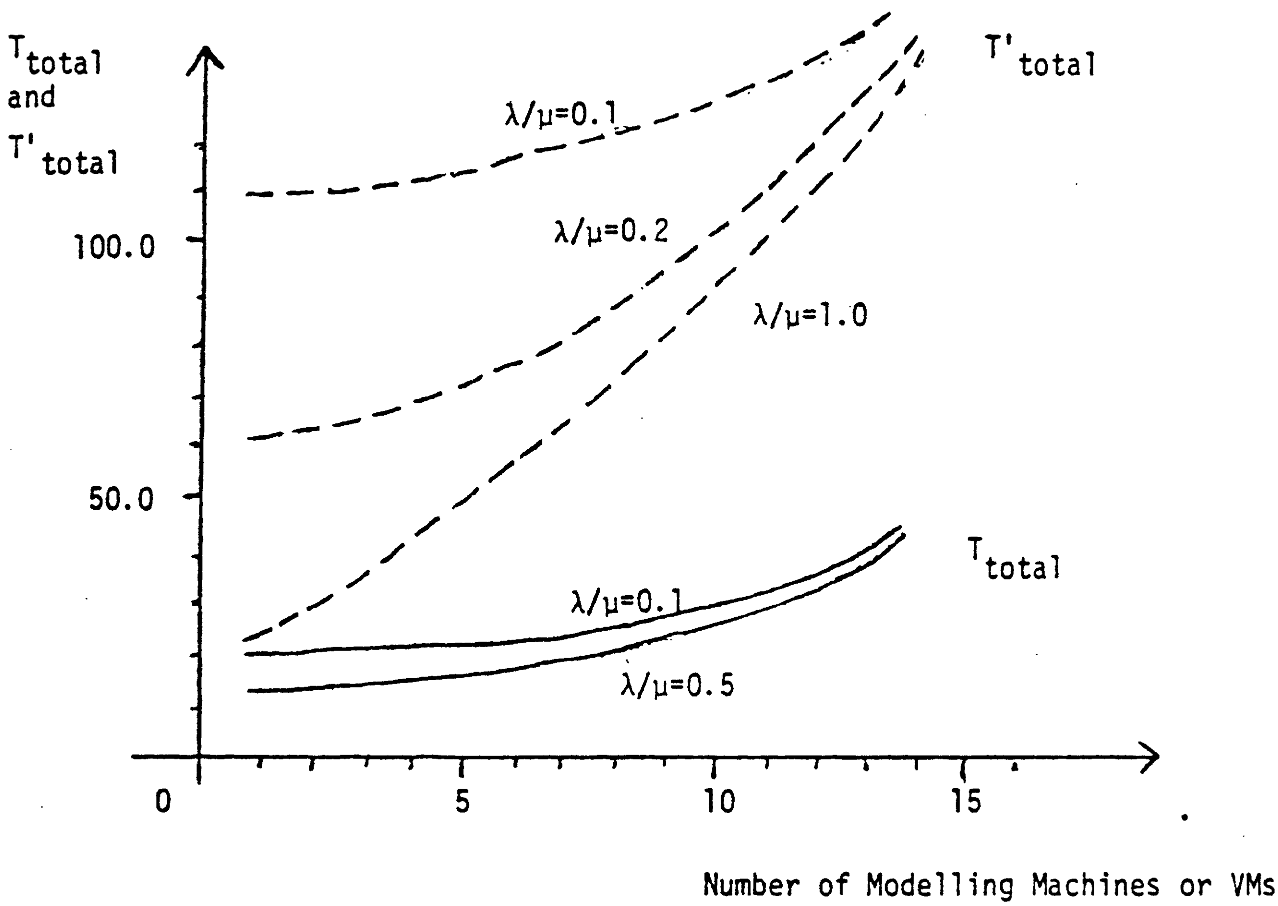


Figure 10

Comparison of Total Elapsed Times for a VM and a non-VM Configuration

Acknowledgment

Want to thank Professor Peter Chen for helping with the analytical contruction and solutions to the performance equations and to Stu Madnick for helping in formulating these equations. We acknowledge the work of Marvin Essrig in applying this VM scheme in developing a system for leading Energy Indicators and in expanding the scheme for creating an environment where several different and potentially incompatible data management systems can all be accessed by the same models or facility.

We acknowledge the assistance of Drs. Stuart Greenberg and Ray Fessel of the IBM Cambridge Scientific Center for their assistance in implementing the scheme here. We acknowledge Louis Gutentag and the MIT students who worked with him for making the system operational.

REFERENCES

1. Buzen, J.P. and U. Gagliardi: "Evolution of VM Architecture", AFIPS Conference Proceedings, National Computer Conference 1975, Vol. 42, pp. 291-299.

2. Buzen, J. P., Peter P. Chen, and Robert P. Goldberg: "Virtual Machine Techniques for Improving System Reliability", Proceedings of the ACM Workshop on Virtual Computer Systems, (March 26-27, 1973).

3. Chamberlain, D. D., and R. F. Boyce: "SEQUEL: A Structured English Query Language," PROCEEDINGS 1974 ACM/SIGFIDET Workshop.

4. Codd, E. F.: "A Relational Model of Data for Large Shared Data Banks," CACM, Vol. 13, No. 6, June 1970, pp. 377-387.

5. Codd, E. F.: "A Data Base Sublanguage Founded on the Relational Calculus," PROCEEDINGS 1971 ACM/SIGFIDET Workshop.

6. Coffman, E.G., Jr. and L.C. Varian: "Further Experimental Data on the Behavior of Programs in a Paging Environment," CACM, vol. 11, No. 7, pp. 471-474.

7. Denning, P.J.: "Virtual Memory," ACM COMPUTING SURVEYS, vol. 2, No. 3, pp. 153-190, September, 1970.

8. Dijkstra, E.: "T.H.E. Multiprogramming System," CACM, May 1968.

9. Donovan, John J.: SYSTEMS PROGRAMMING, McGraw-Hill, New York, 1972.

10. Donovan, John J., and Henry D. Jacoby: "A Hierarchical Approach to Information System Design," Report CISR-5, M.I.T. Sloan School Working Paper 762-75, January 1975.

11. Donovan, John J., Louis M. Gutentag, Stuart E. Madnick, and Grant N. Smith: "An Application of a Generalized Management Information System to Energy Policy and Decision Making: The User's View," Proceedings of Nation Computer Conference, Anaheim, CA, May 1975.

12. Donovan, John J., and Stuart E. Madnick: "Application and Analysis of the Virtual Machine Approach to Computer System Security and Reliability," IBM SYSTEMS JOURNAL, May 1975.

13. Drake, A. W.: FUNDAMENTALS OF APPLIED PROBABILITY THEORY, McGraw-Hill, New York, 1967.

14. Goldberg, R. P.: "Architecture of Virtual Machines," Proceedings AFIPS, National Computer Conference, vo. 42, pp. 309-318, 1973.

15. Goldberg, R. P.: "Survey of Virtual Machine Research," Computer, Volume 7, Number 6, (June 1974), pp. 34-45.

16. Hall, Robert: "TSP Manual," Harvard Technical Report No. 12, Harvard Institute of Economic Research, Cambridge, MA, April 1975.

17. Hatfield, D. J., "Experiments on Page Size Program Access Patterns, and Virtual Memory Performance," IBM Journal of Research and Development, vol. 16, no. 1, pp. 58-66, January 1972.

18. Hogg, J., and P. Madderom, "The Virtual Machine Facility -- How to Fake a 360," Internal Note, University of British Columbia and University of Michigan Computer Center, (1973).

19. IBM, "IBM Virtual Machine Facility/370: Introduction," Form Number GC20-1800, (July 1972).

20. Iverson, Kenneth E., A Programming Language, John Wiley and Sons, 1962.

21. Jacoby, Henry D.: "Final Report, FEA Leading Energy Indicator Project," Energy Lab Working Paper, M.I.T., April 1975.

22. Jacoby, Henry D., and John J. Donovan: "Final Report, New England Energy Management Information Systems," Energy Lab Working Paper, M.I.T., February 1975.

23. Little, John: "A Proof of the Queueing Formula: $L = \lambda\omega$," Operations Research 9, (1961), pp. 383-387.

24. MacAvoy, P., and R. Pendyke: "A Model for Natural Gas," Bell System Journal, 1975.

25. Madnick, S. E.: "Design Strategies for File Systems," M.I.T. Project MAC TR-78, October 1970.

26. Madnick, S. E.: "Time-Sharing Systems: Virtual Machine Concept vs. Conventional Approach," Modern Data 2, 3 (March 1969), pp. 34-36.

27. Madnick, Stuart E., and John J. Donovan: Operating Systems, McGraw-Hill, New York, 1974.

28. Morrison, J. E.: "User Program Performance in Virtual Storage Systems," IBM Systems Journal, vo. 12, no. 3, pp. 216-237, 1973.

29. Parmelee, R. P., T. I. Peterson, C. C. Sullivan, and D. S. Hatfield: "Virtual Storage and Virtual Machine Concepts," IBM Systems Journal, vol. 11, no. 2, pp. 99-130, 1972.

30. Popek, G. J., and C. Kline: "Verifiable Secure Operating Systems Software," AFIPS Conference, 1974 NCC.

31. PRIME Computer, Inc., "DOS/VM Reference Manual," (1974).

32. Schober, F., EPLAN- An APL-based Language for Econometric Modeling and Forecasting, IBM Philadelphia Scientific Center, 1974.

33. Smith, Grant N., "Internal Intermediate Language, Version 2," M.I.T. Sloan School Report CISR-6, November 1974.

34. Satty, T. C.: Elements of Queueing Theory, McGraw-Hill, New York, 1961.

35. TROLL REFERENCE MANUAL, National Bureau of Economic Research, 1972.

Code for a General Purpose System Identifier and Evaluator (GPSIE)

DAVID W. PETERSON AND FRED C. SCHWEPPE, SENIOR MEMBER, IEEE

Abstract—The modeling process may be viewed as a three-step iteration: 1) hypothesize a structure, 2) estimate (identify) unknown parameters, and 3) test for consistency between the model and available data. This paper describes a new, publicly-available computer program which performs the second and third tasks. The program, called General Purpose System Identifier and Evaluator (GPSIE), can handle nonlinear, time-varying, multiple input-output systems of arbitrary dimensions. The user supplies an array of data and a subprogram in PL/I or Fortran defining the model structure of interest. GPSIE searches for the maximum-likelihood estimates of any unknown parameters, and computes statistical measures of consistency between the model and the data. Options allow the user to deal efficiently with many kinds of systems.

I. Purpose of GPSIE and What it Does

THE modeling process in engineering and social sciences may be viewed as a three-step iteration:

Step 1: Hypothesize a structure.
Step 2: Estimate (identify) unknown parameters.
Step 3: Test the model for consistency with data.

If a model fails Step 3, the modeler returns to Step 1 to hypothesize a new structure. This paper describes a new, publicly available computer program, called the General Purpose System Identifier and Evaluator (GPSIE). The modeler may use GPSIE to perform Steps 2 and 3 of the above iteration, and related subtasks (see Table I). The next paragraphs relate the three steps of the modeling iteration to the methods implemented in GPSIE.

Step 1—Model Structures: Many of the model structures currently hypothesized in engineering and the social sciences fall into the class of nonlinear, stochastic, time-varying, multiple input-output systems, of the form:

$$x(n) = f(x(n-1), u(n), w(n), n), \qquad n = 1,2,\cdots,N$$

$$z(n) = h(x(n), u(n), v(n), n), \qquad \text{for some subset of the above } n\text{'s}$$

where the notation is of the standard type, in which the vector $z(n)$ is the nth sample of data, $x(n)$ is the state of the system, $u(n)$ are exogenous inputs or controls, $w(n)$ and $v(n)$ are Gaussian, white processes, and f and h are

Manuscript received February 19, 1974; revised August 1, 1974. This work was supported in part by the National Science Foundation under Grant GI-39150.

D. W. Peterson is with the Department of Electrical Engineering and the Energy Laboratory, Massachusetts Institute of Technology, Cambridge, Mass. 02139.

F. C. Schweppe is with the Department of Electrical Engineering, the Energy Laboratory, and the Electrical Power Systems Engineering Laboratory, Massachusetts Institute of Technology, Cambridge, Mass. 02139.

TABLE I
USES OF GPSIE

Parameter Estimation (System Identification):
Estimate all parameters in a model, including characteristics of driving and measurement noise.
Estimate some parameters, taking others as known *a priori*.
Confidence Tests:
Test for consistency between model and data.
Choose among alternate models (hypothesis testing).
Feasibility:
Determine if the estimation of parameters in a model is possible, using given data (related to econometric identification).
Determine the kind and amount of data needed to estimate unknown parameters (via experimentation with simulation data).
State Estimation and Simulation:
Estimate unknown inputs and the state trajectory (straight Kalman filtering).
Simulate the model (deterministic or noise-driven).

nonlinear vector functions. As is well known, the above form is quite general; for example, the following kinds of systems and special cases can usually be reduced to it [1]:

Autoregressive and moving-average models.
Static systems.
Sensor dynamics.
Nonwhite, non-Gaussian noise.
Perfect observations.
Unknown inputs.
Uncertain or unknown initial conditions.

Step 2—Identification (estimation of parameters): Unknown parameters (and to some extent, structure) in such models may be estimated by 1) computing the log-likelihood function with an extended Kalman filter, and 2) using nonlinear programming algorithms to search for the parameter values which maximize the likelihood function [1]–[5].

Step 3—Consistency and Confidence Tests: The consistency between data and a model may be estimated by computing the whiteness of the residual (innovation) process of the extended Kalman filter [1], [6], [7], the covariance of the estimated state, the information matrix and its inverse, and by comparing the size and shape of the likelihood function with theoretical expectations [1], [8], [9].

The increasing popularity of the above methods has resulted in many special-purpose, one-application computer programs [2], [10]. To aid our own work, and possibly that of others, we tried to write a general-purpose, user-oriented program, flexible enough to handle a wide variety of systems, and powerful enough to handle the general case. The result, GPSIE, allows the user to concentrate more on his specific problem and less on the debugging of filter and search codes. The remainder of this paper sketches the overall structure of GPSIE and lists some of its features.

II. Structure of GPSIE

GPSIE is a precompiled program which consists of 1) matrix equations for computing points of the likelihood function, 2) several nonlinear-programming algorithms for maximizing the likelihood function over the space of unknown parameters, and 3) control logic for handling input-output, options, and special cases.

For each model, the user writes a subprogram which describes the model of interest, its dimensions, and its linearization. The user subprogram may be written in either PL/I or in Fortran. The user subprogram is compiled, linked with the precompiled GPSIE, and loaded. GPSIE then accesses the user subprogram to learn the dimensions of the model and data, the initial guesses of any unknown parameters, and user options (choice of search algorithm, stopping-rule parameters, etc.). From then, on, the user program and GPSIE interact as shown in Fig. 1.

III. Features and Options

Search Options

Numerical maximization of the likelihood function often requires versatility of approach. GPSIE includes as options Newton–Raphson, Gauss–Newton, Davidon–Fletcher–Powell, Marquardt, and Powell algorithms, as well as a manual search, in which the user specifies the sequence of parameter values whose likelihood is to be evaluated.

Least-Squares Initialization

It is often helpful to first approximate the maximum-likelihood solution via a least-squares approach, which is computationally cheaper. GPSIE includes least-squares capabilities as options.

Steady-State Filter

It is sometimes desirable to assume the Kalman filter is in steady state after a given sample, either because it *is* in steady state, or to simplify computation early in a search (as in the least-squares initialization).

Cross-Sectional or Regional Data

Especially in social systems, data is often available from several systems operating in parallel and sharing the same unknown parameters. GPSIE can process such data.

Simulation Capability

GPSIE includes noise generators and control logic for simulating the model of interest and storing the results, toward debugging the user's subprogram or for experiments on the identifiability of the system.

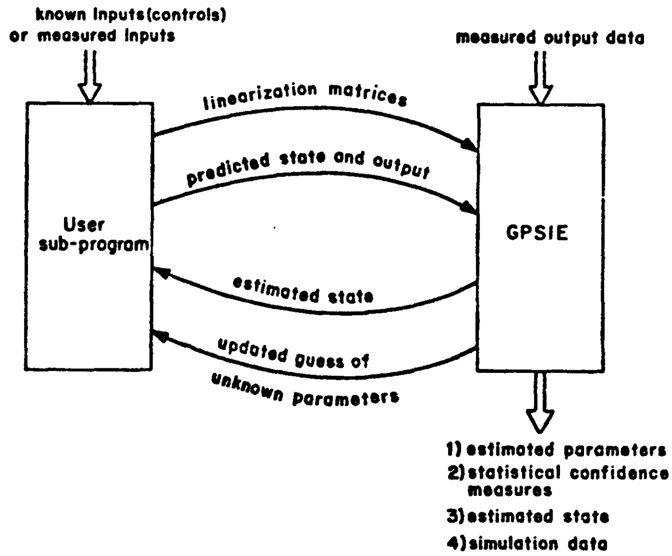


Fig. 1. Interaction of GPSIE with the user subprogram.

Alternate Filters

The Kalman filter is available in GPSIE in two mathematically equivalent forms, one of which allows noiseless measurements to be processed, and the other of which is relatively immune to numerical loss of significance.

Whiteness Test

GPSIE, as an option, computes the normalized correlation matrices needed to estimate the whiteness of the residual process, along with the matrices' expected values and the variance of each element.

Other Statistical Confidence Tests

GPSIE also estimates the information matrix and its inverse, as well as several covariance matrices, for use in determining probable error and the consistency between model and data. In addition, we hope by press time to have included the translation of some of these tests to standard econometric forms, such as R^2 and Durbin–Watson.

Computation of Derivatives

GPSIE can compute both gradient and Hessian of the log-likelihood function by finite differences, in order to handle the most general case. An option is included for an ergodic approximation of the Hessian, which is more efficient, but may work well only for steady-state systems.

Noise in the Measurements

The algorithms of GPSIE can operate under assumptions of

- noiseless measurements;
- measurement noise of known characteristics;
- measurement noise with characteristics taken as unknown parameters.

Partial Measurements

The model may contain variables for which there are no direct measurements. Thus, econometric models processed by GPSIE may contain some variables for which there is no data.

Integration Interval Independent of Data

GPSIE can conveniently integrate the model equations more than once between data points.

Missing Data and Varying Sampling Intervals

The data processed by GPSIE may be distributed unevenly in time and space. For example, some variables may be sampled monthly, others yearly, with partial data or no data at some sample times. Similarly, the data availability may vary from region to region, in the case of cross-sectional or regional data.

A Priori Information on Parameters

Parameters may be
completely unknown;
unknown, with *a priori* mean and variance;
a priori known;
or any combination thereof.

IV. Major Limitations

In spite of its generality, GPSIE has two major limitations, inherent in the use of the extended Kalman filter (in addition to the obvious problems of finding global maxima with hill-climbing techniques).

1) In nonlinear systems, excessive noise in the measurements or initial conditions, or a poorly observable system structure, may allow the estimated state of the system to drift too far from the true state. Under such conditions, the linearization of the filter may become invalid.

2) The variable-dimensioning feature of GPSIE allows it to handle systems of any reasonable size, but requirements of computer time or storage may obviously become extravagent for some systems. However, even relatively large systems may be treated as *a priori* known (no estimation of parameters) and tested for consistency with data.

V. Availability

Subject code of GPSIE, user's manual, and full documentation are available through the authors, c/o the Energy Laboratory, Massachusetts Institute of Technology, Cambridge, Mass. 02139.

References

[1] F. C. Schweppe, *Uncertain Dynamic Systems*. Englewood Cliffs, N.J.: Prentice-Hall, 1973.
[2] R. K. Mehra and J. S. Tyler, "Case studies in aircraft parameter identification," in *Identification and System Parameter Estimation*, P. Eykhoff Ed. New York: Elsevier, 1973.
[3] K. J. Astrom and S. Wenmark, "Numerical identication of stationary time series," presented at the 6th Int. Instruments and Measurements Cong., Stockholm, Sweden, Sept. 1964.
[4] R. L. Kashyap, "Maximum likelihood identification of stochastic linear dynamic systems," *IEEE Trans. Automat. Contr.*, vol. AC-15, pp. 25–34, Feb. 1970.
[5] R. K. Mehra, "Identification of stochastic linear dynamic systems using Kalman filter representation," *AIAA J.*, vol. 9, Jan. 1971.
[6] R. K. Mehra and J. Peschon, "An innovations approach to fault detection and diagnosis in dynamic systems," *Automatica*, vol. 7, 1971.
[7] R. K. Mehra, "On the identification of variances and adaptive Kalman filtering," *IEEE Trans. Automat. Contr.*, vol. AC-15, pp. 175–184, Apr. 1970.
[8] K. J. Åström, "On the achievable accuracy in identification problems," presented the 1967 IFAC Symp.
[9] A. W. F. Edwards, *Likelihood*. New York: Cambridge, 1972.
[10] R. Moore and F. C. Schweppe, "Adaptive control for nuclear power plant load changes," presented at the 5th World Congr. IFAC, Paris, 1972.

David W. Peterson was born in Ashland, Kan., on April 2, 1946. He received the B.S. and M.S. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1969 and 1972, respectively.

From 1969 to 1972 he was on the Research Staff of the Sloan School of Management, M.I.T., applying simulation and control theory to the design of top management policies for industrial corporations, and lecturing on the application and implementation of system dynamics. Since 1972 he has been a Research Assistant in Electrical Engineering at M.I.T., associated with the Energy Laboratory. His interests include modeling of social systems, parameter estimation, filtering, and stochastic control. He has applied the above techniques to areas of corporate management and energy policy.

Mr. Peterson is a member of Eta Kappa Nu and Tau Beta Pi.

Fred C. Schweppe (S'55–M'59–SM'74) was born in Minneapolis, Minn., on November 18, 1933. He received the B.S.E.E. and M.S.E.E. degrees from the University of Arizona, Tucson, in 1955 and 1956, respectively, and the Ph.D. degree in electric engineering from the University of Wisconsin, Madison, in 1958.

From 1958 to 1966 he was with the Lincoln Laboratory, Massachusetts Institute of Technology, Lexington. He is presently Associate Professor of Electrical Engineering, Electric Power Systems Engineering Laboratory, Mass. Inst. Technol., Cambridge. His general research interests are in the areas of central control and system planning for interconnected electric bulk power systems, and economic, environmental, and reliability (security) tradeoffs in the operation and planning of electric power systems.

Dr. Schweppe is a member of Tau Beta Pi, Sigma Pi Sigma, Sigma Xi, Pi Mu Epsilon, and Phi Kappa Phi.

Reprinted by permission from IEEE TRANSACTIONS ON AUTOMATIC CONTROL
Vol. AC-19, No. 6, December 1974, pp. 852-854