ipea
INSTITUTO DE PLANEJAMENTO ECONÔMICO E SOCIAL

Brazilian Economic Studies

5

The Institute of Economic and Social Planning (Instituto de Planejamento Econômico e Social — IPEA) is a foundation established by the Brazilian Federal Government in 1967. Its main activities are related to economic research, planning, government budgeting and training.

IPEA is under the responsibility of a president and comprises a research institute (Instituto de Pesquisas — INPES), a planning institute (Instituto de Planejamento — IPLAN), a budget institute (Instituto de Programação e Orçamento — INOR) and a training center (Centro de Treinamento para o Desenvolvimento Econômico — CENDEC).

The outcome of the work done at IPEA is regularly made available, in Portuguese, in five series of publications. In addition, IPEA issues an economic journal entitled *Pesquisa e Planejamento Econômico.*

The papers published in *Brazilian Economic Studies* refer not only to research conducted at IPEA, but also to work carried out by scholars at other Brazilian institutions.

All correspondence should be addressed to:

IPEA/INPES
Caixa Postal 2.672 (ZC-00)
Rio de Janeiro — BRASIL

INSTITUTO DE PLANEJAMENTO ECONÔMICO E SOCIAL
INSTITUTO DE PESQUISAS — INPES

Brazilian Economic Studies

N.º 5 — Edited by Fernando Rezende

IPEA/INPES. Rio de Janeiro, Brasil. 1979.

IPEA — 061 — 79006 ISSN 0100 — 2910

Brazilian economic studies. — Rio de Janeiro (Caixa Postal 2.672): Instituto de Planejamento Econômico e Social, Instituto de Pesquisas, 1975 (n.º 1) -23 cm.

Leading indicators for the industrial sector *

Claudio R. Contador **

1 — Introduction

Forecasting is crucial to economic policy, since predicting reversals makes it possible for government decisions to partially modify future developments. Although forecasting is meant to limit the risks of decisionmaking, the fact that is based on prior estimates implies error, regardless of the approach. As a rule, the more invested in forecasting, the more accurate the estimates, and consequently the fewer the oversights and the less risky the decisionmaking process.

It was with this in mind that IPEA undertook an extensive study [1] to select better criteria for the construction of economic leading indicators capable of providing reasonably accurate estimates at relatively low costs in terms of time and money. While more precise statistics — generally *ex post* — are prepared by specialized institutes such as the Fundação Instituto Brasileiro de Geografia e Estatística (FIBGE) and the Fundação Getulio Vargas (FGV), their costs are substantially higher and the data become available only three months later or more.

Most economic variables — real product, employment, price levels, rate of capital accumulation, etc. — are characterized by fluctuations. Not only Brazilian economic history is marked by periods of intense expansion, others of modest growth, and a few of stagnation or negative growth. Other countries, be their economies

* This article is a summary of *Ciclos Econômicos e Indicadores de Atividade no Brasil*, Coleção Relatórios de Pesquisa n.º 35 (Rio de Janeiro: IPEA/INPES, 1977). I am grateful to Herval Aluísio Mota Cardoso for his assistance in updating the original and reviewing the data.

** Of the research institute of IPEA.

[1] Cited in the preceding note.

centralized or market-oriented, developed or underdeveloped, have experienced even stronger fluctuations. The fact that changes in levels of real output, employment and prices are recurrent and widespread phenomena have led to the concept of the "cycle" becoming one of the focal points of modern macroeconomic theory.

Business-cycle discussion lies between two poles. At one pole are those who argue that cycle phenomena, especially the longer-lasting, have no relationship to the structure of the economy; instead, recessions and revivals are due to purely random disturbances. From this standpoint, analysis of the known cycles contributes little to the prediction of subsequent cycles. At the other pole are those who contend that the occurrence of cycles is solely attributable to economic forces in interaction with sociocultural and natural-resource factors. Hence, for this group, understanding past cycles is of major importance to economic policy and economic history. The majority of writers stand between these extremes. Abramovitz, for example, agrees that the longer cycles have been correctly described and statistically documented, but adds that their causes remain mostly controversial. For Abramovitz and other economists, what ultimately matters is that the similarity of business cycles renders their analysis important to economic theory and policy, and hence justifies the development of statistical tools.[2] In our study we adopted a similar middle approach, not seeking the causes of Brazilian business cycles, but rather a set of relatively stable statistical "laws" that might be used to forecast future cycles.

The cycles of each economy present peculiarities that differentiate them from those of other economies. With reference to Brazil, two facts stand forth. In the first place, inflation has reigned even during recessions, when deflation is generally the rule. Rarely have prices remained stable or fallen, and even in these instances deflation has been slight and short-lived.[3] Thus, the outstanding characteristic of Brazilian economic history and price cycles has been the stability,

2 U. S., Congress, *Hearings before the Joint Economic Committee,* 86th Cong., 1st sess., part 2, pp. 411-66.

3 See M. H. Simonsen, "Inflation and the Money and Capital Markets of Brazil," in *The Economy of Brazil,* ed. H. Ellis (Berkeley: University of California Press, 1969), pp. 134-35; A. V. Villela and W. Suzigan, *Política do Governo e Crescimento da Economia Brasileira, 1889-1945,* Série Monográfica, n.º 10 (Rio de Janeiro: IPEA/INPES, 1975), table VII, pp. 424-25; O. Onody, *A Inflação Brasileira, 1820-1958* (Rio de Janeiro, 1960), pp. 25 and 117-19; M. Bucscu, *300 Anos de Inflação* (Rio de Janeiro: APEC, 1973), table LX, p. 223. For an econometric analysis of the determinants of the rate of inflation since 1861, consult C. R. Contador and C. L. Haddad, "Produto Real, Moeda e Preços: A Experiência Brasileira no Período 1861-1970," *Revista Brasileira de Estatística* 6 (July-September 1975): 407-440. The text contains some printing errors in the time series on real income, prices and inflation listed in the appendix.

rises and falls not in price levels but in inflation rates. In the second place, Brazilian economic growth has been practically uninterrupted in the twentieth century, with the possible exception of two short recessions during World War I and in 1929.

Besides being nearly constant, the economic expansion of Brazil in the last two decades achieved notable average rates previously unobserved in the country's history. This tendency was especially marked in the late sixties and early seventies, as the average annual rate of growth of GDP moved from just under 6% in 1947-1967 to over 9% in 1968-1970 and almost 11% in the next three years, at a pace comparable only to West Germany and Japan. However, the optimism generated by the so-called "Brazilian miracle" was countered by a drop in the rise in income from 9.8% in 1974 to a modest 5.6% in 1975, renewed by an increase to 9.2% in 1976, and again offset by a fall to 4.7% in 1977. The most pessimistic foresee a continuance of the modest growth of today, maintaining that the expansion of 1968-1974 was atypical and irrecoverable, and contending that Brazil is destined to a slow, hard-won growth that may, or may not, lead to developed-country status.

Fortunately, the modest or repressed growth attested in 1975, 1977, and 1978 is not permanent. What is likely to happen is that the current phase will be followed by another of high growth rates which, unhappily, will not last forever either.

Once one accepts that economic growth is in itself an unstable process, interest turns to three problems: (a) the causes and mechanisms of business cycles; (b) their signals; (c) the means of controlling them. Since there exists an abundant literature on the causes and ways that business cycles are extended, as well as on anticyclical measures, we may avoid discussing these themes. We shall embark, instead, on as critical and objective a discussion as possible of our recent experience with business-cycle forecasting indicators in Brazil.

Although the reasons behind business cycles are often unknown, especially in Brazil, economists from different schools agree that *ex ante* prediction of the future behavior of the economy allows that appropriate decisionmaking offset undesirable cyclical reversals within time and at minimum losses. Clearly, the more reliable the indicators and the earlier the diffusion of the forecasts, the greater the utility of the predictions to economic policy. Since our theme is still in the exploratory stage in Brazil and consequently subject to controversy, we will not follow any given theoretical approach. However, the existing theories will be used to the extent that they simplify the methodology without damaging the quality of the indicators, and will be judged on the accuracy of their *ex ante* performance.

2 — Defining business cycles

A business cycle is understood to be a regularly recurring sequence of expansions and contractions that affects an economy in many aspects, especially national product and employment levels.[4] Strictly, this concept is a pure abstraction. Recurrence implies that economic activity cannot follow the same path indefinitely. Moreover, any deviation tends to reinforce movements in the same direction until opposing forces push the economy the other way. This is a mechanical view of what is meant by a cycle.

However, even if cycles are accepted as recurrent fluctuations, there is neither theoretical nor empirical evidence that they follow the same pattern with respect to duration and amplitude. Some are severe, others hardly noticeable. In some cases the contraction lasts longer than the expansion; in others, the contraction is mild while the expansion is intense and short. Nonetheless, despite the specific nature of each, and the multitude of causes and characteristics, all business cycles are governed by certain processes or laws of formation which, once identified, allow for the construction of forecasting devices. This fact will be taken into account in making the necessary statistical assumptions. After all, if each cycle were considered different from all others in the past or future, the very notion of prediction would lose practical meaning.

The concept of the cycle has undergone modifications to the point that there are now three sets of criteria for identifying and dating the phases: (1) the "classical" definition, (2) the "revised" definition, and (3) the "growth-cycle" definition. Generally speaking, the timing, amplitude, lags, leads and so forth vary according to the definition used.

Under the classical concept, the cyclical phases and their chronology refer to peaks and troughs in the levels of aggregate series, the most common being National Product. According to the revised definition, the phases and their chronology relate to deviations from the trend of the series.[5] Lastly, the growth-cycle concept uses the minimum and maximum growth rates in the series as the criterion for identifying phases and their timing.[6] The implications of each concept are not purely academic, as we will see below.

[4] According to A. F. Burns and W. C. Mitchell, *Measuring Business Cycles* (New York: NBER, 1946). Also see V. Zarnowitz, *The Business Cycle Today* (New York: Columbia University Press, 1972).

[5] I. Mintz, *Dating Postwar Business Cycles*, NBER Occasional Paper, n.º 107 (New York: Columbia University Press, 1969).

[6] This is the concept suggested by M. Friedman and A. Schwartz, "Money and Business Cycles," in *The Optimum Quantity of Money and Other Essays* (Chicago: Aldine Publishing Co., 1969), pp. 189-235.

The classical definition identifies beginnings of depressions only if there are decreases in the absolute levels of product or employment. Clearly, it has lost ground since 1940, as such decreases have become rare in the majority of economies. Its utility is now limited to individual sectors and activities, for which it may hold even when not applicable to aggregate product.

The revised definition calls for prior recognition of the trend of each economic series. The fluctuations then correspond to the deviations relative to the trends. If the trend is horizontal, the cycles and their timing will obviously be the same according to the classical and revised concepts. However, if it is not (and in general the economic time series present positive trends), the trend may be influenced by certain factors that generate the cycle. For this reason, the method chosen for recognizing trends may affect the identification of business fluctuations. Although this choice is arbitrary, it is usually possible to graphically depict a trend and the deviations therein. But the existence of different tendencies in given subperiods requires special attention, for such tendencies may actually correspond to phases in a prolonged cycle.[7]

The revised concept, with some modifications, has come to be used for measuring the so-called "GNP gap". Since it is important to know how far the economy is from full employment before taking steps, and since there is no direct means of gauging full employment, simplified methods have been developed. The most widespread is that adopted by the U. S. Council of Economic Advisors.[8] This method determines potential product by interlinking peaks with the trend line. The tangents hypothetically represent full (or near full) employment, and the difference between potential and actual product stands for the idle capacity of the economy. As a rule, cyclical phases are inversely associated with idle capacity and unemployment; during a contraction idle capacity increases, and during an expansion it decreases. If the trend used to measure the gap is the same as that utilized to render the series stationary,

[7] For a detailed discussion, see R. A. Gordon, *Business Fluctuations* (New York: Harper & Row, 1961), pp. 253-57.

[8] "The Gap between Actual and Potential GNP," *1965 Annual Report of the Council of Economic Advisors*, pp. 81-84, reprinted in *Macroeconomic Readings*, ed. J. Lindauer (New York: The Free Press, 1968), pp. 244-47; "Realizing the Economy's Potential," *1969 Report of the Council of Economic Advisors*. For a complete description of the method, see L. R. Klein and R. S. Preston, "The Measurement of Capacity Utilization," *American Economic Review* 57 (March 1967): 34-58.

the shape of the gap will be precisely opposite that of the revised cycle, despite the fact that the average values may differ.[9]

Lastly, the growth-cycle definition identifies cycles in terms of rates of change. This concept is highly applicable to the Brazilian case, for the rapid expansion of the economy has eliminated classical cycles, and frequent changes in the trend have made the revised concept unsuitable.

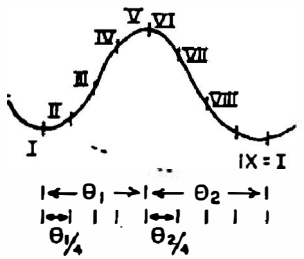
Regardless of the criteria used, a complete cycle has two phases: an expansion and a contraction. Pre-Keynesian theory further divided the expansion into periods of recovery and prosperity, and the contraction into periods of recession and depression. However, these subdivisions proved confusing, and the distinctions purely semantic and subjective.[10] Retaining the expansion-contraction cleavage, the National Bureau of Economic Research (NBER) has designed a reference cycle in which the interval between peaks and troughs is arbitrarily divided into four subphases.[11] The expansion covers subphases I to V, all of equal duration, and the contraction subphases V to IX. Subphase IX corresponds to subphase I of the following expansion. The duration of each subphase depends on the duration of each phase. Thus, in a given cycle, the four expansion intervals are of equal length, as are the four contraction intervals. The average duration of an expansion subphase may differ, however, from that of a contraction subphase.[12] Likewise, the duration of expansions and contractions may vary from cycle to cycle. The NBER reference cycle will be useful for understanding

[9] This method is applied to the Brazilian economy by C. R. Contador in "Crescimento Econômico e o Combate à Inflação," *Revista Brasileira de Economia* 31 (January-March 1977): 131-68.

[10] Or political. When the economy suffers a setback, for example, the party in power terms it a "slight recession" while the opposition calls it a "depression".

[11] For a good summary, see W. I. Greenwald, *Statistics for Economics* (Columbus, Ohio: Charles E. Merrill Books, Inc., 1963), chap. 2.

[12] In pre-Keynesian literature, recovery included subdivisions I to III; prosperity, III to V; recession, V to VII; and depression, VII to IX (= I).



the differences between the three concepts discussed above, as well as their applicability in dating Brazilian cycles.

National-accounts data are available as of 1947, when FGV began to gather annual statistics. Unfortunately, three decades is too short a span for detecting well-defined cycles, especially those of longer duration.

However, the official national-income series may be complemented by indirect estimates for the years prior to 1947.[13] As a guideline, we will use a domestic-product series that goes back to the mid-nineteenth century. Although the estimates are preliminary, the income fluctuations are in basic agreement with known historical facts.[14] Certainly the real-product estimates may contain errors, but for our purposes the trend and absolute level of the series are less important than the fluctuations and their chronology.

The GDP estimates are shown in the upper part of figure 1. The negative signs (—) indicate contractions and the positive signs (+) expansions, according to the classical definiton of the business cycle. Some cycles are clearly related to historical events. For example, the contraction in 1865-1868 was due to the Paraguayan War (1865-1869) and the economic crisis which originated in France and England in 1866. The contraction that began in 1874 and lasted for seven to eight years was initiated by the world recession of 1872 and worsened by the drought of 1877. The expansion of the next five years was followed by another contraction in 1889-1898, this time brought on by various political and economic changes and disturbances, including the abolition of slavery in 1888, the proclamation of the Republic in 1889, numerous internal conflicts, the "*Encilhamento*" [15] in 1892, the drop in international coffee prices in 1896, and, lastly, the austerity program adopted by the Campos Sales government in 1898.

13 See, for instance, Contador and Haddad. Another series, more complex but shorter, is that presented by C. L. Haddad in "Crescimento do Produto Real Brasileiro: 1900-1947," *Revista Brasileira de Economia* 29 (January-March 1975): 5-26.

14 The behavior of the domestic-product series estimated by Contador and Haddad is in accord with the historical descriptions of Brazilian economic crises presented by various authors, among whom the following might be cited: R. C. Simonsen, *História Econômica do Brasil, 1500-1820* (São Paulo, 1937); *idem, A Evolução Industrial do Brasil* (São Paulo, 1939); C. Prado Junior, *História Econômica do Brasil* (São Paulo: Editora Brasiliense, 1956); C. Furtado, *Formação Econômica do Brasil* (Rio de Janeiro: Fundo de Cultura, 1964); Villela and Suzigan.

15 The crisis of the *Encilhamento* — the outcome of the economic euphoria and financial speculation of the first years of the Republic — culminated in late 1891 with the bankruptcy of numerous firms that had just been created, mainly in Rio de Janeiro.

The international crises that preceded and marked the beginning of the First World War led to the contraction of 1912-1915, which was followed by an expansion that lasted up to the eve of the 1929 crisis.[16] The end of the economic impact of the Great Depression (in 1932-1933 in Brazil) signaled another expansionary phase, which was to continue until the Second World War, when the export controls imposed by the United States and European nations restricted the Brazilian imports of capital goods and basic materials. Since 1942, the Brazilian GDP has registered uninterrupted growth, though at varying rates. The three decades would therefore belong to a prolonged period of cyclical expansion in terms of the classical concept.

As the chance of absolute decreases in real product has become more and more remote, the classical definition of the business cycle has stopped arousing academic and political concern. Nonetheless, despite the fact that disturbances due to cyclical fluctuations have been increasingly offset by improved economic policy and government intervention, cycles are still observable in other forms. Thus, as the middle part of figure 1 shows, the phases of the revised cycle usually precede those of the classical cycle. Moreover, use of the revised definition reveals additional cycles. Whereas the classical concept indicates constant growth since the Second World War, the revised concept points to one expansionary phase having ended in 1961, and another having extended from 1967 to 1974. Finally, the growth cycles traced in the lower part of figure 1 predate not only the classical cycles, but also the revised cycles. To cite the most recent examples according to the growth-cycle definition, the contraction that began in 1961 ended in 1963 rather than 1967, and the last expansion in 1973 rather than 1974.

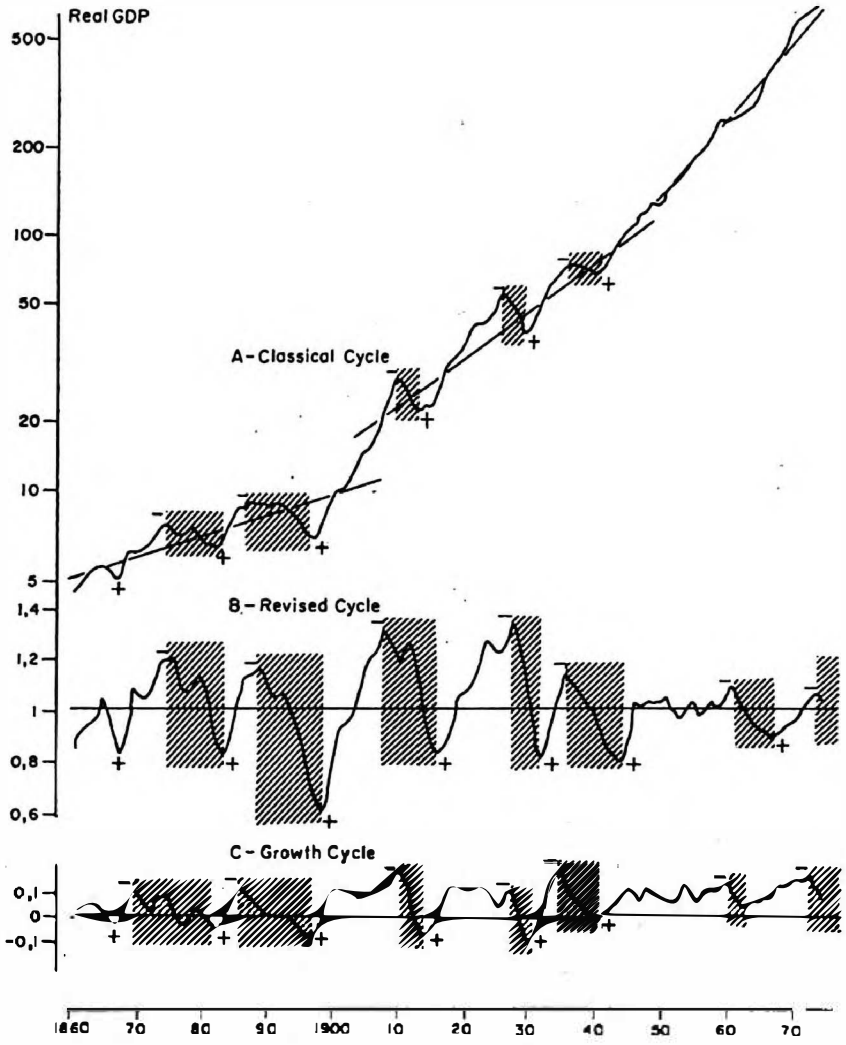
The dating of the cycles according to the three definitions is summarized in table 1. Each line represents one complete cycle, i.e. a contraction and an expansion. As in figure 1, the contraction spans under the revised definition are equal to or longer than those under the classical definition, and the expansions shorter under the former than under the latter. Likewise, the cycles begin earlier according to the growth-cycle concept, except for the contraction which the growth concept dates as having started in 1910, the revised definition in 1908, and the classical definition in 1912.

Rapid economic growth has been the long-standing government aim in Brazil, and the current emphases on fighting inflation and reducing the balance-of-payments deficit are certainly exceptions to

[16] Furtado, pp. 222-23, argues that the Brazilian economy was affected by the crisis as early as 1928, and that it had already begun to recover in 1932, when the U. S. still faced another full year of depression.

Figure I

BRAZILIAN BUSINESS CYCLES, 1860–1970



The dates marked signal the beginning of contractions (–) and expansions (+) according to each definition. The darkened areas indicate periods of contraction.

Table 1

Brazilian Business Cycles According to Three Definitions,[a] 1865-1977

Classical Cycle				Revised Cycle				Growth Cycle			
Contraction		Expansion		Contraction		Expansion		Contraction		Expansion	
Start	Duration (Years)	Start	Duration (Years)	Start	Duration (Years)	Start	Duration (Years)	Start	Duration (Years)	Start	Duration (Years)
1865	3	1868	8	1865	3	1868	7	1863	4	1867	3
1876	8	1884	6	1875	9	1884	5	1874	7	1882	8
1890	8	1898	14	1889	9	1898	10	1886	4	1897	13
1912	3	1915	13	1908	8	1916	12	1910	4	1914	13
1928	4	1932	6	1928	4	1932	4	1927	3	1930	5
1938	4	1942	...	1936	8	1944	17	1935	5	1949	21
—	—	—	—	1961	6	1967	7	1961	2	1963	10
—	—	...	...	1974	...	—	—	1973	2	1975	1
								1976	1	1977	...

a Only the major cycles are identified.

the rule. For this reason, in Brazil the most relevant concept is the growth-cycle definition. The fact that this definition dates cycles relatively early is an advantage that deserves investigation, but a factor that creates certain practical difficulties when constructing leading indicators.

3 — Leading indicators

Economic planning — whether on the part of individuals, firms or the government — is conducted based on forecasts for the future. And since anticyclical measures do not have an immediate impact, their effectiveness depends not only on the follow-up, but also on the prediction, of macroeconomic variables such as growth of real product, employment levels, and inflation rates.

Unfortunately, the necessary data are often made available with considerable delay, when not with serious statistical errors. To overcome these problems, national-accounts and other statistics would have to be gathered sooner, more accurately, and at biannual or quarterly rather than annual intervals. Despite the efforts made in this direction by FIBGE, the task is painstaking and not likely to yield results for some time. Meanwhile, the policymakers will have to continue to rely on certain variables that have come to be accepted as proxies for national income and industrial product, such as industrial electricity consumption, sales of household appliances, sales-tax collections and so on.

Despite its apparent simplicity, the utilization of selected variables is subject to four types of interpretation problems. Firstly, the variables chosen as "economic thermometers" rarely point in the same direction. In any given period, some may be rising, others constant, and others falling. Even during the most severe contractions, some variables will register growth. Under these conditions, it is difficult to identify a contraction or an expansion.

A second problem derives from the existence of leads and lags between variables. If the leads and lags are not known, the growth rates of the variables may be wrongly interpreted. For example, should the economy be in depression, a rising variable that leads real product for several months may be overlooked as an indicator of expansion. Similarly, when the economy is recovering, a variable that lags product may still show a recession. For this reason, it is essential to estimate the leads and lags among variables when constructing leading indicators.

Thirdly, there is a difference between real and nominal changes. Although it is customary to restrict analysis or give more emphasis to the real changes in the variables, the question is controversial, especially in the presence of unstable inflation.

Even if several variables are found to be closely correlated and timed with real income and employment, they would not be useful leading indicators. Such variables may serve as current or coincident indicators useful for accompanying the movement of the group, but unsuited to predicting fluctuations in the coming months.

The last problem concerns the ability of the indicators to forecast changes in the reference variable (in most cases, the rate of growth of real product) . Here the question involves not only the number of variables — experience reveals that inclusion of too many details complicates the method and that exclusion of some has little effect on *ex ante* performance — but also their quality.

Although the quality of the data used in this analysis was often unsatisfactory, the approach chosen tends to dampen the effects of errors in isolated variables. Thus, the main difficulty encountered in constructing the indicators was lack of statistics and the length of the time series available.

Any forecasting system has three time dimensions: (1) the period or interval covered, (2) the space between this period and the time the forecast is made, and (3) the frequency of forecasts and revisions. The period is the unit of time (month, quarter, etc.) to which the prediction refers. Generally speaking, the shorter the period, the more difficult the gathering of data and the riskier the predictions due to erratic movements. Forecasts for monthly intervals are thus more subject to error than those for longer, e.g. annual or semester, intervals.

The element of error also enters in considering the second dimension. If anticyclical policies are to be effective, measures need to be taken well ahead of time. However, the longer beforehand the predictions are made, the greater the chance of mistakes. It is harder to foretell the fourth-quarter rate of growth of industrial product, for instance, in the first quarter than in the third quarter, when more information is available.

Finally, even though running forecasts should, theoretically, reduce the margin of error, in fact, first estimates are sometimes more accurate than the revised ones, as we shall see later.

A phenomenon initially overlooked in the NBER method is inflation. This is excusable for the U.S. economy, which enjoys relatively stable prices, but would hardly be so in the Brazilian case. We will first discuss forecasting under stable inflation, and then turn to the difficulties involved in constructing forecasting systems for countries such as Brazil whose economies are characterized by unstable inflation.

Stable inflation influences the dating of classical cycles observed in deflated variables. A continuous and constant rise in prices pushes the trend line of a nominal variable such as current GDP upwards.

Expansions therefore tend to be longer and contractions shorter for nominal domestic product than for the deflated variable. Similar differences are not observed in the timing of revised and growth cycles. Since price increases are incorporated in the trend, the chronology of revised cycles remains untouched, and real and nominal fluctuations are superimposed. In turn, because the difference between the real and nominal growth of product corresponds to a constant (the rate of inflation itself), the timing and dating of growth cycles also remain unchanged. Nonetheless, it would be precipitate to conclude that stable inflation can be disregarded in business-cycle analysis.

Nominal changes acquire economic significance only insofar as they affect real variables. However, economic policies aimed at fostering economic activity seek to shift nominal aggregate demand. The extent to which nominal income raises prices and physical product depends on the amount of idle capacity and the degree of factor mobility in the economy. If there is no idle capacity, shifts in aggregate demand can be expected to raise prices only. However, should there be significant idle capacity or a possibility of improving factor allocation, the economy can expand without exerting further pressure on prices.

It is not always easy to differentiate real and nominal changes.[17] Theoretically, the distinction is simple and almost tautological, with real changes roughly corresponding to the difference between the nominal variations and the inflation in the period. But what price index should be used? Should it be seasonally adjusted or not? These and other doubts arise, and the use of different deflators can lead to different answers. Moreover, in the short run, not all nominal variables have the same relationship to the price index. Due to seasonal and other cyclical factors that affect specific variables, deflating may conceal important movements or forge others that are, in fact, nonexistent.[18] Nominal series must therefore be deflated with caution, especially in the construction of leading indicators.[19]

If all the preceding holds for stable inflation, what must be said for unstable inflation? Unfortunately, inflationary processes are rarely stable in real life. There is even evidence that prices

[17] The fact that this difference is not accounted for in the NBER indicators has led to serious problems in understanding cyclical movements in the U.S. economy now that it is subject to inflation.

[18] The stock market provides a good example. The nominal stock-market index does not contain seasonal fluctuations, but if the real stock-price index is obtained through deflating by a general price index, inverted seasonal fluctuations similar to those of the deflator will appear. This does not make sense, for it clearly goes against the hypothesis of profit-maximizing investors.

[19] On this point, see C. S. Greenwald, "A New Deflated Composite Index of Leading Indicators," *New England Economic Review*, July-August 1975, pp. 5-17.

rise faster in the final phases of expansions and increase slower (or even decrease) during contractions. When price accelerations or decelerations are not anticipated, the explanation may lie in short-run trade-offs between inflation and employment or idle capacity, in the pattern of the Phillips curve. Our experience supports this hypothesis,[20] and it would be pointless to negate the instability of Brazilian inflation and its effects on real variables.

Instable inflation affects the dating not only of classical cycles, but also of revised and growth cycles. For example, according to the revised definition, if it is admitted that prices accelerate during expansions and decelerate during contractions, the former will tend to be longer and the latter shorter with nominal series than with deflated series. In the case of growth cycles, the timing is also changed, but in what direction is impossible to predict a priori. All depends on the behavior of the actual rate of inflation in each phase of the cycle.

In sum, in order to analyze and forecast real fluctuations in economies with unstable rates of inflation, it is necessary to deflate the nominal variables despite the problems that this procedure may create. Further on, we will show how to keep deflation from eliminating or generating cyclical movements.

The performance of the forecasting system is based on the relative frequency of mistakes pointed out by *ex post* experience. On the one hand, indicators rank high if they generally score well in the *ex ante* prediction of most cyclical reversals. On the other hand, they lose validity if they systematically lead to "errors of omission" or "false signals".[21] In the first case, they fail to predict cyclical turns; in the second, they foretell turns that do not occur.

It is difficult to avoid errors of omission because the behavior of leading indicators is determined by that of the variables of which they are comprised. Since leads and lags vary depending on the duration of the cycle, some unexpected fluctuations are inevitable. Thus, in order to minimize errors of omission, the indicators should be composed of variables that are evenly distributed among cycles of different lengths. Some should lead the reference variable in the shorter cycles of three to five months, others in the cycles of six to 12 months, and so forth.

False signals are both harder to overcome and more serious due to the false optimism or pessimism they may generate. Owing to their erroneous warnings, the government may relax measures designed to raise aggregate demand before a turn actually occurs,

20 Contador, "Crescimento Econômico e Combate à Inflação."

21 For a short discussion, see M. C. Lovell, *Macroeconomics: Measurement, Theory, and Policy* (New York: John Wiley and Sons, Inc., 1975), pp. 390-92.

or, should the false warning be given during an expansion, create conditions capable of accelerating prices. Since false signals are often the result of misinterpretation of strong random movements, the more stable the behavior of the indicator, the lower the probability of mistakes.

The construction of leading indicators begins with prior analysis of the leads and lags between the variable to be forecast — that is, the reference variable — and the others to be utilized. This assumes that the patterns observed in the past will hold true in the near future. Once the leads and lags are identified, the variables can be classified as leading, coincident or lagged.

The method by which leads and lags are identified depends on which definition of the cycle is used. The simplest method cross-correlates past and future changes in two economic series. Variable X leads variable Y (industrial product, for example) if the highest correlations are between the current value of Y and the past values of X. The two variables are coincident if the highest correlation is for the nonlagged values. Finally, X lags Y if the correlations are highest between the past values of Y and the current values of X. Under the revised definition of the business cycle, the correlations refer not to levels, but to deviations relative to the trend; analogously, according to the growth-cycle concept, the correlations are between the rates of change of X and Y.

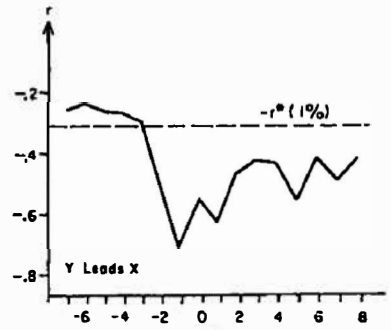
Figure 2 illustrates how this criterion works for the growth-cycle definition. The examples refer to the Brazilian economy in the period January 1960 — December 1975. On the vertical axes are shown the coefficients of correlation between the 12-month rates of change of the reference variable (i.e. industrial product) and the leads and lags of variable X over the same period. The leads (+) and lags (—) are expressed in months. The values at which the correlations are significantly different from zero at the 1% level are indicated by the horizontal lines.[22]

Theoretically, for leads to have economic significance, the correlation between the behavior of changes in industrial product and those in the variable under analysis must be recognized beforehand. Receiverships, for instance, are expected to be inversely associated with real product; in other words, the two variables should be negatively related or opposite in phase, since industrial slowdowns increase the number of receiverships and vice versa. Hence, the variable "receiverships" will be identified as lagged, coincident or

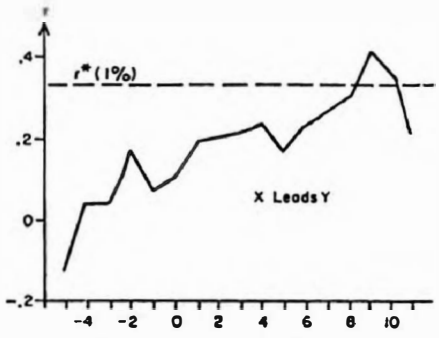
22 Throughout this article, a coefficient of correlation is considered significantly different from zero if its absolute value is higher than .32 at the 1% level (the value used in the figures) or higher than .25 at the 5% level.

Figure 2

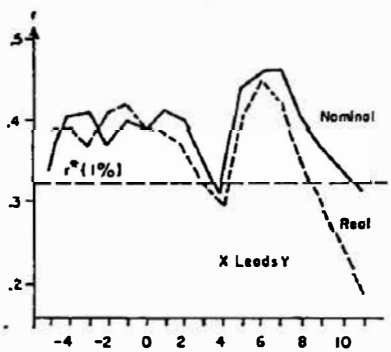
CROSS-CORRELATION BETWEEN INDUSTRIAL PRODUCT (Y) AND SELECTED VARIABLES(X), JANUARY 1960 – DECEMBER 1975 (LAGS [–] AND LEADS [+] IN MONTHS)



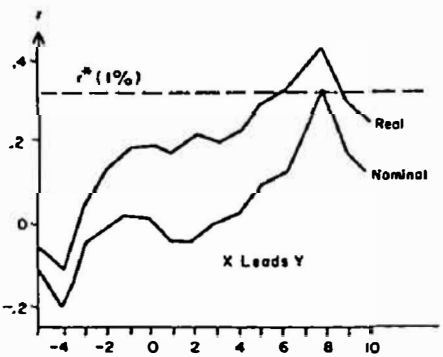
A – Receiverships: São Paulo



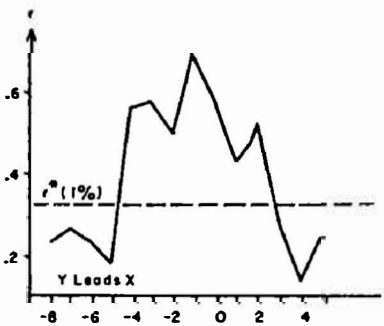
B – Nonresidential Building Permits: Square Units



C – Money Supply: Definition M_2



D – Ex Post Federal Expenditures



E – Industrial Consumption of Electricity: Rio de Janeiro – São Paulo

leading depending on whether the highest negative correlation is during lags (months with a negative sign), in the current month (month zero), or during leads (months with a positive sign), respectively. As the figure shows, receiverships tend to lag industrial product by one month. As our interest lies in forecasting future cyclical reversals, this and similar variables may be disregarded in the construction of leading indicators.

Before turning away from lagged variables, a glance should be taken at the diagram on industrial electricity consumption (figure 2). The one-month lag confirms the findings of an earlier study.[23] Since electricity is a basic industrial input, this variable should lead or coincide with industrial product. The fact that it does not may be owing to statistical shortcomings or to delays in the publication of data.[24]

The other diagrams in figure 2 trace the behavior of leading variables. "Nonresidential building permits" leads industrial activity by approximately nine months (in comparison to three months in the study cited in fn. 23), "money supply M_2"[25] by six months, and "*ex post* federal expenditures" by eight months.[26] The latter two variables are presented in both nominal and real terms (deflated by the general price index). The correlations are positive and significant, with money supply registering some advantage.

Taken together, the diagrams in figure 2 sufficiently illustrate the empirical method most commonly used to classify variables as leading, coincident or lagged, and to quantify the respective leads and lags in relation to the reference variable. Multiple-regression models, in which the reference variable is explained by distributed lagged variables or orthogonal polynomials of the Almon type,[27] are variations on the method described.

That each variable goes through numerous cycles underlines the fact that the hypothesis of a single lead is a gross oversimplification, despite its widespread and unqualified acceptance. Nonetheless,

23 C. R. Contador, "Indicadores de Atividade Econômica no Brasil," *Pesquisa e Planejamento Econômico* 6 (April 1976): 1-60.

24 This case is explored by J. A. P. Peixoto, "Avaliação do Índice de Crescimento de Energia Elétrica como Indicador de Crescimento Industrial," *Revista Brasileira de Estatística* 36 (July-September 1975): 531-40.

25 Currency plus demand, time and savings deposits, and certificates of deposit.

26 The above results are not meant to compare the lags between monetary and fiscal policy. For this comparison to be made, other government expenditures such as those of autarkies and government companies would have to be included.

27 S. Almon, "The Distributed Lag between Capital Appropriations and Expenditures," *Econometrica* 33 (January 1965): 178-96.

acceptance of the notion that economic series are comprised of cyclical components of distinct durations does not imply adherence to any theory that postulates endogenous self-sustaining cycles such as the investment cycle of Hicks and Samuelson or the stock-cycle proposed by Metzler.

There is convincing evidence that economies respond to disturbances and stimuli in the form of oscillations that generally fade over time. Depending on the type of disturbance, the response may resemble a cycle, with the economy being simultaneously affected by numerous variables, each with its specific characteristics, range and duration of impact.

In constructing indicators, another factor that must be taken into account is the recurrence of the fluctuations to be forecast. If the government is concerned with long-range prospects only, there is no need to cover short-range fluctuations. But if the policy aim is to stabilize short-term growth of product, the indicators must be duly adjusted.

From what has been said about lags and cyclical movements, it is possible to establish guidelines for choosing variables that can be used as indicators or included in composite indicators. Such variables should (1) consistently conform to known fluctuations in economic activity, (2) present stable leads relative to the overall economy, (3) be dominated by nonseasonal movements rather than erratic oscillations, and (4) have importance to the reference variable or to the economy as a whole.[28] Although these criteria place leading, coincident and lagged indicators on the same plane, the first are unquestionably the most important to economic policy, since they are the only ones capable of predicting fluctuations in income and product. Coincident and lagged indicators can do no more than serve as a check on the leading indicators.

Mitchell and Burns [29] list additional criteria that help forestall errors of omission and false warnings, but due to problems such as difficulty in obtaining longer time series, we were not able to follow their suggestions. Thus, in the Brazilian case, an indicator will receive a higher ranking:

(a) the longer and more stable its lead relative to the economic activity or employment in each cycle. In other words, holding the

[28] V. Zarnowitz and C. Boschan, "Cyclical Indicators: An Evaluation and New Leading Indexes," *Business Conditions Digest*, May 1975.

[29] W. C. Mitchell and A. F. Burns, *Statistical Indicators of Cyclical Revivals*, Occasional Papers, n.º 69 (New York: NBER, 1938), p. 173.

next three conditions constant, an indicator that leads economic activity by four months is preferable to one that leads it by three months, or less.

(b) the more stable and uniform its lead for cycles with different frequencies. The ideal indicator would lead activity by a constant period, say four months, in any cycle. This condition would be satisfied by a "pure lag", but the possibility of such a lag is remote.

(c) the more stable and uniform the association between the indicator and economic activity. For example, if regressions were performed for the fluctuations in the indicator and those in the level of economic activity, disaggregated into cycles with different intervals, the coefficients of regression should be identical among the several cycles.

(d) the wider the range of activities covered by the indicator. What matters is that the variable chosen in constructing the indicator be strongly associated with other important aggregate variables in order that it satisfactorily approximate a broader process or other economic movements.

When the variables are affected by strong seasonal elements, i.e. by 12-month cyclical movements, it may be harder to recognize leads and lags. Even if variables X and Y are influenced by completely different seasonal factors, cross-correlation will produce a lag whose distortion increases with the importance of the seasonal cycles in the formation of the series. On the one hand, if the fluctuations in X and Y are dominated by seasonal changes, the average lag will be basically determined by the lag between the seasonal movements, which is often generated artificially and is irrelevant from the economic standpoint. On the other hand, if neither of the variables (or only one) is subject to seasonal changes, the average lead will not be affected. Thus, special care must be taken to identify seasonal elements, for cross-correlations habitually marked by such elements tend to produce indicators that give false warnings.

A statistical analysis (not reproduced in this article) [30] showed that industrial product and employment, the two reference variables most important for our purposes, as well as the majority of series available for the construction of indicators, are characterized by strong seasonal factors. This emphasizes the need to remove these factors before aggregating the variables into composite indicators.

30 Contador, *Ciclos Económicos*.

4 — Construction of the indicators

4.1 — Methodology

The technique used for constructing leading indicators for Brazil was that developed by the National Bureau of Economic Research.[31] This method involves analysis of the behavior of a large number of variables which are observed at various phases of the business cycle and aggregated on the basis of homogeneity of timing relative to the cycle. The variables are then classified as leading, coincident or lagged depending on whether their turning points precede, coincide with, or follow those of the reference variable. It is further assumed that the variables that antedated the majority of reversals in the past will continue to do so in the future, and can consequently be used to form composite leading indicators.

A certain degree of selectivity is required, however. Since seasonal variations are regularly recurrent and may be generated or extended by the process of deflation, for instance, such variations are of little economic concern and should therefore be removed from the time series that comprise the indicators.

Let us assume that time series $X(t)$ includes four independent logarithmic additive components: a trend (X^T), a seasonal component (X^S), other cyclical components (X^C), and a random component $(e[t])$. That is,

$$Log\ X(t) = Log\ X^T(t) + Log\ X^S(t) + Log\ X^C(t) + e(t) \qquad (1)$$

Our main interest is in X^C, which we know to be formed of a large number of cycles. Through statistical techniques we will try to eliminate the effects of the trend and the seasonal variations, and, if possible, to dampen the impact of random fluctuations. If the components of the time series are expressed in logarithmic terms, as in equation (1), simple differentiation will eliminate the trend.

In turn, the series may be affected by seasonal components formed through a moving-average and/or auto-regressive process

31 This method dates from W. C. Mitchell's pioneer work *Business Cycles* (Berkeley: University of California Press, 1913), followed by his previously cited studies together with A. F. Burns, *Statistical Indicators of Cyclical Revivals* and *Measuring Business Cycles*. The technique and its forecasting ability have been the subject of numerous discussions, some of which have become classic. See, for instance, T. C. Koopmans, "Measurement without Theory," *Review of Economics and Statistics* 29 (August 1947): 161-72; and R. Vining, "Koopmans on the Choice of Variables to Be Studied and of Methods of Measurement," *Review of Economics and Statistics* 31 (May 1949): 77-94.

and/or degree of seasonal differentiation. For convenience sake, we shall disregard the first two possibilities and assume that simple first-order seasonal differentiation is capable of eliminating the trend and the 12-month cycles. That is,

$$(1 - B^{12})\ Log\ X(t) = \mu + (1 - B^{12})\ Log\ X^{C}(t) + (1 - B^{12})\ e(t) \quad (2)$$

where $(1 - B^{12})$ is the filter formed by the seasonal differentiation, and B the lag operator

$$B^{j}\ X(t) = X(t - j) \quad (3)$$

The rate of growth corresponds to the logarithmic derivative of the trend and is equal to the constant μ. Manipulation of (2) transforms $X(t)$ into a stationary series free of seasonal elements. To facilitate notation, let x_i represent the stationary series [32] of variable X_i.

$$x(t)_i = (1 - B^{12})\ Log\ X(t)_i \quad (4)$$

Strictly speaking, x_i is a rough approximation of the growth rate of X_i over the last 12 months, when the variation is close to zero. But, given the variations estimated, it might be better to obtain x_i using the rate of growth over the past 12 months. However, for the sake of consistency, we will maintain notation (4).

In order to keep the more volatile variables from dominating the indicator, fluctuations x_i need to be standardized for a determined unit. Prior normalization of the variables with their means and standard deviations is the most common criterion used to compare the behavior of relative changes. In other words,

$$z(t)_i = \frac{x(t)_i - \overline{x}_i}{\sigma_{x_i}} \quad (5)$$

where z_i is the series standardized by the average $\overline{x}_i$ and the standard deviation σ_x. Another form of standardization is based on the moving average of the absolute value of x_i.

In fact, this method was used in the previous study, [33] but the continuous need to estimate new moving averages each time new

32 Another form of linear filter based on a moving average is that used in the Census X-11 Program. See W. P. Cleveland and G. C. Tiao, "Decomposition of Seasonal Time Series: A Model for the Census X-11 Program," *Journal of the American Statistical Association* 71 (September 1976): 581-87. A more specific filter could be developed using the general ARIMA model of Box and Jenkins.

33 Contador, "Indicadores de Atividade Econômica"; *idem*, "Queda e Recuperação do Ritmo de Crescimento Econômico," *Conjuntura Econômica* 30 (April 1976): 94-99.

information is obtained shows the shortcomings of the method. Moreover, if the distribution of the rates of variation is relatively stable and the number of observations is large, the normalized estimates obtained through the average and the standard deviation of the sample should also be quite stable and similar to the estimates for the universe. Under these conditions, the standardization through (5) can be grounded on the same average and standard deviations, and additional calculations made only occasionally.

The next step is to obtain the mean aggregate rate of variation for the last 12 months, in which the standardized rates are weighted as follows:

$$y(t)_j = \sum_{i=1}^{n} w_{ij}\ B_j^{l_{ij}}\ z(t)_i = \sum_{i=1}^{n} w_{ij}\ z(t - l_{ij}) \tag{6}$$

where $y(t)_j$ stands for the rate of change in the leading indicator for the reference variable y_j; w_{ij} for the weighting of variable X_i in the formation of y_j; and l_{ij} for the average lag (in months) between variables X_i and y_j. B represents the lag operator given in equation (3). By definition,

$$\sum_{i=1}^{n} w_{ij} = 1 \tag{7}$$

Series $y(t)_j$, which is comprised of standardized series, corresponds, in turn, to an "approximately" standardized variable. The "destandardization" is the result of multiplying $y(t)$ by the standard deviation of the reference variable to be forecast and adding its average to

$$y^*(t)_j = \bar{y}_j + y(t)_j\ \sigma_{x_j} \tag{8}$$

where y^* is the indicator expressed in growth rates and in dimension comparable to the growth of reference variable y_j.

The weights of the standardized variables are based on the correlation between the reference variable and each variable x_i,

$$w_{ij} = \frac{r_{ij}}{\Sigma |r_{ij}|} \tag{9}$$

Weighting variables in the above manner, i.e. in terms of their correlation with the reference variable, is a means of bypassing subjective criteria such as those employed first by Moore and Shiskin and subsequently by Zarnowitz and Boschan.[34] In both

[34] Zarnowitz and Boschan, "Cyclical Indicators."

cases, the variables were chosen and weighted according to a subjective consensus as to their economic significance, statistical quality, regularity of cyclical timing, conformity and absence of random components, and ready availability of data. Equation (9), like any weighting system, involves some arbitrary choices, but it has the advantage of placing associations between observed movements over subjective preferences.

Finally, the average lag $\bar{l}$ of indicator y^* is the average of the lags between variables X_i $(i = 1, 2, \ldots N)$ and the reference variable y_j, weighted by the coefficients of correlation r_{ij}:

$$\bar{l}_j = \frac{\sum_i l_{ij} |r_{ij}|}{\sum_i |r_{ij}|} \tag{10}$$

This is the approach used in constructing aggregate leading indicators for the Brazilian economy.

4.2 — Results of the leading indicators

For observing the economy and predicting future developments, the aggregate variables of most interest are real product, employment, and the rate of inflation. Of the three sectors that contribute to national product, industry, and secondly services, are best suited to the construction of leading indicators. Despite the importance of the agricultural sector, two types of difficulties limit forecasting of its short-term behavior. In the first place, production and factor-utilization decisions are subject to seasonal restrictions. Between planting and harvesting little or nothing can be done to alter the quality and volume of the product, and there is always the chance of crops being affected by pests or adverse climate. In the second place, since decisions are made at appreciably longer intervals than in industry, it is hard to obtain data on short-term fluctuations of the variables (which, in any case, are different from those that comprise industrial indicators). For these reasons, we chose to concentrate on the industrial sector.

Monthly indicators were constructed for variables that portray two interrelated phenomena: employment and product. For the level of industrial employment, the reference variable is the series mounted for Greater São Paulo, as of 1965, by the industrial federation of the state of São Paulo (Federação das Indústrias do Estado de São Paulo — FIESP). For industrial product, the data used are the national statistics gathered by FIBGE since 1969.

Analysis of the cross-correlations (not reproduced here)[35] for more than 80 variables revealed that 38 lead industrial employment and that 32 lead aggregate industrial product. On the basis of the value of the correlations and the number of months of the leads, it was possible to reduce to 11 and 14, respectively, the number of variables included in the leading indicators for employment and product.

The composition and other information on the leading indicator for industrial product in Brazil are given in table 2. The average lead, weighted as in equation (10), was estimated at five to six months, in comparison to the average lead of just over two months produced by a preliminary IPEA indicator.[36] The variables cor-

Table 2

Aggregate Industrial Product: Composition of Leading Indicator

Variable	Average Deviation[a]	Standard Deviation[a]	Correlation with Reference Variable[b]	Lead (Months)[c]
Nonresidential Building Permits (in square units)	.1415	.3717	.306	9—10
Residential Building Permits (in square units)	.0321	.2053	.421	9
Real Value of Textile Production	.0448	.0745	.700	9—10
Financing and Refinancing of Real Estate[e]	.3065	.1035	.478	8
Ex Post Federal Expenditures[e]	.1717	.2224	.431	8
Rediscounts to Commercial Banks: Manufactures[d]	.2080	.2326	.688	7
Exports: Other Products (in US$)	.3511	.3554	.545	7
Value of Checks Cleared[d]	.2879	.1658	.587	6
Credit Requests for Purchases of Consumer Goods: São Paulo (total)	.0303	.2295	.903	6
Loans through Bills of Exchange[d]	.2760	.1532	.561	5
Laminate Production (total)	.0744	.0815	.873	4
Value of Bills Protested: São Paulo[d]	.2662	.3507	—.547	4
Employment Demand for Technicians: Greater São Paulo	.2314	.2046	.715	2
Money Supply: Definition M_2[d,e]	.1924	.1075	.628	2

a Average and standard deviation in the 12-month rate of growth.

b The reference variable is the index of the real value of industrial product estimated by the Fundação Instituto Brasileiro de Geografia e Estatística (FIBGE). All the correlations are significantly different from zero at the 1% level or higher.

c The average weighted lead is 5.6 months.

d Deflated by the general price index — internal availability.

e Currency plus demand, time and savings deposits, plus certificates of deposit.

35 Contador, *Ciclos Econômicos*.

36 Contador, "Indicadores de Atividade Econômica." At first sight, there is no point in comparing the composition of the two leading indicators since the reference variables are different. However, for industrial product in the state of São Paulo (the reference variable in the first study), the conclusions are the same.

respond to investment decisions in construction; the behavior of the textile industry; manufactured exports; credit and solvency in the private sector; laminate production; employment demand for technicians; government expenditures; and the money supply. Only three of these variables are common to the preliminary and the current indicator — nonresidential building permits, the real value of loans through bills of exchange, and the index of the employment demand for technicians.

The estimates for industrial product are shown in table 3 and figure 3, which compares the estimates of the leading indicator and the actual growth rate of real industrial product. Unfortunately, the length of the monthly series allowed for confirmation of the dating of only three cyclical reversals — one in October 1973, when

Figure 3

INDUSTRIAL PRODUCT: COMPARISON BETWEEN THE GROWTH OF THE OBSERVED INDEX AND THE ESTIMATES OF THE LEADING INDICATOR, 1966-1977

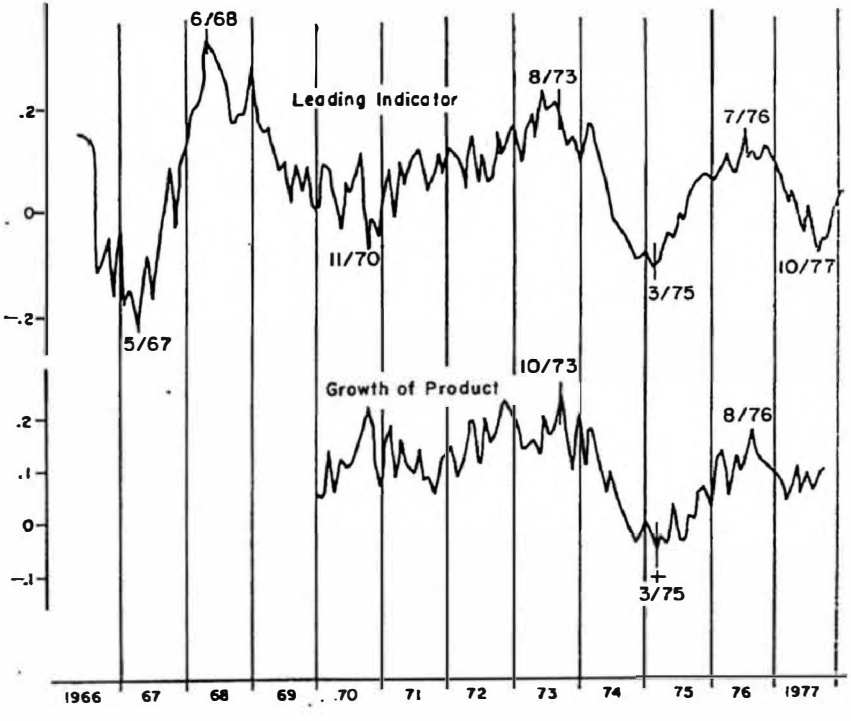


Table 3

Leading Indicator for Industrial Product: Annual Rates[a]

Year	January	February	March	April	May	June	July	August	September	October	November	December
1966	...	...	...	...	...	...	15.1	15.2	10.4	—11.8	—10.5	— 5.5
1967	—16.6	—31.5	—18.0	—15.6	—20.7	—21.4	— 8.8	—17.7	— 0.8	— 0.7	8.8	— 4.0
1968	8.5	10.3	18.7	19.5	22.2	32.5	30.0	28.0	26.5	16.8	16.5	18.2
1969	18.5	20.7	16.3	13.9	16.4	11.8	6.8	8.0	.8	8.4	3.8	7.7
1970	12.9	10.2	0.0	7.5	— 1.0	— 4.5	4.4	3.1	7.7	11.3	— 5.9	— 2.1
1971	— 5.0	2.4	7.1	— 1.6	8.7	4.0	0.5	11.2	8.3	2.6	5.3	10.0
1972	5.9	11.5	11.2	9.4	2.7	13.4	5.1	0.7	5.0	6.1	14.2	10.1
1973	13.6	15.2	7.7	15.1	10.8	13.4	21.8	18.2	20.3	19.5	14.9	11.7
1974	13.4	7.8	15.1	15.5	0.8	5.1	1.6	— 3.3	— 4.3	— 5.4	— 8.2	—10.5
1975	— 9.7	— 9.3	—11.8	—10.9	— 0.4	— 5.2	— 5.6	— 1.8	— 2.8	2.8	4.7	5.2
1976	5.0	7.6	12.1	8.5	7.0	12.3	16.4	7.6	9.8	8.1	0.2	11.0
1977	9.8	2.4	— 2.8	.1	— 5.8	—10.4	— 3.9	— 6.9	— 9.7	—12.2	—10.8	—12.0

[a] In units roughly approximate to the annual growth of the reference variable.

growth rates began to fall, another in March 1975, when they started to rise, and a third in August 1976, when they again declined. The indicator accurately dated the March 1975 reversal, but erred by two months and one month, respectively, with regard to the 1973 and 1976 turns. There is no evidence that the indicator gave any false warnings or failed to predict important reversals. Nonetheless, it will have to be tested further before decisive conclusions can be drawn as to its forecasting performance.

Tables 4 and 5 and figure 4 refer to industrial employment in São Paulo. The leading indicator is comprised of 11 variables, eight of which are also included in the indicator for industrial product, though with different leads and weights. With an average lead

Table 4

Industrial Employment in São Paulo: Composition of Leading Indicator

Variable	Average Deviation[a]	Standard Deviation[a]	Correlation with Reference Variable[b]	Lead (Months)[c]
Financing and Refinancing of Real Estate	.3065	.1035	.494	10
Value of Bills Protested: São Paulo	.2662	.3507	—.639	10
Rediscounts to Commercial Banks: Manufactures	.2780	.2326	.599	10
Credit Requests for Purchase of Consumer Goods: São Paulo	.0303	.2295	.837	9
Real Value of Textile Production	.0448	.0745	.715	9
Sales of Household Appliances	.1713	.1066	.472	9
Laminate Production (total)	.0744	.0815	.856	8
Real Value of Food-Industry Production	.0856	.0751	.451	8
Residential Building Permits (in square units)	.0321	.2963	.326	8
Loans through Bills of Exchange[d]	.2780	.1532	.728	7
Real Value of Chemical-Industry Production	.1742	.1050	.701	6

a/d See notes to table 2.

b The reference variable is the index of industrial employment in São Paulo.

c The average estimated lead is 8.6 months.

Table 5

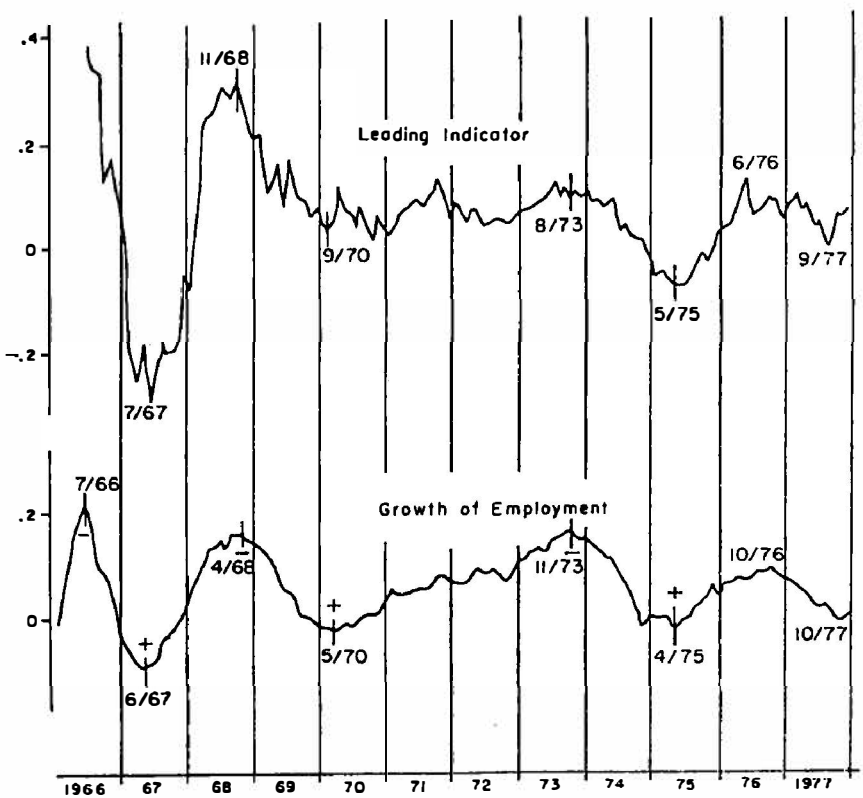
Leading Indicator for Industrial Employment in São Paulo: Annual Rates [a]

Year	January	February	March	April	May	June	July	August	September	October	November	December
1966	...	...	...	...	...	...	...	...	...	30.0	22.3	17.2
1967	9.1	6.2	8.7	-20.5	-25.5	-17.9	-27.6	-22.6	-17.5	-20.4	-20.0	-18.6
1968	-5.3	-7.9	3.9	17.3	24.6	25.2	28.3	30.5	29.5	28.0	31.9	27.9
1969	21.8	21.3	21.5	10.2	12.0	15.6	7.6	16.8	12.2	9.2	8.7	5.6
1970	6.9	5.4	2.9	4.7	11.7	7.0	7.0	3.8	7.6	5.4	.9	5.2
1971	4.2	3.3	4.1	6.0	6.8	7.8	8.6	8.0	8.7	9.2	12.6	10.0
1972	6.4	8.3	7.3	4.4	7.0	7.4	4.1	5.0	5.4	5.2	5.5	4.8
1973	6.8	6.9	7.2	7.5	8.1	8.9	9.9	12.0	9.1	10.7	8.9	10.1
1974	9.3	10.0	7.6	7.7	7.4	8.3	8.7	3.4	3.7	1.5	1.0	1.2
1975	- .9	-5.9	-5.3	- 5.3	- 7.3	- 8.5	- 7.8	- 6.7	- 4.8	- 2.6	- 2.0	- 3.0
1976	-1.5	2.9	3.0	4.2	6.1	9.9	6.5	6.1	8.8	9.5	8.9	8.4
1977	4.4	5.9	8.5	4.8	5.9	2.0	.3	1.5	- 5.2	- 4.9	3.0	3.3

a In units roughly approximate to the annual growth of the reference variable.

of over six months, the indicator performed extremely well, recognizing eight reversals in the period 1966-1977 — in July 1966, June 1967, November 1968, May 1970, November 1973, April 1975, October 1976, and October 1977. The last two reversals nearly coincided with turns in industrial product. The *ex post* accuracy of the estimates is clear from figure 4, which also depicts the fluctuations in the actual level of industrial employment. Only in the case of the November 1973 reversal was the indicator imprecise.

Figure 4
INDUSTRIAL EMPLOYMENT IN SÃO PAULO: COMPARISON BETWEEN THE GROWTH OF THE OBSERVED INDEX AND THE ESTIMATES OF THE LEADING INDICATOR, 1966 – 1977



Even a glance at the short cycles observed between 1966 and 1977 shows that the indicators erred only slightly. In table 6, the dates of reversal forecast by the leading indicators are compared to the effective indices for industrial product and employment. The 1970-1973 expansion began in May 1970 according to the employment data and in January 1971 according to the product data. In terms of employment the indicator dated the recovery six months too late, and in terms of product two months too early. The subsequent contraction in industrial activity (1973-1975) started in October or November, according to product and employment statistics, respectively. In both cases, the estimates of the leading indicators preceded the actual turns by three months. As to the recovery that began in March-April 1975, the product indicator was accurate and the employment indicator erred by only one month. The next reversal, in September-October 1976, was predicted two months early in the case of product and four months early in the case of employment. Finally, the (short) recovery witnessed in October 1977 has so far been confirmed by employment data only, but with an error of just one month. It must be emphasized that since the information required for the indicators is made available an average five to six months beforehand, the predictions concerning the 1976 and 1977 turns were reasonably accurate.

The forecasting ability of the leading indicators employed can hardly be assessed on the basis of their *ex post* confirmation of the cycles up to 1975. Their *ex ante* predictions for 1976 and 1977, however, are proof of their performance.

5 — Concluding remarks

The aim of this paper has been to outline a method for constructing economic indicators so as to reduce uncertainty concerning the future and thereby aid policymakers. Like any other, the method has its shortcomings. Hopefully these shortcomings have been limited by examination of the available data on the most important cycles in the recent past.

The leading indicators for product and industrial employment were constructed on the basis of 80 monthly series. The period covered is relatively short, with the longest series going back to 1965. However, the Brazilian economy has undergone far-reaching changes in the last two decades, and the cause-and-effect mechanisms among the variables have most surely been extensively altered. If this is the case, the relationships and mechanisms operative in the distant past are therefore capable of revealing little about those most influential at present and in the future.

Table 6

Cyclical Reversals in Industrial Product and Employment, 1966-1975

Reversal	Industrial Product				Industrial Employment		
	FGV[a]	FIBGE[b]	Indicator[c]	Error in Forecast[d] (Months)	FIESP[e]	Indicator[f]	Error in Forecast[g] (Months)
1	1966—[h]	...	...	...	July/66—	...	...
2	1967+[i]	...	May/67+	...	June/67+	July/67—	—1
3	1968—[j]	...	June/68—	...	November/68—	November/68—	0
4	...	January/71+	November/70+	+2	May/70+	November/70+	—6
5	...	October/73—	July/73—	+3	November/73—	August/73—	+3
6	1975[k]	March/75+	March/75+	0	April/75+	May/75+	—1
7	1976[m]	September/76—	July/76—	+2	October/76—	June/76—	+4
8	1977[n]	...	October/77+	...	October/77+	September/77—	+1

Note: Reversals are indicated by + (expansion) and — (contraction).

[a] According to the annual aggregate growth rate estimated by the Centro de Contas Nacionais of the Fundação Getúlio Vargas.
[b] According to the 12-month growth rate of the real value of industrial product estimated by FIBGE.
[c] According to the 12-month growth rate of the leading indicator listed in table 3.
[d] Difference between the timing of the reversal as pointed to by the FIBGE data and the leading indicator. A positive (negative) value reveals that the indicator erred by predicting the date of the reversal too early (late).
[e] According to the 12-month growth rate of industrial employment in São Paulo estimated by the Federação das Indústrias do Estado de São Paulo.
[f] Leading indicator listed in table 5.
[g] Difference between the timing of the reversal as pointed to by the FIESP data and the leading indicator.
[h] Annual growth rate estimated at 11.7%.
[i] *Idem*, 3.0%.
[j] *Idem*, 15.5%.
[k] *Idem*, 6.2%.
[l] *Idem*, 10.8%.
[m] *Idem*, 2.9%.

For this reason, plus the lack of statistics and the difficulty of obtaining uniform series prior to 1965, to have extended the analysis to earlier periods would have been too time-consuming and too costly in terms of computation expenses. Even so, we sought to provide an historical overview of the more important and longer-lasting cycles of the last 100 years. Although the data are admittedly poorer than those currently available, the timing and duration of the cycles are at least suggested. Judged in terms of recorded history, our estimates identified the major cycles with reasonable accuracy.

The method adopted for constructing the indicators follows that used by the National Bureau of Economic Research. The indicators registered an average lead of five to seven months, and scored well in forecasting the main cyclical reversals. Their performance was appraised according to two criteria: their ability to predict reversals, and their accuracy in dating these reversals. With respect to the first, they indicated precisely the same behavior as the reference variables for product and industrial employment. As to the second, they proved satisfactory, though not perfect, in dating the more recent reversals.

The structure of industrial wages in 1970 *

Paulo Vieira da Cunha **
Regis Bonelli **

1 — Introduction

Although a much researched topic, the determination of wages is still a subject of controversy, periodically calling forth a new wave of divergent opinion concerning: (1) the levels and rates of change in individual wages over time, (2) the share of wages in total income, and (3) the distribution of average wages among the various economic activities. Despite the links between the three distributions — that is, the personal, functional and sectoral distributions — the first two have received considerably more attention than the third. Indeed, after a flourish of contributions stemming from the industrial organization literature of the fifties, it is only recently that the analysis of average wages by sector has come under renewed interest.

The study of average wages is facilitated by the ready availability of data. Moreover, average wages can be analyzed together with the more usual economic indicators that have to do with firms or sectors, such as profit rates, the stock of capital, the productivity of labor and the size of industrial establishments. The very concept of individual wage determination presupposes some theory about the link between an individual's wages and those variables whose

* This paper was first presented at the 5th annual meeting of the Associação Nacional de Pós-Graduação em Economia (ANPEC), Rio de Janeiro, December 1977, and subsequently published as "Estrutura de Salários Industriais no Brasil: Um Estudo sobre a Distribuição de Salários Médios em 1970," *Pesquisa e Planejamento Econômico* 8 (April 1978): 117-67. The translation is a revised version of the original text, which was substantially improved thanks to the comments of D. Werneck, A. L. O. de Almeida, A. Mello e Souza, E. J. Reis and R. Ekerman.

** Of the research institute of IPEA.

roles can only be determined at the aggregate level — and this is, in itself, the subject of the debate we wish to explore. Even if individual wages reflect the overall productivity of labor, it is extremely difficult, except in the case of the most rudimentary production processes, to ascertain the productivity of a single worker as something distinct from, albeit related to, his occupation, age, sex or educational background.

In this study we consider two alternative, though complementary, hypotheses about the factors underlying the variation in average industrial wages: a structural hypothesis that is fundamentally based upon sectoral and regional differences in the organization of production, and a personal-characteristic hypothesis built around the distribution of human-capital attributes in the labor force. It should be emphasized at once that our intention is not to assess the superiority of one approach or the other (a task which, even at the empirical level, would require additional data and more precise statistical procedures), but merely to highlight some aspects that are frequently overlooked in the discussion concerning the distribution of wages.

2 — Analyses of interindustry wage structure: an overview

Interindustry wage distribution is important to understanding income distribution as a whole. Some of the first studies on the theme, which date back three decades, focused on the impact of unionization [1] or the level of economic activity [2] on the wage structure. Hiring policies were also seen to be relevant.[3] The fact that these policies were conditioned not only by the qualifications of the workers, but also by the economic structure of the firms (e.g. the ratio of the wage bill to the income generated, or the profit per unit sold) soon indicated that the simple theoretical models were in need of revision.[4]

The conclusions drawn by Ross and Goldner succinctly reveal the nature of the initial research in this field. Having examined

1 See, for example, A. M. Ross, "The Influence of Unionism upon Earnings," *Quarterly Journal of Economics* 62 (February 1948): 268-76.

2 See, for instance, A. M. Ross and W. Goldner, "Forces Affecting the Interindustry Wage Structure," *Quarterly Journal of Economics* 64 (May 1950): 254-81.

3 Similarities in wage structures in different countries, in various regions and cities in the United States, and in certain industries over time are examined by S. Lebergott, "Wage Structures," *The Review of Economic Statistics* 32 (1950): 279-85.

4 See S. H. Slichter, "Notes on the Structure of Wages," *Review of Economics and Statistics* 32 (February 1950): 80-91.

the behavior of average wages in approximately 50 industries over the period 1933-1946, they ascertained the following. (1) The average increase had been US$.60 per hour, and the deviation from the average was less than US$.10 in 35 industries. Wages had therefore accompanied economic growth as a whole rather than the expansion of individual industries. (2) Relative wage increases had been greater in those industries that were less unionized at the beginning of the period than in those that were more so. (3) Wages had risen most in the activities in which employment had also grown most.[5] Increases had been higher in heavy industry, with its oligopolistic market, than in those activities with competitive markets such as those for consumer goods.[6] Since unionization, growth of employment and oligopoly were observed in the same group of industries, the authors conclude that wage gains can surpass the average only in the presence of unionization, but that this factor alone is not sufficient to bring this about.[7]

Some 10 years later, Reder attempted to bring the discussion into line with conventional wage theory, noting that the literature was abundant in ad hoc hypotheses, some of which were in agreement with neoclassical price theory, but many of which were not.[8] The importance of Reder's argument, which concerns a competitive labor market where the various industries in the same locale offer the same pay to equally qualified workers, lies in the distinction between short and long-range features. At short term, wages should rise with increased employment due to labor-supply inelasticities. The association between growth of employment and differences in qualifications is therefore positive, since skilled labor should become relatively scarce. At long term, however, wage differentials should depend on skill mix alone, reflecting neither capital investments (total or per capita) nor the volume of employment.

As one might expect, Reder concentrates on applying the "competitive hypothesis"[9] summarized above and on showing the incon-

5 The opposite conclusion is reached in the well-known study by the OECD, *Wages and Labor Mobility* (Paris, July 1965), which questions labor competition theories and the efficiency of labor markets in allocating wages.

6 Concerning the influence of product markets on wages, see J. W. Garbarino, "A Theory of Interindustry Wage Structure Variation," *Quarterly Journal of Economics* 64 (May 1950): 282-305.

7 Ross and Goldner, pp. 280-81.

8 M. W. Reder, "Wage Differentials: Theory and Measurement," in *Aspects of Labor Economics* (New York: NBER, 1964), pp. 257-99. Reprinted in *Readings in Labor Market Analysis*, ed. J. F. Burton, Jr. *et al.* (New York: Holt, Rinehart and Winston, 1971), pp. 281-309. In this article, page references are to the latter.

9 "... the hypothesis that prices and quantities behave as though they were in long-run equilibrium under conditions of pure competition." *Ibid.*, p. 296, fn. 45.

sistencies in the "heterodox" hypotheses. His findings point, for example, to weak long-range correlations between average hourly wages and productivity (total, à la Kendrick) and profit rates. Furthermore, he denies that noncompetitive industries raise wages more readily than competitive industries (on the basis that the further monopolization of a group of industries between 1899 and 1937 was negatively, though not significantly, associated with increases in average wages). Finally, he notes a positive association between the interindustry wage hierarchy and the wealth (sic) of the skill mix in different activities.

The crux of Reder's argument is that previous studies had bypassed skill mix in favor of structural variables such as productivity, unionization, market structure and concentration, profits and hiring policies. Taking skill mix into account makes it possible to reconcile these heterodox interpretations with the competitive hypothesis, especially at long range. At short range, the performance of this hypothesis is less favorable due to the evidence being less clear and to the need for examination of specific cases.

Reder's approach is not supported by the OECD comparison of wage structures in the United States and 10 European countries.[10] In this study, the main conclusion is that the average wages in different industries are linked to relative changes in employment neither at short nor at long term. It is also contended that wage scales remain stable for long periods of time, whether viewed from the occupational, regional or interindustry standpoint. Furthermore, wage differentials are shown to exert little effect on labor transfers, and to be dependent on market concentration and profit levels.

The latter results were qualified by Weiss, who notes the theoretical ambiguities: the monopolist may offer relatively low wages due to production and employment restrictions, but should pay more owing to high profits.[11] Empirically, he verified that more concentrated industries pay higher wages to workers in similar occupations but obtain more skilled labor and workers with "socially desirable" (e.g. racial) characteristics.

As time passed, the approach changed. Thus, in the fifties, when both monopolies and unions gained increasing power in the United States, the demand for labor was the center of interest; accordingly, the industrial structure was analyzed from the point of view of its differential impact on wages. In the sixties, concern

10 OECD, especially pp. 85-118.

11 This may encourage unionization. Weiss notes that the degree of concentration is strictly related to the degree of unionization. "Concentration and Labor Earnings," in *Readings in Labor Market Analysis*, ed. J. F. Burton, Jr. *et al.*, pp. 344-61.

shifted from the structural (demand) side to the individual or human-capital (supply) side, with special emphasis on low-wage employment.[12]

An interesting attempt to reconcile these approaches was made by Watchel and Betsey,[13] who attribute to Bluestone[14] the dual-labor-market model that they adapt and estimate. Though simple in construction, it is difficult to provide empirical estimates of this model, which may be briefly described as follows. Assuming, as in figure 1, that the demand for labor on the part of three industries is represented by curves I_1, I_2 and I_3,[15] and that the differences in human capital are depicted by supply curves H_1, H_2 and H_3 (where H_3 stands for greater endowments than H_2, etc.), two models can be analyzed at the same time. According to one, the labor supply is stable and wages are determined by the industrial structure. According to the other, the demand for labor remains at the same level and wages are influenced by differences in skill.

The supply curves (H) reach the positions observed in the figure due to more qualified individuals receiving higher wages, as well as to their being relatively few in number. Thus, a worker endowed with skill H_3 would earn wage w_{13} in industry *1*, w_{23} in industry *2*, and w_{33} in industry *3*. Equally qualified workers therefore receive different pay depending on the industries in which they are engaged, and wages may vary within a given industry depending on the skill of its employees. An important implication is that less qualified workers, if engaged in certain activities, may command the same wages as more qualified persons in other activities. This is owing to the separation of labor markets and to a lack of mobility that may be considerably prolonged by artificial educational require-

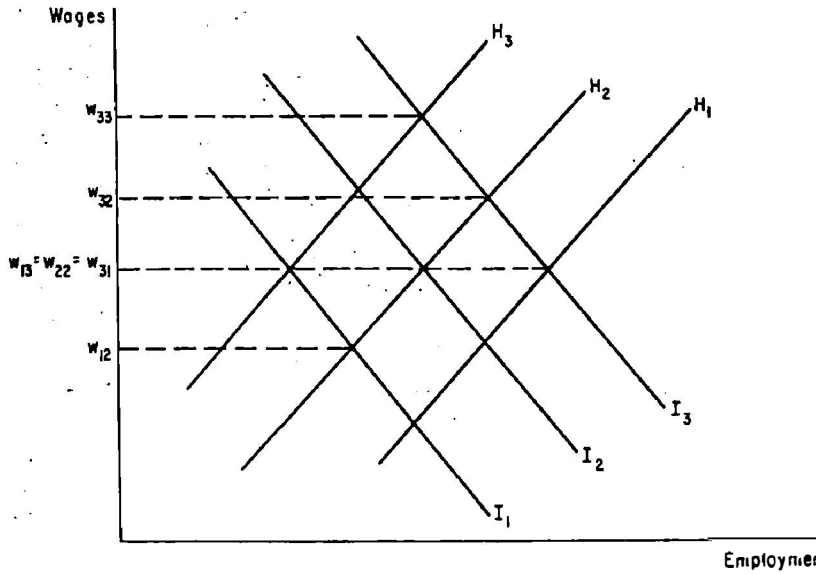
12 Among which might be cited the well-known studies by Schultz and Becker, as well as the overview by J. Mincer, "The Distribution of Labor Incomes: A Survey with Special Reference to the Human Capital Approach," *Journal of Economic Literature* 8 (March 1970): 1-26.

13 H. M. Watchel and C. Betsey, "Employment at Low Wages," *Review of Economics and Statistics* 54 (May 1972): 121-29.

14 B. Bluestone, "The Tripartite Economy: Labor Markets and the Working Poor," *Poverty and Human Resources*, July-August 1970. The concept of a dual labor market was actually developed by P. Doeringer and M. Piore, mainly in *Internal Labor Markets and Manpower Analysis* (Washington, D.C.: Heath and Co., 1971).

15 The differences between the demand curves reflect the differences among industries as to technology, profit rates, market concentration, unionization, etc.

FIGURE I



ments, scarcity of information about the job market, the financial cost and risks of moving from one place to another, discrimination, union regulations, and so forth.

Using these considerations as their point of departure, Watchel and Betsey developed a wage-determination model capable of decomposing demand effects (i.e. structural variables) and supply effects (personal characteristics). So as to minimize the problem of multicollinearity, a two-stage regression analysis is performed. In the first stage, the effects of personal characteristics (job tenure, color, age, sex, education, and marital status in this case) on wages are eliminated, and an attempt is made to assess the influence of structural variables on this wage variable from which the impact of personal characteristics has been removed. Then, in the second stage, another regression is performed in which the dependent variable is the residual of the previous regression, and the independent variables are structural (i.e. industry, region, size of the city where the firm is located, occupation, union affiliation). The process may

subsequently be reversed, by beginning with the structural variables and using personal characteristics as the independent variables in the second stage.[16]

The results obtained in the first stage point to the relative importance of education, job tenure and color, which together explain 34% of the variance in wages. The outcome of the second-stage regression (in which the residual of the preceding regression serves as the dependent variable) indicates the importance[17] of industry and occupation, and, secondarily, of region and city size; the multiple correlation coefficient is .28, and all the variables are significant at the 1% level. On the basis of this evidence, the authors conclude that once the effects of personal characteristics have been removed, industrial structure explains wage differentials to a substantial degree. Reversing the process (i.e. introducing the structural variables in the first stage) yields similar results. Industry and occupation, and secondarily region, city size and union affiliation, jointly account for 35% of the variance in wages. The relative importance of color and education stand out in the second stage.[18]

Watchel and Betsey conclude their work by drawing some interesting implications with respect to low-wage employment. In particular, they note that education and training programs are based on the premise that workers have not invested enough in themselves — an approach which resembles that prevalent in the nineteenth century, when poverty was considered the result of personal failure or sheer laziness.[19] Thus, the emphasis has been on improving human capital, rather than on how to employ this capital in the labor market. Most important, in these programs, the structural characteristics of the labor market and their influence on wage levels have been slighted. Although the study by Watchel and Betsey is open to criticism from several angles,[20] it must be recognized that the authors make a coherent attempt to show the complementary nature

16 All the variables except wages are explicit, and were specified using dummies.

17 Measured by the beta coefficients.

18 At this stage, the correlation coefficient is .25. We are considering this analysis in greater detail owing to its similarity to our model.

19 This is akin to blaming the victims for their own plight. It might be noted in passing that similar attitudes still prevail in certain circles in Brazil.

20 See E. Kalachek and F. Raines, "The Structure of Wage Differences among Mature Male Workers," *The Journal of Human Resources* 11 (Fall 1976): 484-506. In addition to the problem of identification, the model contains some theoretical inconsistencies. If the labor markets were truly segmented, for example, figure 1 would have to include a different set of labor supply curves for each demand curve. Moreover, if the markets were not segmented, the disequilibrium shown in the figure would not occur because competitive forces would equalize a given wage on all markets.

of structural variables and personal characteristics, and that their conclusions are based on observation of real economies rather than on textbook theories.

Turning to the Brazilian case, the only studies of which we are aware are those by Senna,[21] parts of works on industrial labor by Moura Castro and Mello e Souza,[22] and, later, by Macedo.[23] Macedo analyzes average wages at the two-digit level as a function of (1) a Theil index of concentration, (2) an economy-of-scale parameter, (3) the level of effective protection of the industry, and (4) its capital/labor ratio (measured in terms of horsepower per worker). Of the four, only effective protection was found to be significant.

Moura Castro and Mello e Souza, basing their analysis on data from the Ministry of Labor (Lei dos 2/3[24]), consider interindustry wage differentials as a function of average education and the size of the industry (measured by the proportion of workers engaged in firms having less than 50 employees). Although the analysis is limited to 21 observations, average education and size of industry together account for 73% of the average-wage differentials, and singly respond for 68% and 35%, respectively. The rate of return to education rises significantly with the size of the firm, thereby suggesting a notable interaction between schooling and the size of the firm in the formation of average wages. This is especially true of training courses such as SENAI,[25] and to a lesser extent of traditional academic curricula.

21 J. J. Senna, "Análise dos Diferenciais de Salários entre os Diversos Ramos da Indústria Brasileira," in *A Economia Brasileira e Suas Perspectivas, 1976*, Estudos APEC, n.º 15 (Rio de Janeiro, 1977), pp. 283-93. Two other analyses of the Brazilian case are being bypassed because they focus on different aspects of the issue — the breakdown of average wages and the demand for labor, respectively. The first is an article by M. da Mata and E. L. Bacha, "Emprego e Salários na Indústria de Transformação, 1949/1969," *Pesquisa e Planejamento Econômico* 3 (June 1973): 303-340. The second study is by R. B. M. Macedo, "Models of the Demand for Labor and the Problem of Labor Absorption in the Brazilian Manufacturing Sector" (Ph.D. dissertation, Harvard University, 1974).

22 C. M. Castro and A. Mello e Souza, *Mão-de-Obra Industrial no Brasil*, Coleção Relatórios de Pesquisa, n.º 25 (Rio de Janeiro: IPEA/INPES, 1974), pp. 270-82.

23 R. B. M. Macedo, "Distribuição Funcional na Indústria de Transformação: Aspectos da Participação do Trabalho" (Ph.D. dissertation, Universidade de São Paulo, 1977), pp. 169-74.

24 Promulgated in the late 1940s, this law stipulated that two-thirds of firms' employees had to be of Brazilian origin. To guarantee adherence to the law, firms were surveyed annually.

25 SENAI (Serviço Nacional da Aprendizagem Industrial) is the official nationwide industrial training program.

Senna investigates the structure of average wages in 18 manufacturing industries, also using the Lei dos 2/3 as his data source. From the outset he notes that, given the results of a previous study,[26] the differences in the levels of average education in the various industries should be important in explaining wage differentials; this is confirmed by the regression analysis, which indicates that average education accounts for 43% of the variance in wages. That a measure of industrial concentration is the next variable introduced reveals that Senna opted for a model which brings together human-capital and structural characteristics. The inclusion of this variable raises the coefficient of correlation to 67%.[27] A third variable — technology (expressed as the ratio between royalties paid and industrial value added[28]) — further raises the coefficient to 71%, though the coefficient of this variable itself is significant at only 10%. In turn, replacing technology with the growth rate of product over the previous three years increases the regression coefficient to 73%. When the four explanatory variables are considered together, R^2 rises to 76%, but there is a drop in the significance of technology (now nonsignificant) and growth of product (significant at the 10% level). Thus, although Senna stresses the role of education,[29] his model also reveals the impact of structural variables on wages.[30]

3 — Average industrial wages: explanatory models

From the preceding discussion, it should be clear that empirical models for analyzing average wages have generally included variables intended to reflect the differences in the structures of production and in the distribution of personal characteristics among the various sectors. Nevertheless, inclusion of both factors is not in itself a

26 J. J. Senna, "Escolaridade, Experiência no Trabalho e Salários no Brasil," Ensaios Econômicos da EPGE, n.º 22 (Rio de Janeiro: Fundação Getulio Vargas, 1976) (Mimeographed). In this study, based on individual wage data, the formal education variable explains about 25% of the wage differential. Inclusion of the experience variable raises the figure to 36%.

27 Concentration is measured by the share in total sales of the four largest firms in the industry. Alternative measures, expressed in the form of dummy variables, yield similar results. The positive value of the coefficient confirms the association between concentration and wages, and is therefore in accord with the hypotheses of Weiss and Garbarino.

28 Valor da transformação industrial (VIT).

29 Senna contends that the role of education "comprises additional evidence of the importance of human-capital models to the analysis of the problem of income distribution, with the relevance of formal schooling standing out" (p. 292).

30 It would be interesting to know what would happen if a variable other than average education were introduced first.

theory of wages; their juxtaposition merely brings one nearer to an empirical reality that may support many — often contradictory interpretations. The underlying reasoning may be expressed as follows.

Let us, for the moment, leave aside what happens to the pool of unemployed labor and as yet unoccupied new vacancies. In this case, there would be a direct correspondence between the distribution of wages, reflecting the structure of jobs, and the personal qualifications of the workers who hold these jobs. Note that while it is conventional to argue that wages are a function of personal characteristics, it would then be just as valid to postulate the inverse — that is, that wages are set for specific jobs and vary only slightly (and erratically) with differences in the qualifications of the workers who hold them. From this viewpoint, wages would be influenced by personal qualifications only to the extent that these are the means by which the worker gains access to the desired position. To differentiate between cause and effect, it would therefore be necessary to analyze the access mechanisms themselves — in other words, the means by which the labor supply is transformed into an active labor force.

Nevertheless, in an analysis of average wages, the object of study cannot be the market mechanism itself, if only because it is unobservable. Rather, it should be the immediate result of this mechanism. Since the data itself, the average wage, is an outcome of unobserved mechanisms, ambiguities will arise when interpreting experimental results based on this information. In other words, we are faced with an identification problem; there are at least two ways to conceptualize and decompose the average-wage differentials, but the reduced forms are, as we shall see, the same. Thus, if there is anything original in our approach, it lies in our demonstrating that the two conclusions (that wages are a function of personal characteristics or that they are set for given jobs) can be reached using the same variables. Specifically, our aim is to criticize an interpretation which, due either to force of habit or lack of alternatives, has become common in labor market analysis: the insistence on seeing a cause-and-effect relationship between workers' qualifications and wages.

This association has been made so often that it has come to be taken as fact rather than theory. However, it is only recently, with the development of human-capital theory, that the causal relationship between investments in education or training and returns in the form of wages (operating through changes in the productivity of labor) has been fully formalized. It is therefore justifiable to link the argument "wages are a function of personal characteristics" to this theory. In doing so, we can, without loss of generality, limit

the range of characteristics to those that effectively embody the worker's human capital — that is, general knowledge and specific training.[31]

In contrast, the link between wages and jobs has received much less attention, and more in industrial relations literature than in the usual economic treatises. This may in part be due to the fact that this is a rather unattractive theory to those who are habituated to the liberal apology of equal opportunity with shared goals for labor and management. Both notions would be seriously compromised were it admitted that the division of labor is the outcome of the structure of production — "of specialization to serve the machines" — rather than the other way around. If wages are a function of jobs, the employment structure, which is controlled not by the workers but by those who hire them, functions as an all-important mediator between personal characteristics and wages. From this standpoint, a theory of wages should begin from the organization of the work process and then extend to the qualifications of the labor force necessary to this process. The political significance of this approach is therefore clear. It questions the feasibility of redistributing wages, and ultimately income, via changes in the labor supply.

Having made these observations, we shall proceed as follows. First, we consider how and why structural characteristics are introduced into the usual human-capital model. Second, we summarize some of the criticisms that have been directed, in this context, against human-capital theory. Finally, and on the basis of these remarks, we outline the principal arguments of what could be, but is so far an admittedly tentative, alternative theory.

In empirical applications of human-capital theory, structural characteristics are included primarily for control purposes. They should serve to delete (1) transitory effects due to fluctuations in demand,[32] (2) nonpecuniary and non-wage benefits,[33] and, above all, (3) the legal-institutional factors that influence the operation of market forces.

[31] We are not overlooking innate ability, but assuming that it exerts more influence on the production of human capital than on wages. See Z. Griliches and W. Mason, "Education, Income and Ability," *Journal of Political Economy* 80 (May/June 1972): S74-S103.

[32] S. Danziger and M. Weinstein, "Employment Location and Wage Rates of Poverty Area Residents," *Journal of Urban Economics* 3 (April 1976): 127-45.

[33] C. J. Duncan, "Earnings Functions and Nonpecuniary Benefits," *Journal of Human Resources* 11 (Fall 1976): 462-83; R. Lucas, "Hedonic Wage Equations and Psychic Wage Returns to Schooling," *American Economic Review* 67 (September 1977): 549-58.

Furthermore, even in the human-capital model, structural characteristics are important determinants of the price of labor whenever there is a systematic relation between technology and the amount of on-the-job training required for efficient production. A persistent problem in the theory is the difficulty (some would say the impossibility) of measuring, independent of age and experience, the amount and value of non-formal training. Considering the overall importance of on-the-job training as a form of qualification,[34] it is necessary to include variables in the empirical model that indicate the effect of such training on wage differentials.[35] Thus, in the usual wage equations, it is not surprising to find a positive and statistically significant association between the volume of capital (a standard proxy for technological complexity) and the level of wages — a result that, in the interpretation given by the theory, reflects the amount of specific training in the industrial occupations.[36]

When included in this manner, technological determinants are exogenous to the model. Consequently, where markets operate efficiently, it can be assumed that there is a strict and measurable relationship between the worker's wage and his (marginal) contribution to the total earnings of the firm.[37] Since this approach rests on relating the worker's relative productivity[38] to his stock of human capital, accurate measurement of the latter would make

[34] According to Mincer, "there is evidence that work experience is much more important than age in affecting productivity and earnings." He views "productivity-augmenting work experience as an investment phenomenon. The assumption of costless opportunities for augmenting productivity, which is sometimes implied in the notion of 'learning by doing,' cannot be descriptive of labor markets where labor mobility is the norm rather than the exception." J. Mincer, *Schooling, Experience and Earnings* (New York: Columbia University Press, 1974), p. 65.

[35] "While measures of on-the-job training exist, they are simplistic. Age or job tenure can hardly be expected to provide an accurate proxy." M. L. Wachter, "Primary and Secondary Labor Markets: A Critique of the Dual Approach," *Brookings Papers on Economic Activity*, 1974, n.º 3, p. 654.

[36] The theory behind this approach is presented by S. Rosen in "Learning and Experience in the Labor Market," *Journal of Human Resources* 7 (Summer 1972): 326-42. Rosen argues that workers choose jobs not only for their pay and non-pecuniary benefits, but also for their training opportunities. Since observed wages have different work/training ratios, they must be adjusted when estimating the income gradient.

[37] Leaving aside short-term fluctuations on the markets for goods and labor, as well as possible lags associated with training that extends over more than one period.

[38] That is, the performance of workers with different qualifications in relation to the same tasks. Note, however, that neither wage levels nor the total wages in the economy are being considered. These, as Mincer recognizes, are determined by the interaction of supply and demand forces in a way that is yet to be duly explained.

explicit consideration of job characteristics unnecessary. Each wage would correspond to a level of productive capacity established by the personal characteristics of the worker in competition with those of the other members of the labor force.[39]

This belief that the worker himself sets his tasks and consequently his wage through the competitive process, even in modern industrial firms, is the theoretical foundation of all human-capital models.[40] It is from this perspective that Eckaus calls attention to the fact that in order to understand how these models operate, it is necessary to imagine a sort of production process in which the firm profits most not from designing (and imposing) a relatively fixed structure of jobs (and of social relations uniting them), but from a constant flow of labor among a wide range of tasks, each pursued by, and created for, a specific "type" of labor, with each "type" being defined by a given level of formal education or training and a corresponding wage.[41]

These hypotheses allow for the following simplification of the wage-formation process.[42] Once the price of each of the personal characteristics demanded in the market (i.e. by the universe of competitive firms) has been determined, the rate of return on specific investments in human capital is simultaneously established. Moreover, considering that in the short run new investment projects can affect the stock of human capital only marginally, it can be said that the supply of labor is, in itself, a distribution of wages. It is only through changes in the rates of return to human-capital investments that there are changes in wages, and these occur only as a result of the confrontation of aggregate supply and demand for

[39] The same argument is applicable to the average-wage differential. The industries that require more qualified labor offer higher average wages, in direct proportion to the intrasectoral distribution of the types of manpower.

[40] This premise stands even in the "complex" models that take the organization of labor into account. With regard to assembly lines, for example, Akerlof "surrealistically" (sic) hypothesizes that "velocity" replaces "working conditions" and "educational level". The better the worker, the higher the velocity he can tolerate. At equilibrium, therefore, no worker will want to transfer from one assembly line to another that operates at a different speed. See G. Akerlof, "The Economics of Caste and of the Rat Race," *Quarterly Journal of Economics* 90 (November 1976): 599-617.

[41] R. Eckaus, *Estimating Returns to Education* (Berkeley: Carnegie Foundation for the Advancement of Teaching, 1973), pp. 64-65.

[42] The introduction of stochastic elements and the imperfect measurement of the variables does not affect the bases of these propositions. The use of indices and signals allows for equilibrium at more than one point. However, as Spence maintains, wages are always equal to their marginal products. M. Spence, *Market Signaling* (Cambridge: Harvard University Press, 1974), p. 176.

labor. In other words, the wage distribution has been delineated without regard to the types of jobs existing in the economy — jobs that, as far as we know, are logically established in line with the production requirements of the firm, rather that with the personal characteristics of the workers.

Unquestionably, the worker can choose among employment opportunities within the limits of his own attributes and the prevailing market conditions. Once engaged by a firm, however, the decision as to where and how to apply his labor is no longer his.[43] It is in this sense that the labor relationship is strictly skewed and coercive. Skewed because labor is not productive unless it is exercised in a given production structure; hence, productivity cannot be bargained outside the labor relationship. Coercive because the firm alone allocates workers to any number of hierarchically organized positions that are determined beforehand and as a result of its production structure.

What is important, therefore, is to select the best-suited worker for each job. To the extent that the firm is unable to contract workers with the desired traits at the wages currently offered (or, on the contrary, to the extent that there is an oversupply of such workers), the firm will modify its wage structure. Thus, as in the preceding argument, there is allowance for competitive market forces, though with a fundamental difference: the dependence of labor on wages as its only means of survival implies that it is subordinate to production.

What is striking in the modern industrial enterprise, however, is not only its employment of a large number of formally prepared workers, but its transformation of labor into the skills and attitudes needed for the production effort and adapted to the maintenance and legitimization of its specific work process. The result, as Thurow[44] points out, is that the labor market is not the locus of competition for given jobs. Rather, it is a market that brings together the supply of trainable workers and a demand for services that is, simultaneously, a supply of training. Under these circumstances, training opportunities are linked to the creation of jobs that require specific skills. In other words, the supply and demand curves are not independent because the definition of jobs antecedes the demand for training; without access to jobs, there is no opportunity for training.

[43] Strictly speaking, power over the wage worker is limited by labor legislation alone. Nevertheless, since firms usually strive to legitimize their positions before their employees, they rarely act in an arbitrary manner or show open discrimination.

[44] L. Thurow, *Generating Inequality* (London: Macmillan, 1976), pp. 79 ff.

Despite what has been said so far, internal job structures are not the immediate result of a given production process or a given technology. Technology alone does not determine the labor process; it merely defines the range of possibilities. Even so, internal job structures are not flexible, at least not in the sense that they respond smoothly to external or internal pressures. The very idea of a production structure militates against the supposition that the same goods can be arbitrarily produced in many ways, or that factors can be used in different proportions depending on price changes. Labor-market patterns, especially when examined at the level of a specific configuration of firms, should be viewed as the historical outcome of a continuous process that, among other things, transforms potential labor into an active labor force.[45]

It has been repeatedly emphasized that cost minimization requires periodic reorganization of the labor process, and that such reorganization tends to follow the adoption of new technologies. However, what is often overlooked is that some technologies are introduced not as a result of market competition, but as a means towards increasing control over the labor process. Legal restrictions on the length of the work day, together with safety regulations, have diminished, though unfortunately not eliminated, the practice of extensively augmenting labor's contribution during the production process. Its intensive utilization has therefore become increasingly important; the worker has been obliged to perform his tasks at a faster pace, which calls for greater efforts in the form of mental concentration and/or manual dexterity.[46]

Work hours are transformed into productive labor through the creation of control systems based on ever more bureaucratic, hierarchical and authoritarian relations. This signifies an internal distribution of responsibility and prestige, and a corresponding wage scale. Once this distribution is linked to the technical production requirements of the firm, the job structure is defined. Thus, in hiring its employees, a firm aims not only to attend to its production needs, but also to find workers with the behavioral and disciplinary traits best suited to its jobs.[47]

45 See, for example, K. Stone, "The Origins of Job Structures in the Steel Industry," in *Labor Market Segmentation*, ed. R. Edwards *et al.* (Lexington: D.C. Heath, 1975), pp. 27-84, or the broader analysis by H. Braveman. *Labor and Monopoly Capital* (New York: Monthly Review Press, 1974), pp. 139-52.

46 It should be pointed out that firms often try to achieve the same results by allowing the worker to co-opt, e.g. for flexible hours in the case of office personnel.

47 A more detailed discussion, together with an empirical example, is presented by R. C. Edwards, "Individual Traits and Organizational Incentives: What Makes a 'Good' Worker?" *Journal of Human Resources* 9 (Winter 1976): 51-68.

From the preceding, it is clear that the relative position of the worker depends more on his job than on his personal characteristics. Nonetheless, specific jobs are typical of a single firm or a group of firms belonging to the same sector. The distribution of jobs, and the average wage data that reflect this distribution, are thus determined by the structure of the industry.

Even so, it would be erroneous to conclude that there is no systematic relationship between job distribution and professional qualifications. Some skills are acquired on the job, and this process is facilitated by previous formal education. In fact, this is one of the ways through which private capital appropriates the benefits of publicly provided services, in this case, subsidized general education programs. In addition, differences in formal education legitimize the social structure within enterprises.

The job structure of the firm, together with its selection and promotion procedures, reinforces the social differences established through the educational process. For instead of acting as a social leveler, this process in fact reproduces pre-established differences as it prepares students — technically and psychologically — for different positions in the production structure.[48] Still, although formal education is one of the most important factors in job selection[49] (and consequently in explaining wage differentials), it is obviously more relevant for distinguishing new entrants into the labor market than in classifying the existing pool of experienced workers. For these, what is likely to be relevant is the acquired position in the job structure, the record of work experience. Once workers conform to the organizational pattern of a particular firm or sector, to its required attitudes and skills, they can contribute more to the collective effort, thereby minimizing costs and maximizing productivity. What they get out of this, at best, is a chance to move up the

[48] S. Bowles contends that "the social division of labor — based on the hierarchical structure of production — gives rise to distinct class subcultures. The values, personality traits and expectations characteristic of each subculture are transmitted from generation to generation through class differences in family socialization and complementary differences in the type and amount of schooling ordinarily attained by children of various class positions." "Unequal Education and the Reproduction of the Social Division of Labor," in *Social Mobility*, ed. A. Coxon and C. Jones (Harmondsworth: Penguin Education, 1975), p. 273.

[49] This is partly due to self-selection on the part of the workers, i.e. to their access to information concerning ob vocancies. On this point, see S. Bowles, "Understanding Unequal Economic Opportunity," *American Economic Review* 64 (May 1973): 346-58.

hierarchical (and wage) scale.[50] This is undoubtedly to an individual's gain, but it is also a means for furthering the goals of the enterprise, a mechanism that allows it to elicit the most from its labor.

In sum, according to both this and the preceding interpretation, average sectoral wages are the outcome of a combination of structural and personal characteristics. The result is the same; the implications, however, are vastly different.

4 — Estimates and analysis of the results

Throughout this paper we have argued that the formation of average wages rests on both the distribution of skills and on the characteristics of the industrial structure. In the following empirical analysis, the first group of variables measures the distribution of age, education and sex.[51] As is customary in studies of this kind, the age variables (I_m) are used to indicate experience *and* skills acquired on the job. It also serves to capture the effects of changes due to the life cycle of the workers.[52] In a strict sense, the educational distribution (E_n) classifies workers according to their formal qualifications, but it also reflects the influence of socio-economic background. Sex composition (H) is included to capture qualitative differences in the employment relationship, an index for possible job discrimination. By itself, however, this variable cannot separate wage differentials that arise from different levels of qualification from those due to different pay for equal work.

The relationship between the industrial structure and average wages is gauged through the following variables: the proportion of the sectoral labor force engaged in production (P), the average size of the firm in the sector (T), gross profits (M), and fixed capital per worker (K).[53] Inclusion of the first variable (P) is justified by the fact that we are not interested in conducting separate analyses of the wages of production and nonproduction workers.

50 In a study of the São Paulo automotive industry, it is concluded that "even in the modern firm, where specialized labor is the focal point of the production process, profissional qualifications are correlated with wages." J. Gonçalves, "Perfil do Operariado numa Empresa de Indústria Automobilística de São Paulo," *Contexto*, n.º 3 (July 1977): 34.

51 The measurement of the variables is described in appendix 1.

52 Estimation by age group makes it unnecessary to include a quadratic term to cover the negative effects of the higher age brackets on wages.

53 Although it would be desirable to include other structural characteristics, such as capital ownership and the concentration of production, data restrictions preclude this possibility.

Moreover, despite importante differences in the distribution of the two types of labor, they are complementary and can be treated as a heterogeneous aggregate. The size variable (T) is measured by the average volume of employment per firm in the sector, and serves to underline the relationship between size, the division of labor, and the workings of a hierarchical production structure. Gross profits per worker (M) may reflect a variety of factors. Assuming a single rate of profits for the whole economy, it simply mirrors the capital endowment of the sector. In the presence of an oligopolistic market structure (as is the Brazilian), it reflects differences in the power of the various sectors and states. On the one hand, therefore, gross profits are related to the organization of the markets; that is, it indicates the disparities in profits that are characteristic of a heterogenous industrial structure. On the other hand, since the surplus is a bargainable share of production and is the object of labor-management confrontations, gross profits should also indicate the relative strength of the organizational power of labor. The stock of capital per worker (K), though difficult to measure, is included for two reasons. To begin with, the hypothesis of equal profit rates for all sectors (which would lead to a high degree of collinearity between this variable and gross profits per worker) is highly questionable.[54] This alone is an important indication of differences in the technological structure of the economy. Moreover, because the stock of capital is a rough measure of the average productivity of labor, different production processes should be related to significant differences in K.

In addition to these variables, a final one was included to capture qualitative differences between the various states, such as variations in the minimum wage, the degree of its enforcement, availability of social overhead capital, in sum, the relative level of development.

The complete model, to be estimated by ordinary least squares, may be written as follows:[55]

[54] See appendix 2 for the coefficients of correlation between the independent variables.

[55] There are two reasons for using logarithms. In the case of the dependent variable, they clarify the distribution of wages in the sample. In the case of the independent variables, logs help correct the far from normal distributions, which are roughly log-normal. Logarithmic transformation also brings the effect of these variables (especially "average size of the firm") into line with expectations; thus, an increase from 20 to 120 workers will have a greater impact on wages than an increase from 10 020 to 10 120 workers.

$$Y_{jk} = a + \sum_{m=1}^{5} b_m I_{jkm} + \sum_{n=1}^{6} c_n E_{jkn} + dH_{jk} +$$

$$+ eP_{jk} + fT_{jk} + gM_{jk} + hK_{jk} + i\ SM_{jk} + u_{jk} \qquad (1)$$

where

J = industry ($j = 1, \ldots 19$)

k = state ($k = 1, \ldots 21$)

Y = average wage (in logs)

I_m = percentage of workers in each of five age groups (20-29, 30-39, 40-49, 50-59, 60+)

E_n = percentage of workers in each of six educational groups (incomplete or complete primary, secondary or university)

H = percentage of male workers

P = percentage of production workers

T = average size of enterprise — number of workers per firm (in logs)

M = gross profits per worker (in logs)

K = capital/labor ratio (in logs)

SM = regional minimum wage as a percentage of the highest minimum wage in effect in the country

According to the preceding arguments, the signs of the estimated regression coefficients should all be positive, with the exception of those for the oldest age group (I_5), the lowest educational level (E_1), and the percentage of production workers (P).

One of our aims is to empirically evaluate the separate contributions of the two groups of variables. The most convenient way to do this is to estimate not the full model, but rather two partial ones, each for a set of variables. The result will be a partition of the total variance models, although, of course, the estimated coefficients are likely to be biased due to the misspecification problem. However, in estimating the personal-characteristic equation, we observed a very high degree of multicollinearity, not a surprising result when dealing with income or average-wage functions. What

is more, the problem was particularly acute in the case of the educational variables, as can be seen from the results shown in table 1.

It is interesting to note that, although the model is misspecified and marked by multicollinearity, the variables included account for sizable shares of the variance in average wages: 64% in Brazil as a whole, 51% in the North-Northeast (N-NE) and 57% in the South-Southeast (S-SE). The sum of the beta coefficients reveals

Table 1

Average Wages and Personal Characteristics: Beta Coefficients[a] *and Levels of Significance*

Variable	Brazil (N = 327)		North-Northeast[b] (N = 200)		South-Southeast[c] (N = 127)	
	Beta	%	Beta	%	Beta	%
% Age 20-29	.338[d]	21	.420[d]	27	.432[d]	37
% Age 30-39	.173[d]	11	.189[d]	12	.221[d]	19
% Age 40-49	.193[d]	12	.171[d]	11	.344[d]	30
% Age 50-59	.089[d]	6	.091[d]	6	.231[d]	20
% Age 60 +	—.158[d]	—10	—.152[e]	—10	—.181	—16
% Primary: Incomplete	.061	4	.081	5	—.304	—26
% Primary: Complete	.466[d]	30	.182[d]	12	.148	13
% Secundary: Incomplete	.055	3	.138	9	—.035	— 3
% Secundary: Complete	.010	1	—.026	— 2	—.149	—13
% University: Incomplete	.133[e]	7	.149[e]	10	.087	8
% University: Complete	.235[d]	15	.228[d]	15	.316[d]	27
% Males	—.001		—.064	4	.050	4
Σ	1.574	100	1.535	100	1.160	100
R^2	.641		.511		.569	

a Beta = $\frac{b\sigma_x}{\sigma_y}$, where b is the coefficient of regression estimated by least squares, σ_x stands for the standard deviation of the independent variable, and σ_y for that of the dependent variable.

b Includes the states of Amazonas, Pará, Mato Grosso, Goiás, Maranhão, Piauí, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Alagoas, Sergipe, Bahia and Espírito Santo.

c Includes the states of Minas Gerais, Rio de Janeiro, Guanabara, São Paulo, Paraná, Santa Catarina and Rio Grande do Sul.

d Significant at the 1% level.

e Significant at the 5% level.

the expectedly high multicollinearity,[66] and the *t* tests show that a fair number of the coefficients (especially those of the educational variables and in the S-SE) are not significantly different from zero. With regard to the individual contribution of each variable to the total variance, it is illustrative to glance at the second column of the table. These percentages indicate the decomposition of the sum of the beta coefficients. For Brazil as a whole, "complete primary education" and "age 20-29" are by far the most important variables on this score. The latter also stands out in the N-NE and in the S-SE. Particularly in the S-SE, the coefficients of "age 20-29", "incomplete primary education" (with a negative sign, as expected) and "complete university education" are also high.

Table 2

Average Wages in Brazil: Beta Coefficients and Levels of Significance of Personal and Structural Characteristics

(N = 327)

Variable	Equation (1)		Equation (2)		Equation (3)	
	Beta	%	Beta	%	Beta	%
% Age 20—29	.211[a]	11	.321[a]	22		
% Age 30—39	.087[a]	5	.185[a]	13		
% Age 40—49	.114[a]	6	.216[a]	15		
% Age 50—59	.052	3	.070	5		
% Age 60 +	—.033	— 2	—.220	—15		
Education 1	.199[a]	10	.298[a]	20		
Education 2	.138[a]	7	.147[a]	10		
Education 3	.094[b]	5	.436[a]	29		
% Male Workers	.158[a]	8	.030	2		
% Production Workers	.248[a]	13			.191[a]	17
Log *K*/Worker	.052[b]	3			.087[a]	8
Log Workers/Firm	.163[a]	8			.185[a]	16
Log Gross Profits/Worker	.221[a]	11			.258[a]	29
Relative Minimum Wage	.267[a]	14			.342[a]	30
Σ	1.971	100	1.483	100	1.127	100
R^2	.838		.604		.719	

a Significant at the 1% level.

b Significant at the 5% level.

[66] In the absence of multicollinearity, the sum of the beta coefficients is equal to *R*. Note that, in table 1, there is less multicollinearity in the equation for the S-SE than in the others.

Even so, it is difficult to draw definite conclusions from these results. To improve the quality of the estimates, we should try to diminish the impact of multicollinearity at least among the educational variables. We can do this through the use of principal components,[57] that is, by altering the variables by first transforming them into orthogonal components and then selecting the first three as our new educational variates (education 1, 2 and 3).[58]

The relative impact of personal versus structural variables can now be estimated. The results are presented in tables 2, 3 and 4, which refer to Brazil, the N-NE, and the S-SE, respectively. In

Table 3

Average Wages in the North and Northeast: Beta Coefficients and Levels of Significance of Personal and Structural Characteristics
(N = 200)

Variable	Equation (1)		Equation (2)		Equation (3)	
	Beta	%	Beta	%	Beta	%
% Age 20—29	.341[a]	15	.399[a]	30		
% Age 30—39	.134[a]	6	.219[a]	16		
% Age 40—49	.150[a]	7	.202[a]	15		
% Age 50—59	.097[b]	4	.080	6		
% Age 60 +	—.001	—	—.184[b]	—14		
Education 1	.167[a]	8	.212[a]	16		
Education 2	.151[a]	7	.292[a]	22		
Education 3	.114[a]	5	.176[a]	13		
% Male Workers	.157[a]	7	—.046	— 3		
% Production Workers	.339[a]	15			.295[a]	27
Log K/Worker	.009	—			.052	5
Log Workers/Firm	.116[b]	5			.120[b]	11
Log Gross Profits/Worker	.297[a]	13			.397[a]	37
Relative Minimum Wage	.142[a]	6			.216[a]	20
$\sum$	2.213	100		100	1.080	100
R^2	.784		.482		.628	

a Significant at the 1% level.
b Significant at the 5% level.

57 See B. T. McCallum, "Artificial Orthogonalization in Regression Analysis," *Review of Economics and Statistics* 52 (February 1970): 110-13.

58 See table 8.

Table 4

Average Wages in the South and Southeast: Beta Coefficients and Levels of Significance of Personal and Structural Characteristics

(N = 127)

Variable	Equation (1)		Equation (2)		Equation (3)	
	Beta	%	Beta	%	Beta	%
% Age 20—29	.043	2	.389[a]	21		
% Age 30—39	.106[b]	5	.200[b]	11		
% Age 40—49	.033	2	.336[a]	18		
% Age 50—59	.009	—	.192	11		
% Age 60 +	—.010	—	—.185	—10		
Education 1	.318[a]	15	.443[a]	24		
Education 2	.266[a]	12	.141[b]	8		
Education 3	.129[a]	6	.228[a]	13		
% Male Workers	.274[a]	13	.080	4		
% Production Workers	.038	2			—.151[b]	—15
Log *K*/Worker	.257[a]	12			.254[a]	26
Log Workers/Firm	.351[a]	17			.488[a]	49
Log Gross Profits/Worker	.044	2			.101	10
Relative Minimum Wage	.289[a]	13			.297[a]	30
Σ	2.160	100	1.824	100	.990	100
R^2	.863		.559		.656	

a Significant at the 1% level.

b Significant at the 5% level.

addition to the estimates for each group of variables, those for the complete model (which covers both personal and structural variables) are given in the first column of each table. Turning to table 2, for example, the second equation shows findings relative to the partial model that includes only personal characteristics, with the educational components standing out. Note that, in contrast to the coefficients in table 1, those for the educational components and the first three age variables are now significant. Taken together, the nine personal-characteristic variables account for 60% of the average-wage differential in Brazil, an impressive share considering the limitations of the available data.

Even more surprising is the fact that the structural variables alone (equation [3] in table 2) account for about 72% of the average-wage differentials, and all of them are highly significant.

Those that weigh most heavily are "gross profits per worker" and the "relative minimum wage", followed by the "percentage of production workers" [59] and "average size of the firm". "Capital per worker" has, by this criterion, the weakest influence on the variance of average wages.

In the complete model (equation [1]), the share of the variance of the dependent variable due to the independent variables is 84%, which is certainly high in terms of the cross-section analysis. Furthermore, only two of the 14 variables are nonsignificant (even at the 5% level), and 10 are significant at the 1% level. The quantitatively most important variables, i.e. those that account for at least 10% as revealed by their share in the sum of the beta coefficients, are the "relative minimum wage", the "percentage of production workers", "gross profits per worker", "age 20-29" and "education 1", in decreasing order.

Interestingly, the age and educational variables make a contribution that is equal to that of the sectoral characteristics ("average size of firm", "capital per worker" and "gross profits per worker"). Combined, they account for 67% of the total variance in the model.

This result, when viewed together with the highly satisfactory performance of the partial-equation estimates, emphasizes the importance of structural variables in analyzing the distribution of average wages. It reminds us, as well, of the importance of complementarities between the distribution of qualifications and the production structures of the various sectors. There is a striking difference between the coefficients of the general equation and those of the partial models, especially with regard to the educational variables. When measured alone, education unquestionably captures the contribution of other factors that have a systematic, and significant, effect on average-wage levels.

Before discussing tables 3 and 4, which report the regional estimates, attention should be drawn to an important aspect that these two have in common with table 2. In all three, the beta coefficients (the sum of which, albeit imperfectly, is an indicator of multicollinearity among the independent variables) decrease when we move from the first to the second and third equations. Although this is partly explained by the fact that equations (2) and (3) have fewer independent variables than equation (1), it is significant that the degree of multicollinearity is lower in equation (3), which includes the structural variables alone. This is especially true of the S-SE: the sum of the beta coefficients is .99, and the coefficient of determination (R) is .81.

[59] The positive coefficient of this variable for Brazil is due to its weight and behavior in the N-NE.

The findings for the N-NE and the S-SE regions both confirm some of the earlier results, and reveal important differences in the performance of certain variables. On the whole, the results for the N-NE are similar to those for the entire country. Although the correlation coefficients (R^2) are lower for the N-NE, the structural variates equation explains 63% of the average-wage differential, compared to 48% for the personal-characteristic equation. In terms of their weight in the sum of the beta coefficients, the order of the individual variables is almost the same in both the aggregate and regional estimates, with the exception of "capital per worker", which in the N-NE is quantitatively irrelevant and does not have a significant coefficient. The variable "percentage of production workers", besides being one of the most relevant in the N-NE, has a sign opposite from that expected (also true of the result for Brazil).[60] This does not mean, however, that a higher percentage of production workers is associated with a higher average wage. It reflects, instead, the nature of the firms in the region. A fair number are small establishments where owners and partners account for a relatively high share of total employment.[61] Since payment to this type of labor does not assume the form of wages, it follows that the higher the percentage of owners (i.e. the lower the percentage of production workers), the lower the average wage. The positive coefficient of the variable in the equation for Brazil as a whole is due to the weight of the N-NE in the total number of observations.[62] In the S-SE, the variable carries the expected sign in the structural equation, and is not significant in the complete model.

There are other, more substantial differences between the results for the N-NE and the S-SE. Among these, the following deserve special attention:

(i) In equation (2), the set of age variables as a whole is equally relevant in the N-NE and in the S-SE, but the tails of the distribution (20-29 and 60+) weigh more heavily in the former than in the latter. Curiously, however, the resemblance does not hold in the complete model (equation [1]). As expected, the share of the age variables diminishes in both regions, but appreciably more in the S-SE. Whereas these variables account for over 50% of the total variance in equation (2), they do not reach the 10%

60 This is because the average wage of production workers is lower than that of administrative personnel, and markedly below that of owners active in the firm.

61 See table 9, where their share of employment in a Northeastern state (Sergipe) is compared to their share in a Southern state (Paraná).

62 Note that all the observations (cells) have the same weight, e.g. "textiles" in São Paulo has the same weight as "machinery" in Piauí.

mark in equation (1). It is difficult to find an explanation for this behavior – it may be that once the structural variables are included, they capture the effects originally attributed to the age distribution, probably owing to promotion and seniority-ladder wage mechanisms.

(ii) The educational variables are more relevant in the S-SE than in the N-NE, and also more so in the complete model than in the partial equations.[63] This may be the consequence of regional disequilibria in the rate of return to schooling, but may also reflect the growing credentialist importance of education in the more developed regions. Allegedly objective hiring criteria may, in fact, be fostering job discrimination. That this is less noticeable in the N-NE than in the S-SE is only to be expected. Despite the fact that it also has an overly abundant supply of labor, the Southeast is the most dynamic region of the country, the undisputed leader in technological and organizational changes that are consistent with the growth of credentialism.

(iii) The estimates of the complete model leave little doubt as to the presence of sexual discrimination in the labor markets. Since there is no evidence that the quality of education (and/or work experience) is lower for women – actually it may be higher – this seems to imply that the discrimination is of the kind most usually found: unequal pay for relatively equal capacity to work.

(iv) The capital/labor ratio proved to be significantly related to wages in the S-SE, but not in the N-NE. This may be due to the fact that in the N-NE the density of capital not only is lower (on the average), but also has a much smaller variation. Moreover, in the S-SE the coefficient of this variable rises as we move from equation (3) to the complete equation (1). This implies that the density of capital is not merely a variable that complements skill and education, but is indeed one of the factors that determines the wage structure.

(v) The opposite is true of the gross profits per worker ratio, which is relevant in the N-NE but not in the S-SE. One explanation for this may lie in the large proportion of family-owned-and-operated firms in the N-NE. In such firms, there is rarely a formal wage-determination mechanism. The remuneration of hired labor is not independent of revenues – and it is likely to assume the form of payment for piecework. This usually means a lower level of income, reported in the census data as wages. In the S-SE, the lack of a strong relationship between gross profits and wages suggests that

[63] The same was found by V. Gibbon, "Taxas de Retorno dos Investimentos em Educação no Brasil: Uma Análise Desagregada," *Revista Brasileira de Economia* 29 (July-September 1975): 109-133.

workers do not share in the productivity gains of the enterprises.[64] No doubt, this reflects the very limited bargaining power of the unions, historically constrained by explicit anti-labor legislation.

Table 5

Average Wages in Brazil: Beta Coefficients and Levels of Significance of Personal and Structural Characteristics
(N = 327)

Variable	Equation (1)	
	Beta	%
% Age 20—29	.185[e]	8
% Age 30—39	.080[e]	3
% Age 40—49	.104[e]	4
% Age 50—59	.049	2
% Age 60 +	—.045	2
Education 1	.201[e]	9
Education 2	.133[e]	6
Education 3	.129[e]	6
% Male Workers	.163[e]	7
% Production Workers	.244[e]	10
Log *K*/Worker	.052[f]	2
Log Workers/Firm	.160[e]	7
Log Gross Profits/Worker	.213[e]	9
Region A[a]	.079[e]	3
Region B[b]	.169[e]	7
Region C[c]	.278[e]	7
Region D[d]	.448[e]	12
Σ	2.366	100
R^2	.845	

Note: The regional comparisons do not cover Maranhão, Piauí, Ceará, Rio Grande do Norte, Paraíba, Alagoas or Sergipe.

a Includes Amazonas, Pará, Mato Grosso and Goiás.

b Includes Pernambuco, Bahia and Espírito Santo.

c Includes Paraná, Santa Catarina and Rio Grande do Sul.

d Includes Minas Gerais, Rio de Janeiro, Guanabara and São Paulo.

e Significant at the 1% level.

f Significant at the 5% level.

[64] This is due to the fact that productivity and gross product per worker are strictly correlated by definition.

(vi) The average size of firms is important in both regions, but especially in the S-SE. In this region, there is the concentration of large and very large enterprises, and therefore of production processes where the greater division of labor allows for higher productivity but demands a more intricate work hierarchy that is related to higher wages. The number of relatively well-paid supervisory and quasi-managerial positions increases both within and outside the production sphere.[65]

(vii) While the importance of the relative minimum wage is nearly the same in the two regions, it is a particularly relevant variable in differentiating average wages within the S-SE region. This is not only (or even principally) the result of arbitrary institutional interferences, but reflects a whole set of historical factors that influence the average cost of labor. However, since it could be argued that the variable captures only the impact of the regional (state) minimum wage on average-wage levels, in another experiment we substituted it with a vector of dummies, one for each of the minimum-wage areas. The results presented in table 5 not only confirm the previous findings (see table 2), but can be interpreted as a rejection of the above mentioned hypothesis. In fact, all the coefficients are significant, and gradually rise as we move from the poorer to the wealthier states.

5 — Conclusions

An exercise in the decomposition of variance, such as the one presented in this paper, cannot be used as a forecasting model. Neither does it have an immediate use from the standpoint of economic policy. For it to serve this purpose, it would be necessary for the analysis to include a set of independent variables capable of being manipulated by the decisionmakers. In addition, government policy is traditionally concerned not with average-wage levels or with the distribution of wages, but with absolute minimum wages and the level of employment. Even so, our results have interesting implications with regard to the role of economic policy.

In part, our findings on average-wage differentials in Brazilian industry substantiate previous hypotheses. The simple regression models explain from 78 to 86% of the variation in the dependent variable. Nevertheless, even though the distribution of personal

[65] These findings may also be reflecting the lower cost of labor selection in small firms. See G. Stigler, "Information in the Labor Market," in *Readings in Labor Market Analysis*, ed. J. F. Burton, Jr. *et al.*, p. 241.

characteristics accounts for a large share of this total, our findings indicate that structural characteristics by sector and state comprise an equally significant, and quantitatively more important, factor.

When the personal-characteristic variables are treated independently, they are responsible for 72% of the 83.3% of the national wage differential explained by the complete model. The remaining 28% is attributable to structural characteristics (see table 2).[66] On inverting the order of the exercise, no less than 86% of the differential is associated with the structural variables, and only 14% with the distribution of individual traits.[67]

These results lead to another conclusion of the study — namely, the complementarity between the two groups of variables in the decomposition of the variance in average wages. When the coefficients of one group of variables are estimated independent of the other, they tend to capture the effects of the variables excluded from the model. Thus, if one group is considered alone, its contribution is overestimated. In a study of the American industrial structure, Kalachek and Raines ascribe this complementarity to the indirect contribution of personal characteristics.[68]

Our findings, and the conclusions of these writers, suggest that this indirect effect has a stronger impact on wages than does the direct effect of education. Since this indirect effect can only be explained by differences in industrial structure — which are beyond the control of the worker and therefore of the human-capital investment cycle — it appears that in order to improve the wage distribution greater emphasis will have to be placed on policy instruments other than those related to workers' qualifications.[69] This conclusion was,

[66] These results agree with those of Kalachek and Raines. Although the coefficient of determination for the sample of American workers is much lower (in 1966, $R^2 = .574$ in a regression with 58 variables and 3 595 observations), the contribution of the 27 personal-characteristic variables, when considered separately, is 73%. See Kalachek and Raines, tables 1 and 2.

[67] For the two regions, if personal characteristics are taken first, they account for 62% in the N-NE and 65% in the S-SE. When structural variables are considered first, the corresponding shares are 80% and 76%, respectively.

[68] "[Education] raises [the worker's] productivity over the range of activities he was hitherto capable of performing, but it also broadens the set of feasible work activities. Education qualifies workers for more skilled occupations and more skill-intensive industries. If occupations and industries are used as proxies for market differentials and are entered into the same wage regression, they will preclude the coefficient of education capturing the indirect role of education in routing workers into different pursuits." Kalachek and Raines, pp. 487-88.

[69] The same conclusion is reached in a recent study by the World Bank. See J. P. Jallade, *Basic Education and Income Inequality in Brazil: The Long Term View*, Staff Working Paper, n.º 268 (Washington, D. C.: World Bank, June 1977).

in fact, anticipated in section 3, where it was shown that wage differentials can be basically understood in terms of the organization of production rather than in terms of the acquisition of human capital. After all, low wages cannot be attributed to educational levels alone; there remains the question of how educational qualifications are used in the labor market. Once this question is asked, one is obliged to examine the influence of the industrial structure (or the demand for labor) on wages. Our results indicate that structural factors are not only present, but weigh more heavily than personal characteristics, in the formation of average wages.

Appendix 1: measurement of the variables

In assessing the results of this study, two points must be borne in mind with respect to the information used. One refers to the aggregation criteria, and the other to the comparability of the data taken from the more readily accessible secondary sources (the 1970 industrial and demographic censuses).

Throughout the study, the units of observation are industrial sectors and states, aggregated at the two-digit level. Both aggregates are imperfect – the first because it includes in the same category firms with disparate characteristics (with different types of organization, technological sophistication, degree of market control, etc.), and the second because it does not adequately distinguish specific labor markets within each state. Even so, the aggregates are presumably accurate enough to indicate the magnitude and direction of the major influences.

The comparability problem arises due to the lack of information on education/skill in the industrial census. This omission was overcome by using data from special tabulations of the demographic census and classifying the industrial labor force by sector, once again at the two-digit level.[70] However, the two census populations are not always the same. Whereas the industrial census covers only registered firms in sectors that account for considerable shares of the value of industrial output,[71] the demographic census includes self-employed and unpaid workers, in addition to the owners of small family firms. The differences in the composition of the two census populations, especially in those sectors that engage the largest number of workers, were minimized through utilization of the percentage

[70] In the special tabulations of the demographic census, "perfumes, soap and candles" is not distinguished from "pharmaceuticals", and the "miscellaneous" category, included in the industrial census, does not appear.

[71] Instituto Brasileiro de Geografia e Estatística, *Censo Industrial do Brasil, 1970*, Série Nacional, 4: xviii.

distribution of the labor force by age group and educational level. In addition, industries that employ a very small percentage of the total labor force in given states were excluded from the analyses of the respective states, as were sectors on which there is no information in the industrial census.

Table 6

Brazil: Employment by Industry According to the Industrial and Demographic Censuses, 1970

Industry	Industrial Census[a] (Total) (1)	Industrial Census (Production Workers) (2)	Demographic Census[b] (3)	(3)/(1) (%) (4)
Nonmetallic Minerals	236 506	183 968	252 874	106.9
Metallurgy	266 928	227 134	445 120	166.8
Machinery	180 431	152 607	159 777	88.6
Electrical and Communications Equipment	115 485	98 053	73 578	63.7
Transport Equipment	158 336	134 708	130 854	82.6
Wood	135 979	109 528	201 715	148.3
Furniture	105 322	80 933	219 479	208.4
Paper and Cardboard	66 994	57 019	61 666	92.1
Rubber	32 868	27 994	26 912	81.9
Leather Products	26 392	22 294	39 156	148.4
Chemicals and Petroleum Derivatives	104 367	81 658	135 052	129.4
Pharmaceuticals[c]	49 961	34 860	46 440	92.9
Plastics	42 566	36 598	37 466	88.0
Textiles	342 839	314 537	386 789	112.8
Clothing and Footwear	164 512	142 494	197 359	120.0
Food Products	372 401	289 157	435 091	116.8
Beverages	58 619	41 230	66 577	113.6
Tobacco	14 509	12 524	28 490	196.4
Printing and Publishing	97 087	72 019	126 881	130.7
Total	2 572 097	2 119 315	3 071 276	119.4

[a] Reference date: 31 December 1970.

[b] Reference date: 1 September 1970. The figures include self-employed workers.

[c] Includes perfumes, soap and candles.

Table 7

List of Industries[a] Excluded from the Analysis by State, 1970

State	Industry	State	Industry
Amazonas	Machinery Electrical Equipment Paper and Cardboard Pharmaceuticals Plastics Clothing and Footwear Tobacco	Piauí	Electrical Equipment Transport Equipment Paper and Cardboard Rubber Pharmaceuticals Plastics Tobacco
Pará	Paper and Cardboard Leather Products Pharmaceuticals Plastics Tobacco	Rio Grande do Norte	Electrical Equipment Paper and Cardboard Rubber Pharmaceuticals Tobacco
Mato Grosso	Electrical Equipment Paper and Cardboard Rubber Leather Products Pharmaceuticals Plastics Textiles Tobacco	Paraíba	Rubber Pharmaceuticals Plastics Tobacco
		Alagoas	Electrical Equipment Transport Equipment Paper and Cardboard Rubber Pharmaceuticals Plastics Tobacco
Goiás	Paper and Cardboard Rubber Chemicals Pharmaceuticals Plastics Tobacco	Sergipe	Electrical Equipment Paper and Cardboard Rubber Pharmaceuticals Plastics Tobacco
Maranhão	Electrical Equipment Paper and Cardboard Rubber Plastics Tobacco	Espírito Santo	Rubber Pharmaceuticals Plastics Tobacco
		Rio de Janeiro	Pharmaceuticals Tobacco
		Paraná	Pharmaceuticals Tobacco
		Santa Catarina	Pharmaceuticals Tobacco

[a]Perfumes, soap and candles are included in the pharmaceutical industry because there is no distinction between these sectors in the special tabulations of the demographic census. Also, there is no information on the Distrito Federal.

The dependent variable in the regression models is the average wage, measured as the total wages paid in 1970 divided by the number of workers employed on December 31st of the same year.[12]

72 Despite the fact that the numerator refers to the annual total and the denominator to a specific day, no systematic distortions were revealed by the test to check for a bias.

The data on personal characteristics (age, education and sex) are from the special tabulations of the demographic census. To facilitate empirical estimation and minimize the above mentioned comparability problem, each characteristic was treated as a percentage of the respective total, e.g. the percentage of workers with incomplete primary education or in the 20-29 age group.[73] It would have been possible to have used the opposite approach: first calculating the average educational level and the average age in each industry/state, and then forming the variable "average experience of the labor force".[74] On the one hand, this specification would have the advantage of allowing for computation of the average rate of return on investments in education by type of industrial employment,[75] as well as the explicit inclusion of experience (instead of age) in the breakdown of the wage differential. On the other hand, it would introduce new, and very serious, problems with respect to the measurement of education and experience. For those who have not completed the various levels, it would be necessary to impute some measure of the number of years of schooling, but there is no information on the average number of years studied according to state, industry and educational groups. Likewise, information needed to calculate average years of experience, i.e. the age of entry into the labor force, is not available.[76]

The data on structural characteristics (size of firm, capital density, percentage of production workers, and gross profits per worker) are from the 1970 industrial census. The reference date is December 31st. The percentage of production workers is a share of total employment, as reported directly in the census, and the

[73] As discussed in the text, due to multicollinearity among the education variables, it was decided to use a linear transformation, obtained through the principal components method.

[74] Conventionally, experience = age − years of schooling − age of entry into the education system, with work lapses being disregarded. See Mincer, p. 75, for example.

[75] According to the model, $lnYj = lnYo + rSj + u_j$, where Y_i stands for the average wage in sector j, S_j for the average level of education, and u_j for the residuals. The derivation of this model is presented in B. R. Chiswick, *Income Inequality: Regional Analyses within a Human Capital Framework* (New York: NBER, 1974), pp. 35-42.

[76] An alternative would be to work with the average age at which each educational level is completed. However, these figures are not available at the necessary level of disaggregation, and arbitrary values would have to be assigned for workers with no formal schooling. See J. R. Velloso, "Human Capital and Market Segmentation: An Analysis of the Distribution of Earnings in Brazil, 1970" (Ph.D. dissertation, Stanford University, 1975), pp. 51-52.

density of capital is simply the capital invested[77] divided by the total number of workers. The average size of the firms is gauged as the average number of workers per establishment.[78] Gross profits per worker were determined by dividing the production surplus[79] by the total number of employees.

The regional minimum wage is a percentage of the highest minimum wage in effect in 1970.

Appendix 2: complementary tables

Table 8

Education Variables: Major Components

Component[a]	Eingenvalue[b]			Variance (%)		
	Brasil	North-Northeast	South-Southeast	Brasil	North-Northeast	South-Southeast
1	3.476	3.252	3.808	57.9	54.2	63.5
2	.993	1.018	1.351	16.6	17.0	22.5
3	.586	.597	.484	9.8	9.9	8.1
4	.432	.549	.484	7.2	9.1	3.1
5	.365	.416	.118	6.1	6.9	2.0
6	.147	.168	.056	2.4	2.8	.9

[a] Only the first three components were included in the regression analysis. They correspond to the variables "Education 1", "Education 2" and "Education 3".

[b] These values were calculated using the six variables that measure the proportions of workers with incomplete or complete primary, secondary or university education.

77 Unpublished special tabulations based on the 1970 industrial census. It is recognized that this measure underestimates the amount of capital invested, but assumed that the degree of underestimation is the same for all industrial sectors.

78 Utilization of other size measures does not change the results. See, for instance, R. Ekerman, "The Wage Share and the Size Structure of Firms," Fundação Instituto de Pesquisas Econômicas, Universidade de São Paulo, 1974 (Mimeographed), table A2-1.

79 VIT less wages and miscellaneous expenditures.

Table 9

Pecentage Share of Proprietors[a] in Total Employment: Sergipe and Paraná, 1970

Industry	Sergipe	Paraná
Nonmetallic Minerals	19.7	7.4
Metallurgy	35.3	3.8
Machinery	44.4	3.4
Electrical Equipment	64.3	6.0
Transport Equipment	12.0	5.7
Wood	25.9	2.4
Furniture	44.5	6.4
Paper and Cardboard	—	.1
Rubber	—	2.3
Leather Products	45.4	4.3
Chemicals	13.3	.3
Pharmaceuticals	15.6	—
Plastics	—	—
Textiles	.4	.7
Clothing and Footwear	20.2	3.8
Food Products	9.8	11.9
Beverages	28.1	3.4
Tobacco	—	—
Printing and Publishing	4.4	2.5
Total	11.4	4.7

Source: Fundação Instituto Brasileiro de Geografia e Estatística (FIBGE), *Censo Industrial*, 1970.

[a] The figures cover only those who are actively engaged in the firms they own, plus the unpaid members of the families of these owners and partners.

Rapid growth, equity and the size of the public sector *

Rogério L. Furquim Werneck **

In a previous paper,[1] we argued that in the current situation of the Brazilian economy, it may be possible to maintain rapid growth in the future only in the presence of an expanding public sector — understood as an increasingly important public enterprise sector relative to the private sector. We also contended that this link could apparently be broken only if substantial public funds were donated — openly or disguisedly — to private industrial groups. Needless to say, this would have negative effects on the long-run distributions of income and wealth. Thus, there would be a trade-off among what are officially stated as the broad goals of economic policy in Brazil: (a) accelerated growth of aggregate product, (b) maintenance of a certain balance between the expansion of the public and private sectors, and (c) more equitable patterns in the distributions of income and wealth.

The argument was basically microeconomic since the trade-off ultimately arises from private firms having their growth limited

* This paper contains partial results of a project financed by a grant from the national program for economic research (Programa Nacional para Pesquisas Econômicas — PNPE). I wish to thank D. D. Carneiro, F. L. Lopes, P. S. Malan, E. J. Reis and D. F. F. Werneck for their valuable comments. An earlier version of the paper was presented at the 5th National Economics Meeting, held in Rio de Janeiro in December 1977.

** Department of Economics, Pontifícia Universidade Católica do Rio de Janeiro.

1 See R. L. Furquim Werneck, "Estatização, Crescimento Rápido e Eqüidade Distributiva," in *Brasil: Dilemas da Política Econômica;* ed. D. D. Carneiro (Rio de Janeiro: Editora Campus, 1977), pp. 187-52.

by financial constraints.[2] Consequently, those that belong to the more dynamic sectors are unable to expand at the pace required from these activities, given a relatively high target for the growth rate of aggregate product.

This paper is an effort to explore, from the macroeconomic standpoint, the "limits of the possible" as to rapid growth, equity and the size of the public sector. The analysis is somewhat impressionistic in some points — especially on the empirical side — reflecting the embryonic nature of the argument. Nevertheless, we believe that a more rigorous treatment of these points, though obviously necessary, will not significantly change the general outline of our conclusions.

During the last 10 years we have witnessed a process of acceleration of economic growth in Brazil. Whereas the average annual growth rate of GDP was about 7% from the end of World War II until the mid-sixties, it climbed to roughly 12% in the period 1968-1972, though it has slowed down recently.

In order to make the required adjustment in the rate of capital accumulation, the sources of savings in the economy had to be realigned. At the macroeconomic level this became feasible through (a) substantial increases in public savings as well as in retained profits of public enterprises, (b) more extensive utilization of forced savings instruments, (c) an increase in voluntary private savings, and, more recently, (d) a rise in the balance-of-payments deficit on current account.

The excessive foreign indebtedness of the country is already demanding that recourse to foreign savings (about 5% of GDP in recent years) be severely limited. This is bound to underline the relevance of the domestic savings constraint. To better understand the effect of this constraint on growth, we shall briefly survey some characteristics of the major components of domestic savings in Brazil.[3]

As to voluntary private savings, the pattern of income distribution can be expected to be a basic determinant since those in different income brackets have different propensities to save. Moreover, it is reasonable to assume that the average propensity to save should rise with the share of profits in aggregate income, and should thus depend on the functional distribution as well.

2 The idea that financial constraints limit the growth of firms, even when they have access to the stock market, is not recent. See M. Kalecki, "Principle of Increasing Risk," *Economica* (1937); *idem*, *Theory of Economic Dynamics* (London: George Allen & Unwin, 1954), chap. 8.

3 While recognizing that many factors condition the growth of an economy over time, we shall focus on only one: the aggregate savings constraint.

Although the intensity of these effects is open to question it is this kind of assumption that leads to the "concentrate-to-grow" strategy. According to the upholders of this argument, only a more concentrated distribution of income can assure the level of savings required for accelerated expansion of the economy. However, as we shall see below, this is not necessarily the case. To concentrate income may be one way to promote growth (at least as long as one looks only to savings without considering, for example, effective demand problems), but it is certainly not the only way.

It has also been frequently alleged that private savings could be considerably increased if the tax burden, judged to be excessively heavy, were lightened. Nevertheless, if we accept the idea that these savings come from the higher income brackets and from profits, it is easy to conclude that this kind of suggestion ultimately amounts to a policy of income concentration. In addition, the tax burden affects potential savers far less than one might imagine from looking simply at the share of taxes in aggregate income. The regressivity of the Brazilian tax system is well known. Not only are direct taxes of secondary importance, but even with respect to income taxes, readily available fiscal incentives favor savers, whether individuals or firms. Hence, it may be more effective to encourage private saving by raising taxes on superfluous consumption than by lowering income taxes.[4]

Let us now turn to government-controlled savings. One striking characteristic of the Brazilian economy in the last few decades has been the government's ability to generate collective and forced savings.[5] The levels of such savings clearly depend on the amount of resources controlled by the government over a given period, and may thus rise as tax revenues increase. Should this happen, one can

[4] See N. Kaldor, *An Expenditure Tax* (London: George Allen & Unwin, 1955)

[5] It is important to differentiate these two savings concepts. "Collective savings" funds are distinguished by the fact that they are essentially public, i.e. they belong to the state and therefore to society as a whole. The main component of collective savings is what is called government savings in national accounting, and comes basically from taxes. "Forced savings" funds, though also imposed upon society, are merely administered by government, and ultimately belong to individuals and private firms who have limited access to their respective shares (the time of service guarantee fund — FGTS, the program for social integration — PIS, the savings and insurance program for civil servants — PASEP, etc.). The retained profits of public enterprises require more careful classification. A fraction of such profits — equal to the government's share in the equity of these enterprises — should be considered collective savings, and the remainder forced savings. Note that what we are calling government-controlled savings is equal to the sum of forced and collective savings. Both come from the use of government's power to generate savings.

surmise, as a parallel to the "concentrate-to-grow" strategy, an "expand-the-public-sector-to-grow" policy. The level of savings required by a given rapid growth target could be generated by properly manipulating the degree of government participation in the economy and the level of government-controlled savings.

This introduces some questions related to the channeling of government-controlled savings. An increasingly common suggestion is that such savings should be "returned" to the private sector, from which they were allegedly drawn. In principle, it seems reasonable that government-controlled savings should also finance private investment. What does not seem reasonable is that most concrete proposals of this kind amount to open or veiled donation of a vast part of the forced and collective savings funds to certain privileged private groups. In practice, these proposals often carry clauses establishing ceilings on indexation, negative real interest rates, etc. Such suggestions, if widely adopted, would surely have the effect of concentrating the distributions of wealth and income in the long run.

At present, there is apparently a consensus that a move towards a more equitable distribution of wealth is one of the keys to a more equitable distribution of income. It is therefore important to avoid taking measures – in the heat of the "public-sector-size debate" – that would further concentrate the distribution of wealth and thus sacrifice the goal of social justice.

In addition, it should be noted that recourse to government power to generate savings, followed by the granting of a sizable part of these savings to the private sector, might have an even stronger effect on the concentration of income and wealth than would the generation of additional savings via the direct concentration of private income. In fact, it may not be politically feasible to adopt such measures. Firstly, because of the need to assure a non-negative real rate of return to the forced savings funds. Secondly, because there might be an "equity among the few" problem, to paraphrase Fellner. Opposition to donations may arise in the realm of the entrepreneurial class itself, simply due to the fact that there are not enough funds for all.

We will therefore accept the hypothesis that the collective savings funds will continue to be the property of society as a whole, while forced savings funds will remain the property of the shareholders, though managed and controlled by government. It follows that greater emphasis on government-controlled savings will lead to government appropriation of a larger share of national income each

year, as well as to public management or ownership of a greater proportion of the annual increase in total wealth.[6]

Let us now try to formalize certain aspects of the previous discussion. The model to be used is in the tradition of the Cambridge growth models of Kaldor and Pasinetti.[7] In these models, the balance between savings and investment is assured by changes in the average propensity to save. These changes occur in a particular way. Assuming an economy composed of two classes (workers and capitalists) with different propensities to save, the aggregate savings rate of the economy, which is a weighted average of the propensities to save of the two groups, varies as the functional income distribution changes. It is precisely this effect of changes in the functional income distribution on the savings rate of the economy that is the distinguishing feature of these models.

Our purpose is to adapt this approach to construct not a growth model, but a simple economic policy model capable of simultaneously indicating the size of the public sector and the pattern of income distribution that are consistent with a given economic growth target.

Let g represent the target for the annual growth rate of aggregate output and k the capital/output ratio, which is assumed to be constant. Designating total savings S, total investiment I and aggregate income Y, the savings rate required to assure the growth rate g will be given by

$$\frac{S}{Y} = \frac{I}{Y} = gk \tag{1}$$

Supposing a closed economy, total savings is equal to private savings S_p plus government-controlled savings S_g. That is,

$$S = S_p + S_g \tag{2}$$

Private savings can be divided into capitalists' savings S_c and workers' savings S_w, so that

$$S_p = S_c + S_w \tag{3}$$

[6] The structure of property ownership may develop quite differently in the long run, depending upon the emphasis given to collective savings vis-à-vis forced savings, notwithstanding the fact that both involve the same degree of government control over the increments in the economy's wealth.

[7] See, for example, N. Kaldor, "Alternative Theories of Distribution," *Review of Economic Studies* 23.

We also know that

$$Y = Y_p + Y_g \tag{4}$$

and that

$$Y_p = Y_c + Y_w \tag{5}$$

where Y_p is private income, Y_g government-controlled income, Y_c the income of the capitalist class, and Y_w that of the working class.

Let us now assume simple proportional savings functions:

$$S_c = s_c Y_c \tag{6}$$

$$S_w = s_w Y_w \tag{7}$$

$$S_g = s_g Y_g \tag{8}$$

which means that capitalists, workers and government save constant shares of their respective incomes.

Substituting (3) into (2) yields

$$S = S_c + S_w + S_g$$

and substituting (6), (7) and (8) into the above equation gives

$$S = s_c Y_c + s_w Y_w + s_g Y_g$$

Using (4) and (5), we may write

$$S = s_c (Y_p - Y_w) + s_w Y_w + s_g (Y - Y_p)$$

or rearranging and dividing both sides by Y,

$$\frac{S}{Y} = s_g - (s_g - s_c) \frac{Y_p}{Y} - (s_c - s_w) \frac{Y_w}{Y}$$

or, if we prefer,

$$\frac{S}{Y} = s_g - \left| s_g - s_c + (s_c - s_w) \frac{Y_w}{Y_p} \right| \frac{Y_p}{Y}$$

Then, using equation (1) and the last two above:

$$gk = s_c - (s_c - s_g)\frac{Y_p}{Y} - (s_c - s_w)\frac{Y_w}{Y} \tag{9}$$

$$gk = s_c - \left| s_c - s_g + (s_g - s_w)\frac{Y_w}{Y_p} \right| \frac{Y_p}{Y} \tag{10}$$

If we prefer, equations (9) and (10) may be rewritten as follows:[8]

$$\frac{Y_p}{Y} = \frac{s_c - gk}{s_c - s_g} - \frac{s_g - s_w}{s_c - s_g}\frac{Y_w}{Y} \tag{11}$$

$$\frac{Y_p}{Y} = \frac{s_c - gk}{s_c - s_g + (s_g - s_w)\frac{Y_w}{Y_p}} \tag{12}$$

Equations (9) to (12)[9] are different ways of determining which income distribution $\left(\text{measured by } \frac{Y_w}{Y} \text{ or } \frac{Y_w}{Y_p}\right)$ and public sector size $\left(\frac{Y_p}{Y}\right)$ combinations are compatible with a given target for the growth rate of aggregate output. Once this rate has been set, it becomes clear that the more concentrated the income distribution, the smaller the public sector need be. Likewise, a larger public sector allows for a more equitable distribution of income.[10]

From another point of view, the additional savings required for a higher growth rate of aggregate output can be generated in two ways: by enlarging the public sector or by concentrating the income distribution. On the one hand, for a public sector of a given size, the growth rate will rise if the income distribution be-

[8] We are assuming that $s_c > s_g > s_w$, i.e. that government has a higher propensity to save than capitalists, and that these have a greater propensity to save than workers. In addition, we are supposing that $s_c > gk$.

[9] These equations might be said to establish a "growth—private sector size—equity frontier".

[10] Note that we are using an indicator of public sector size which gauges the economic importance of government along a dimension quite different from the one used in the paper cited in fn. 1. In that paper we were interested in the size of the public enterprise sector. In the present paper our attention turns toward the share of national income appropriated and controlled by the state $\left(\frac{Y_g}{Y}\right)$. Still in relation to this point, we shall see further on that our definition of government-controlled income is somewhat particular.

comes more concentrated. On the other hand, for a given income distribution pattern, the larger the public sector, the higher the feasible growth rate.

Let us now introduce, in a very arbitrary way, plausible values for the parameters in equation (10) so as to obtain a rough idea of these trade-offs in the case of the Brazilian economy.

In recent years, the aggregate savings rate has been approximately 25%. Since this rate has been consistent with the 10% average annual growth rate of aggregate output observed over the last decade, it is reasonable to assume a capital/output ratio (k) equal to 2.5. As we are working with a closed economy hypothesis, we should subtract from aggregate savings the balance-of-payments deficit on current account. Of late, this deficit has been on the order of 5% of GDP or one-fifth of total savings. Hence, the aggregate rate of domestic savings should be around 20%, which is consistent with a growth rate of 8% per year, if the capital/output ratio is held at 2.5.

As to propensities to save, it is particularly difficult to obtain satisfactory estimates because the existing data do not fit our definitions of government-controlled savings and private savings.[11] One alternative, but a rather problematic one, would be to infer the sources of savings from information on the sources of fixed investment. According to unpublished statistics of the Fundação Getúlio Vargas (FGV), about 45% of fixed investments are of public origin and the remaining 55% of private origin.[12] Assuming that government-controlled savings are slightly larger than public investments and therefore responsible for financing part of private investment, we decided to work with the hypothesis that savings are equally divided between government-controlled and private savings.

To arrive at an estimate of the government's average propensity to save (s_e), we still need to know what share of national income is controlled by government. For the Brazilian economy, a

11 "Government savings", with estimates available in the national accounts data, constitute only part of what we are calling "government-controlled savings", as defined in fn. 5.

12 According to data used by Baer, Kerstenetzky and Villela, the share of public investment would account for 60.6% of total investment. However, this figure is apparently too high due to the fact that the share of public enterprise investment was estimated from a sample comprised of the largest firms in the country. See W. Baer, I. Kerstenetzky and A. Villela, "Modificações no Papel do Estado na Economia Brasileira," *Pesquisa e Planejamento Econômico* 4 (December 1973): 883-912.

reasonable value for Y_c in 30%.[13] Given a 20% domestic savings rate and the above hypothesis concerning the sources of savings, government-controlled savings would be equal to 10% of total income. Since 30% of this income is government-controlled, we may conclude that s_g would be equal to .10/.30, or approximately 34%.

Now we must divide private income into that of workers and that of capitalists in order to estimate the propensities to save of the two classes. The available data suggest that 55% of private income is appropriated by the workers.[14] Since private income is

[13] As explained in fn. 5, our definition of government-controlled savings comprises both collective and forced savings. Accordingly, our definition of government-controlled income, from which, after all, these savings come, is somewhat particular. In the case of the Brazilian economy, we would have to include not only tax revenues, as defined in the national accounts statistics (which include payroll taxes, FGTS, PIS-PASEP, etc.), without subtracting transfers and subsidies, but also retained profits of public enterprises. The idea is that, at least theoretically, government controls all these resources. Therefore, our definitions of "government-controlled income" and "private income" are quite different from the concepts of "public sector disposable income" and "private sector disposable income", with estimates available in the national accounts. As is well-known, these two aggregates are defined as:

public sector disposable income = indirect taxes − subsidies + direct taxes − transfers + other current revenues

private sector disposable income = GNP − public sector disposable income

However, the definition of what we are calling "government-controlled income" should be the following:

government-controlled income = public sector disposable income + transfers + subsidies + retained profits of public enterprises + public enterprise dividends paid to private shareholders

That is,

government-controlled income = indirect taxes + direct taxes + other current revenues (excluding public enterprise dividends paid to government) + public enterprise profits

Therefore, "private income" should be defined as follows:

private income = private sector disposable income − transfers − subsidies − public enterprise retained profits − public enterprise dividends paid to private shareholders

In fact, a more rigorous definition of "government-controlled income" would require some additional adjustments in "other current revenues". For example, government transfers to public enterprises (considered as belonging to the private sector in national accounting) enter as a negative balance in this account. As soon as both government and public enterprises are included in the "government-controlled sector", an adjustment is in order. For the breakdown of "other current revenues", see FGV/IBRE, *Contas Nacionais do Brasil: Conceitos e Metodologia* (Rio de Janeiro, 1972), pp. 21 and 66-67.

[14] According to the revised 1975 national accounts estimates recently published by FGV, the share of labor in urban income was 52.5%. It is interesting to note that if private savers had been divided not into capitalists and workers, but into the 10% wealthiest and the 90% poorest, the latter would still answer for a similar share: 55% according to the 1970 census data.

being estimated at 70% of total income, this means that 38.5% of the latter would be appropriated by the working class. Analogously, 45% of 70%, or 31.5% of total income, would be appropriated by the capitalist class.

Supposing that s_w equals 5% — so as not to hold it equal to zero — the workers' savings would account for just under 2% of total income. Since private savings are assumed to be equal to 10% of total income, those of capitalists should answer for the remaining 8%. As 31.5% of total income is appropriated by this class, s_c would be equal to approximately 25%.

We now have plausible values for all the parameters, and equation (10) can finally be used. The present situation of the Brazilian economy is roughly described by what is designated case A in the table below. The table includes other cases which simulate some policy experiments.

Case	g	$\frac{Y_w}{Y_p}$	$\frac{Y_p}{Y}$
A	.08	.55	.70
B	.09	.55	.58
C	.09	.37	.70
D	.08	.60	.67
E	.09	.60	.55

In case B, the long-run annual growth target is raised from 8 to 9% and the income distribution pattern is held constant. Under these circumstances, $\frac{Y_p}{Y}$ should fall from 70 to 58%, which means that the share of government-controlled income in total income should increase from 30 to 42%.

The long-run growth target is again raised from 8 to 9% in case C, but this time it is the size of the public sector that is held constant. This requires a brusk concentration of private income, with $\frac{Y_w}{Y_p}$ diminishing from 55 to 37%. Comparison of cases B and C suggests that it might be more reasonable to "increase the public sector to grow" than to adopt a "concentrate-to-grow" strategy.

In case D, the growth rate is held constant but income is redistributed, so that $\frac{Y_w}{Y}$ rises from 55 to 60%. Note that only a slight expansion of the public sector is needed to assure the 8% growth rate. $\frac{Y_p}{Y}$ drops from 70 to 67%.

Finally, in case E, the income redistribution policy is adopted at the same time that growth is accelerated, so that the public sector has to expand substantially, with $\frac{Y_g}{Y}$ falling to 55%.[15]

These effects might be somewhat magnified. This is partly due to the hypothesis that propensities to save are assumed to be constant. To the extent that the government's marginal propensity to save were substantially higher than its average propensity to save — which appears to be the case — smaller adjustments in the size of the public sector would be required. Likewise, changes in the income distribution pattern would be far less marked if the capitalists' marginal propensity to save were somewhat higher than their average propensity.

The model presented above is extremely simple and surely leaves aside various important constraints on the growth of the Brazilian economy. A particularly serious omission arises from the closed-economy hypothesis, which forced us to bypass the capacity-to-import. Nevertheless, we believe that, simple as it is, the model helps to clarify the nature of certain policy options involved in designing the long-run strategy for growth.

The basic idea is that there is more than one way to generate the domestic savings required by a given economic growth target. The choice amounts to further concentrating the distribution of private income, or allowing the government to appropriate and control a larger share of national income. As recourse to external savings becomes more difficult owing to the foreign debt of the country, and since further concentration of income is undesirable, continued rapid growth should depend on the latter option.

[15] It could be argued that at least one other case should be considered, i.e. the possibility of reducing government participation in the economy at the cost of lowering the long-term growth rate. Though logically feasible in the formal context of the model, the case does not seem to be relevant. After all, government performs many important roles in the economy besides generating savings, and most of these roles could not be assumed by the private sector. Consequently, there is little reason to suppose that the $\frac{Y_g}{Y}$ ratio observed at present could be significantly reduced.

Public employment and economic development *

Fernando Rezende **
Flávio P. Castelo Branco **

1 — Introduction

Employment has been a recurrent theme in the discussion of economic policy in developing countries. Although curtailment of unemployment and reduction of underemployment have often been declared prime targets in development plans, the evidence shows that these goals have seldom been achieved.

The weak performance of employment policy has to do with the alleged difficulty of coordinating it with growth policy. However, the conventional assumption that higher growth rates comprise the natural solution to the employment problem has been undermined by the experience of various countries, and increasingly more arguments have been presented in favor of adopting measures specifically aimed at raising employment.

Nonetheless, the measures proposed contain an element of contradiction. On the one hand, it has been suggested that the private sector be encouraged to use more labor via the concession of fiscal or financial incentives to establishments with higher labor/capital ratios. On the other hand, the government itself, in the name of administrative efficiency and lower current expenditures, frequently opts for capital-intensive technologies when providing public services.

* This paper was originally published in Portuguese in the collection *Aspectos da Participação do Governo na Economia*, Série Monográfica, n.º 26 (Rio de Janeiro: IPEA/INPES, 1976), pp. 35-76.

** Of the research institute of IPEA.

The severity of the problem stands out when it is taken into account that, even in the developing countries, the prospect of the private sector generating enough jobs is not very promising, given the rate of growth of the labor force. As Harrington asks of the American economy, if "the private sector becomes less and less able to absorb available manpower and, perhaps more importantly, less and less able to make a socially desirable allocation of that manpower, what then?"[1] The same might be asked about the developing countries, where most studies have pointed to the weak performance of the agricultural and industrial sectors with respect to creating job opportunities, especially for low-skill manpower.

It is in services, therefore, that the possibility of providing jobs for a growing percentage of the economically active population is greatest. Moreover, within the service sector, it is mainly through government activities that the basic needs of society (health, food, education, etc.) are met. Thus, the concern with the financial cost of government activities may conflict with the potential social benefits of expanded public services and of the employment of the additional personnel required to perform them.

If public employment is already a matter for consideration in the developed economies, it deserves even greater attention in the developing ones. In the latter, open unemployment is rarely important, but underemployment is usually widespread. At the same time the industrial sector is failing to generate a sufficient number of new jobs, the modernization of agricultural activities is releasing contingents of previously employed low-skill rural workers. These workers then migrate to the large urban centers, where they aggravate the already existing underemployment and social problems.

Up to now, the employment-policy debate has centered on the alternatives open to the government for encouraging the private sector to use less capital-intensive production techniques. In a recent work on Brazil,[2] for example, it is suggested that the government promote industrial employment by reducing social-security charges in direct proportion to the number of new jobs created per unit of investment. Even leaving aside the question of the efficacy of this type of incentive, it is worth noting that this alternative is more easily accepted than the utilization of manpower in government services, even where the direct and indirect costs of the latter option would be lower than those of the former.

1 M. Harrington, "Introduction," in *Public Service Employment*, ed. A. Gartner, R. Nixon and F. Riessman (New York: Praeger Publishers, 1973), p. xviii.

2 E. L. Bacha, M. Mata and R. L. Modenesi, *Encargos Trabalhistas e Absorção de Mão-de-Obra: Uma Interpretação do Problema e Seu Debate*, Coleção Relatórios de Pesquisa, n.º 12 (Rio de Janeiro: IPEA/INPES, 1972).

Similar observations can be made concerning public employment versus social security. As unemployment and underemployment rise, so do claims for social-security benefits, mainly in the form of financial transfers to the unemployed. Eventually, contributions have to be increased to cover the added outlays. Unless there is a belief that the problem is temporary, the government will gradually be pressured to assume more responsibility, either through the social-security system or unemployment insurance, for those who are openly or hiddenly unemployed. Once again, discussion of the available alternatives reveals some bias. The discussion has involved, for example, the adoption of a negative-income-tax scheme (despite the substantial funds that would have to be raised) as a means of ensuring the whole labor force a minimum income. The possibility of guaranteeing a minimum income by increasing public-employment opportunities has not received equal consideration, in spite of the fact it does not present the possible disincentive to work contained in the negative-income-tax proposals.

Another point worth mentioning is the potential impact of employment policy on internal migration. Rural-urban migration is one of the major causes of rising underemployment in the large cities of the countries experiencing rapid growth. Governments are therefore trying to control migratory flows so as to keep workers from concentrating in metropolitan areas that already have relatively saturated job markets. One way to keep some would-be migrants in their regions of origin might be to offer more public employment in small cities. Providing such employment would be less expensive than further burdening the social-security system and expanding the urban services in the large cities. To date, however, the emphasis on capital expenditures, together with that on administrative efficiency, has limited the effect of local investments on the creation of job opportunities.

The purpose of this study is to discuss the role of public employment as a major instrument of development policy. To this end, formerly unpublished information on the magnitude of government employment and its distribution by sector is examined. The emphasis is on identifying the factors that explain employment in urban public services, and especially on the institutional restrictions that preclude the more intensive use of labor in such services.

2 — The magnitude of public employment in Brazil

As usual in studies of the public sector, this analysis was subject to data restrictions. The only statistics published on public employment are those presented by the Instituto Brasileiro de Geografia e

Estatística (IBGE) in the population censuses and the household surveys (*Pesquisa Nacional por Amostragem de Domicílios* – PNAD). In both (except for the 1950 census, which will be commented upon separately), however, the data cover only the traditional government activities – i.e. administrative and legal services, public safety and national security – and therefore underestimate the employment generated by the government.

Another shortcoming of the data is that they do not allow for disaggregation of employment by occupational level or administrative sphere. Such breakdowns are especially important to understanding the growth of public employment in countries such as Brazil that are characterized by a high degree of administrative decentralization.

The 1950 population census is a partial exception to the preceding observations. This census contains a cross-referenced summary of public employment at the individual level by branch of activity and administrative sphere. Although "public employment" is not closely defined, it is assumed to be near inclusive, for the data cover all those employed by central governments and independent agencies.[3]

Turning to unpublished information, the register of a new social insurance fund for government workers (Programa de Formação do Patrimônio do Servidor Público – PASEP) is of great use in gathering statistics on the volume and structure of public employment. This source lists all those employed by federal, state and local governments and independent agencies, as well as those employed by public enterprises and mixed companies. Coverage is therefore complete, and the data comparable, with some minor exceptions, to those contained in the 1950 census. This comparability makes it possible to draw some conclusions as to the development of public employment over the period.

The PASEP register is not, however, of unquestionable reliability. On the one hand, there may be overestimates due to the multiple listing of some persons, or to the failure to remove from the files some who have retired or left the public sector.[4] On the other hand, since the fund is quite new, there are possible omis-

[3] Many of the entrepreneurial activities of the government are conducted by former independent agencies that already exercised such activities, though under official aegis. The transportation system furnishes an example. At the same time, it has been in the last 20 years that the government has decentralized its administrative apparatus and strengthened its role in business undertakings.

[4] Closed accounts have been recorded in the last two periods, but these do not represent the total multiple entries or cases of retirement or transfer to the private sector.

sions due to some agencies and firms not having registered their employees. It is hoped that the latter has occurred only at the level of the smaller *municípios* (roughly equivalent to counties, *municípios* serve as the unit of local government in this analysis), and is therefore of limited importance.

Another unpublished source that is yet to be fully utilized is the special set of tabulations based on the 1970 population census. These tabulations cover public employees, defined as "persons who receive pay from federal, state or local governments or from independent public agencies".[5] Only those engaged in government entrepreneurial activities are excluded. For the present study, additional tabulations were requested on the 116 *municípios* whose urban populations surpassed 50 thousand inhabitants in 1970. Since the latter do not embrace the whole country, they cannot be used to measure total public employment. Nonetheless, they are of great value in analyzing public employment at the urban level.

In this study, the PASEP data serve as the basis for gauging the current size and structure of public employment in Brazil. With due reservations, these data are compared to the 1950 census data in order to determine the changes that have taken place over the last 20 years. The more detailed analysis of public employment in urban areas, together with the attempt to explain the factors underlying the volume of employment in different types of urban services, is grounded on the special tabulations from the 1970 population census.

The available information indicates that the proportion of public employment is quite high in Brazil. In 1973, the government employed 3.4 million workers, or 9.5% of the labor force and just under one-fifth of the total engaged in nonagricultural activities (table 1).[6] These figures are comparable to — and even lower than — those for the United States. In a recent study, Harrison points out that the American government is now responsible for about one-fifth of the jobs in the country, and that the share of public employment is on the rise: one of every four new jobs is in the public sector, and in some major cities the public employment/total employment ratio reaches one to three.[7]

The comparisons between Brazil and the United States serve to show that public employment is playing an increasingly important role not only in the developing, but in the developed countries

[5] Fundação Instituto Brasileiro de Geografia e Estatística (FIBGE), *Censo Demográfico*, 1970, "Instruções de Coleta," p. 49.

[6] These data are based on the PASEP register.

[7] B. Harrison, *Public Employment and Urban Poverty* (Washington, D.C.: The Urban Institute, 1975), pp. 1-2.

as well. They also reveal that the magnitude of public employment is associated not with the degree of state control over the economy (which is generally high in developing nations), but with the broadening range of government activities, especially social and urban services at the state and local levels.

Analysis of the distribution of public employment in Brazil confirms the rise in the relative importance of the states and *municípios*. As table 2 shows, in the period 1950-1973, public employment grew an average 5.3% per year, but at a much faster rate for the states and *municípios* than for the federal government. Thus, whereas the federal government accounted for the greatest share in 1950 (approximately 50% of total public employment), it has now been supplanted by the states. The share generated by the *municípios* has not changed significantly. This is in part due to states taking over urban services (e.g. water and sewerage, public transport, public lighting) because of the reduced financial capacity of the *municípios*.[8]

Table 1

Brasil: Growth and Share of Public Employment, 1950 and 1973
(1 000 Workers and %)

		1950	1973
Public Employment	(A)	1 027	3 351[a]
Economically Active Population	(B)	17 117	39 371
Urban Wage Employment	(C)	5 333	17 238
A/B (%)		6.0	8.5
A/C (%)		19.3	19.4

Sources: Instituto Brasileiro de Geografia e Estatística (IBGE), *Censo Demográfico*, 1950; PASEP, general register; IBGE, *Pesquisa Nacional por Amostragem de Domicílios*, 1974.

[a] Excluding closed accounts.

[8] According to a survey made for IPEA by the Instituto Brasileiro de Administração Municipal (IBAM), in some *municípios* in the state of Rio de Janeiro, there are few services in which agencies maintained by the *municípios* themselves predominate (sewerage, sanitation, public recreation, and highways). Rather, it is the state government, or agencies within its sphere, that are mainly responsible for providing services. See IPEA/IBAM, "Participação do Setor Público na Economia como Empregador e Prestador de Serviços" (Rio de Janeiro, 1975) (Mimeographed), p. 45.

Table 2

Growth of Public Employment [a] by Level of Government, 1950-1973

	Total Public Employment (1 000 Workers)	Federal Government		States		*Municipios*	
		1 000 Workers	% of Total	1 000 Workers	% of Total	1 000 Workers	% of Total
1950	1 027[b]	506[c]	49.3	318	31.0	186	18.1
1973[d,e]	3 351	1 186	35.4	1 515	45.2	650	19.4
Average Annual Growth Rate (%)	5.3	3.8		7.0		5.6	

Sources: IBGE, *Censo Demográfico*, 1950; PASEP.

[a] Employment in all government agencies and entrepreneurial undertakings.

[b] The total exceeds the sum of the parts due to data problems.

[c] Including all independent agencies, since the majority must have been under federal administration.

[d] Including employment in public enterprises.

[e] Including closed accounts.

Table 3

Growth of Employment in Public Agencies by Level of Government, 1950-1973

	Total Public Employment (1 000 Workers)	Federal Government		States		*Municipios*	
		1 000 Workers	% of Total	1 000 Workers	% of Total	1 000 Workers	% of Total
1950	830	326	39.3	318	38.3	186	22.4
1973[a]	2 041	410	20.1	1 058	51.8	373	28.1
Average Annual Growth Rate (%)	4.0	1.0		5.4		5.0	

Sources: Same as table 2.

[a] Including closed accounts.

Table 4

Public Employment by Level of Government, 1973

	Number of Workers[a] (1 000)	% of Total
Government Agencies	2 041	59.9
Municipios	573	16.8
States	1 058	31.1
Federal Government	410	12.0
Independent Government Agencies and Foundations	660	19.4
Municipios	61	1.8
States	289	8.5
Federal Government	310	9.1
Public Enterprises and Mixed Companies	700	20.6
Municipios	26	.8
States	191	5.6
Federal Government	483	14.2
Nonclassified	3	.1
Total	3 404	100.0

Source: PASEP.
[a] Including closed accounts.

The available data also suggest that the expansion of public employment over the last 20 years is more the outcome of the creation of new agencies than of increased job opportunities in existing ones. The relatively accelerated growth of employment in independent agencies and entrepreneurial activities is partly explained by the gradual decentralization of the Brazilian public sector. This is especially true of the federal government (table 3). As a result of this trend, the joint share of independent agencies and entrepreneurial undertakings rose from less than 20% of total public employment in 1950 to about 40% in 1973.

The distribution of employment in 1973, by level of government and type of organization, is presented in table 4. With respect to public agencies, the federal government was less important (in absolute and relative terms) than the states and *municipios*. At the federal level, only half as many were engaged in government agencies as in public enterprises, mixed companies, independent govern-

ment agencies and foundations. In the states, the ratio was the opposite, with twice as many in public agencies as in the other kinds of activities. The relative weight of independent government agencies and entrepreneurial undertakings in total public employment is important from the standpoint of employment policy, given the nature of these activities and the concern with efficiency that has arisen with the process of administrative decentralization.

3 — Public employment in the major urban centers

3.1 — The magnitude of public employment by class of *município*

The preceding analysis shows that in Brazil the aggregate share of public employment in total urban employment is roughly 20%. The evidence on the North-American case indicates that this share is strictly related to city size, in that the relative importance of the government as employer grows with urbanization. To what point this hypothesis holds true in Brazil can be checked using the special tabulations referring to the 116 *municípios* having more than 50 thousand inhabitants in 1970. These data are presented in table 5 according to four subdivisions: (1) *municípios* in which state capitals are located, (2) others within metropolitan areas, (3) those having over 100 thousand inhabitants (large), and (4) those with 50-100 thousand inhabitants (small).

The relative importance of public employment was significantly greater in the 116 largest *municípios* (15%) than in the country

Table 5

Magnitude of Public Employment by Class of Município, *1970*
(1 000 Workers and %)

	Total	Capital	Metropolitan	Large	Small
Public Employment[a] (A)	1 722	1 195	168	275	81
Economically Active Population (B)	11 094	6 808	1 362	2 268	657
Wage Employment (C)	8 254	5 365	1 002	1 405	392
A/B (%)	15.5	17.5	12.3	12.2	12.8
A/C (%)	20.9	22.3	15.4	19.6	21 4

Source: Special tabulations based on the 1970 population census. The data refer to the 116 largest *municípios* in the country.

a Including workers in public utilities, despite the fact that these persons are not classified as public employees in the 1970 census

as a whole (8.5% according to table 1). And within this group of 116 *municípios*, the share of such employment was highest in those in which the state capitals, which include the most important cities in the nation, are located (17.5% compared to an average of 12% in the other groups of *municípios*). It was also in these *municípios* that the share of public employment in total urban wage employment was highest. As to the small *municípios*, the high rates of government employment may be explained by the lack of other urban job opportunities.

3.2 — Public employment by type of activity

The data in table 6 confirm the earlier observation as to the role of social programs in the expansion of public employment. Entrepreneurial activities withstanding, social programs — mainly education — account for the greatest share of government employment in the principal cities. The traditional government activities (public administration and defense and safety) represent less than half of the respective total.

Once again, disaggregation by class of *município* substantiates the relative importance of those in which the state capitals are located (table 7). On the average, there are twice as many public employees in the *municípios* in which state capitals are located as in the others. While this is mainly explained by the concentration of state and federal administrative activities, services also account

Table 6

Public Employment [a] by Type of Activity, 1970

	Number of Workers (1 000)	% of Total
Social Activities	577	34.7
Education	339	20.4
Medical Care	101	6.1
Public Administration	399	24.0
Defense and Safety	376	22.6
Others	310	18.7

Source: Same as table 5.

[a] Excluding workers in entrepreneurial activities other than public utilities.

Table 7

Average Public Employment per Thousand Inhabitants by Type of Activity and Class of Municipio, *1970*

	Total		Capital		Metropolitan		Large		Small	
	Employment per 1 000 Inhabitants	% of Total	Employment per 1 000 Inhabitants	% of Total	Employment per 1 000 Inhabitants	% of Total	Employment per 1 000 Inhabitants	% of Total	Employment per 1 000 Inhabitants	% of Total
Social Activities	15.4	30.0	20.8	30.4	8.8	24.1	12.3	32.3	13.0	32.8
Education	10.1	20.0	13.5	19.7	6.1	16.7	10.2	26.8	11.1	28.0
Medical Care	3.0	5.8	4.0	5.8	1.8	4.9	1.3	3.4	1.1	2.8
Public Administration	21.2	41.2	29.9	43.7	15.5	42.5	12.8	33.6	12.5	31.6
Urban Services[a]	6.9	13.4	8.0	11.7	6.4	17.5	4.8	12.6	5.2	13.1
Urban Improvements	1.6	3.1	2.4	3.5	1.7	4.7	1.7	4.5	1.7	4.3
Public Utilities	3.1	6.0	3.2	4.7	3.4	9.3	2.1	5.5	2.6	6.6
Others (residual)	7.9	15.4	9.7	14.2	5.8	15.9	8.2	21.5	8.9	22.5
Total	51.4	100.0	68.4	100.0	36.5	100.0	38.1	100.0	39.6	100.0

Source: Same as table 5.

[a] Including nongovernment public-utilities workers.

for significant interclass differences. In the case of education, for example, the average number of employees per thousand inhabitants is twice as high in the capital *municípios* as in the other *municípios* within metropolitan areas. Nevertheless, relative to total government employment, the importance of educational employment increases as city size decreases. This reflects not only recent government priorities, but also the relevance of manpower in the production of this service.

In social activities, there is a marked imbalance in the distribution of employment between educational and health programs. In overall terms, medical services are at an extreme disadvantage, and the high number of persons engaged in such services in the state capitals may be due to the limited role of the local governments in providing medical services.

Urban services may be divided into two categories: urban improvements and public utilities. With regard to urban improvements, employment levels are highest in the *municípios* in which capital cities are located, the per capita figures for the other groups of *municípios* being similar to one another. In the provision of public utilities (which include water, sewerage, gas, and electric energy), economies of scale enter, and large firms often extend their services to several *municípios*.[9] In public utilities, the numbers employed per thousand inhabitants are highest in the capital and metropolitan *municípios,* precisely because of their size and greater availability of funds.

Thus, the data on public employment — both total and by type of activity — reveal striking differences in the number of persons employed in the four groups of *municípios.* These differences are partly explained by city size. However, other factors, some of them institutional, have an equal or greater impact on urban service employment in Brazil. The hypotheses underlying this statement, together with the related empirical test, are presented in the next section.

3.3 — The determinants of public employment

The factors that determine urban public employment may be initially grouped as follows: (a) those that affect the demand for the types of services performed by the government, (b) those that ex-

[9] Considering the negative externalities for the rest of the urban structure, there may even be diseconomies of scale as of a certain size (i.e. extent of services and dimension of facilities). However, to prove or disprove the existence of economies of scale in services is difficult, and impossible on the basis of employment data alone.

plain the capital/labor ratios for these same services, and (c) those that reflect the institutional restrictions on government employment.

In any attempt to incorporate all these factors into a statistical analysis aimed at explaining the differences in the number of persons employed by the government in each class of *município,* data problems inevitably arise. These problems, which refer to the selection and availability of information on persons actually employed (total and in certain service activities), make it impossible to distinguish the effect on employment (1) of an increase in the demand for public services, (2) of substitution between factors of production, and (3) of limitations imposed by the availability and degree of freedom allowed in the use of government funds.

Even so, regression analyses that use economic and demographic factors as independent variables can produce interesting quantitative results. These results, in turn, may serve as policy guides, since they measure the relationships between the volume of public employment and the size and composition of the population, the level of income, and the structure of public revenues. A study of this kind was recently made in the United States,[10] and is here adapted to the Brazilian case.

The model to be tested is essentially of the type $L_i = f\ (P, Y, R, D)$, where L_i stands for the volume of employment in activity i; P for the population (total or urban); Y for the income level; R for revenues (able to be subdivided into own revenues and transfers); and D are dummy variables for factors that are not perfectly quantifiable, such as regional differences in attitudes toward government employment.

It is assumed that the demand for urban public services, and therefore the level of employment in such services, increases with the size and/or density of the population and with the income level. It is further assumed that the availability and source of funds affect the degree to which the production of (and employment in) urban public services rises with demand. Theoretically, the correlation between employment and revenue should be positive, since more funds should mean less financial constraints on the expansion of services. In practice, however, the composition of revenues can produce significant alterations. For example, employment is likely to be restricted to the extent that it rests on transfers, for these are partially earmarked for capital-investment outlays.

10 R. Lehne, "Revenue Sources and Local Government Employment," *Public Finance Quarterly* 3 (October 1975): 400-410.

Due to the large number of equations estimated, the results of the econometric analysis are presented in a single table at the end of this section. Generally speaking, the results are only fair, since the explanatory power of the equation and of the coefficients estimated is rarely very significant.[11] The lack of significance of most of the coefficients of the revenue variable suggests that financial constraints are of little importance in determining the level of urban public employment (with the exception of the large *municípios*, which are better able to overcome the constraints on employment posed by earmarked transfers). It should be pointed out, however, that the financial variables used refer only to funds available at the local level, and therefore reveal nothing about urban employment directly generated by the federal or state governments.

As for the demographic variables, the main conclusion to be drawn from the analysis concerns the probable existence of substantial economies of scale in the provision of public services at the local level.[12] With minor exceptions, the coefficients relative to urban population and population density are less than zero when significant. This suggests that the rate of growth of public employment per capita does not keep pace with urbanization.[13] The major exceptions are administrative employment, which is positively correlated with the degree of urbanization of the large *municípios* and those within metropolitan areas, and local government employment, which is positively correlated with population density in the state capitals and other large cities. The latter is explained by the high yield of property taxes in urban centers.

At first glance, using the state value added tax (ICM) as a proxy for the *município* income, the results do not confirm the assumption as to the relationship between employment and income levels. Except for public utilities, the signs are positive for the small *municípios* and negative for the large and metropolitan *municípios*.

[11] Such limited explanatory power is acceptable in cross-section analyses, since the data are subject to wider variations than in time-series analyses. The coefficients of determination are low because certain factors had to be omitted due to the impossibility of quantifying them; the factors that determine the location of state and federal agencies in given *municípios* provide an example.

[12] Economies of scale were also reported in a recent study on *município* expenditures. See A. B. Araújo *et al.*, *Transferências de Impostos aos Estados e Municípios*, Coleção Relatórios de Pesquisa, n.º 16 (Rio de Janeiro: IPEA/INPES, 1973), p. 202.

[13] The negative coefficients for the population variables do not go against the original hypothesis that the public employment/population ratio should be positive. Since the estimates are based on per capita values, the negative coefficients merely indicate that public employment does not grow as quickly as urban population.

Table 8

Determinants of Public Employment

	Sample	Constant	*Municipio* Funds per 1 000 Inhabitants	Federal Transfers per 1 000 Inhabitants	ICM Collections per 1 000 Inhabitants	Population Density	Urban Population	Regional Dummy	R^2
Municipio Employment per 1 000 Inhabitants	A	—.975	.307[a] (.102)	—	—	.109[c] (.063)	—.250[c] (.132)	—	.517[a]
	B[d]	—	—	—	—	—	—	—	—
	C	.631	.253[a] (.072)	—	—	.114[c] (.061)	—	—	.415[a]
	D	2.622	—	—	.147[b] (.065)	—.136[b] (.056)	—	—	.376[b]
Total Public Employment per 1 000 Inhabitants	A	2.337	.128[b] (.054)	—	—	.070[a] (.023)	—.245[a] (.050)	—	.633[a]
	B	.236	—	—	—.160[b] (.066)	—	.270[b] (.124)	—	.536[b]
	C	—.509	—.100[b] (.073)	—	—.135[c] (.077)	—	—	—.130[b] (.060)	.524[a]
	D	2.835	—	—	—	—.158[b] (.060)	—	—.223[b] (.098)	.372[b]
Public Social Employment per 1 000 Inhabitants	A	2.066	—	—	—	—.098[a] (.025)	—.271[a] (.054)	—.099[c] (.051)	.707[a]
	B[d]	—	—	—	—	—	—	—	—
	C	.381	—	—	—	—	—	—.138[a] (.042)	.301[a]

	D	.162	—	—	—	—	—	—.174[a] (.045)	.459[a]
Public Administrative Employment per 1 000 Inhabitants	A	2.283	.226[b] (.083)	—	—	—	—.219[b] (.078)	—	.467[b]
	B	—.093	—	.406[c] (.214)	—.206[a] (.070)	—	.340[b] (.130)	—	.599[a]
	C	—1.052	.325[a] (.073)	.355[b] (.171)	—	—	.483[b] (.199)	—	.571[a]
	D	—	—	—	.185[c] (.103)	—.328[c] (.087)	—	—	.464[a]
Public Urban-Service Employment per 1 000 Inhabitants	A	.460	—	—	.196[c] (.097)	—.193[b] (.074)	.123[a] (.036)		.482[a]
	B	2.704	—	—	.209[a] (.076)	—.112[c] (.058)	—	—.183[c] (.104)	.467[b]
	C	.040	.288[a] (.042)	—	—.109[b] (.051)	—	—	—	.558[a]
	D[d]	—	—	—	—	—	—	—	—

Notes: The values between parentheses are the standard deviations.

The samples are: A — *municípios* in which state capitals are located (23 observations).

B — *municípios* within metropolitan areas (20 observations).

C — *municípios* with 100 thousand inhabitants and over (46 observations).

D — *municípios* with 50—100 thousand inhabitants (27 observations).

An empty cell means that the variable was not used due to its not reaching the level of significance and/or its not adding to the explanatory power of the regression.

a Significant at the 1% level.

b Significant at the 5% level.

c Significant at the 10% level.

d Neither the coefficients of determination nor the regressions significant at the 10% level.

This apparent contradiction suggests that the ICM is not an adequate income-level proxy. In the case of the small *municípios,* the ICM reflects the effect of budget constraints on total *município* employment and on total administrative employment (which probably consists, in this class, of *município* jobs alone). But in the larger *municípios,* the negative coefficients refer to administrative employment, which loses relative importance with increases in income and diversification of government services, and to total public employment per capita, which decreases with the expansion of job opportunities resulting from development.

Since few of the coefficients for federal transfers are significant, it was not possible to prove that growth of public employment is restricted in *municípios* that depend to a greater degree on such funds, which are partially earmarked for capital expeditures. However, in the large and metropolitan *municípios,* the coefficient that shows the relationship between federal transfers and administrative employment is positive when significant (table 8). Since only part of the funds are earmarked and the volume of transfers is directly proportional to the size of the population, indications are that some of these funds are used to create additional administrative jobs.

The regional dummy is most significant for employment in social activities in small *municípios.* This is precisely the case in which the explanatory power of the other variables is lowest, and the random location of state and federal agencies probably most important.

These results suggest that for a given increase in urban population, population density, or availability of resources at the local level, public employment per capita will grow more rapidly in administrative activities than in social activities or urban services. Moreover, they confirm that capital is being substituted for labor in the performance of government services. This trend is the outcome of two factors: (1) minimization of operating expenditures in the name of efficiency, and (2) the gradual decentralization of government activities. While this trend cannot be quantified on the basis of our estimates, it is reflected in the coefficients of elasticity between employment and available funds being low despite the fact that manpower is the principal factor of production in most public services. It is worth noting that the limited impact of increased funds on government employment in urban services and social activities, especially in the small *municípios,* conflicts with the aim of raising public employment as an instrument of economic policy.

4 — Public employment and economic activity

The usual approach to government employment emphasizes emergency programs specially suited to times of depression. The main objective is to minimize the impact of crises on employment and income by expanding public-works programs so as to provide temporary job opportunities, and thereby either stabilize the economy or hasten recovery.[14] Being short-term, these solutions are not suited to overcoming the employment problems in the developing countries, where employment problems require more *ex ante* intervention in order to eliminate the structural causes of labor-market disequilibria. In these countries, the issue should be approached from the angle of underemployment rather than open unemployment. Whereas the latter affects only 4% of the urban labor force in Brazil, the former ranges from 5.3 to 17.3% in the major urban centers.[15]

Thus, public-employment policy should perhaps stress the creation of permanent job opportunities instead of the adoption of emergency public-works programs. From the point of view of short-term stabilization, such a change would mitigate depression *ex ante*, since public employment would be relatively immune to reductions in the overall demand for the goods and services produced by the private sector. From the standpoint of long-range goals, new possibilities might arise for maintaining high rates of growth while reducing underemployment, distributing income more equitably, and improving the standard of living.

The belief that higher growth rates will automatically solve the employment problem has gradually been abandoned. Not only does industrialization fail to generate a sufficient number of new jobs, but the distribution of those that are created is not in line with the skill of the labor force. As a result, unemployment (underemployment) spreads among the unskilled, and relative wages rise for skilled manpower.

The most common recommendations for resolving the growth-versus-employment conflict are of two types. First, there are those who advocate additional training so as to prepare the labor force to meet the demand for skilled manpower, and thus improve the distribution of wages. Second are the proponents of labor-intensive

[14] This is the approach used during major droughts in the Brazilian Northeast.

[15] See H. C. Tolosa, "Subutilização e Mobilidade da Mão-de-Obra Urbana," in *Política de Desenvolvimento Urbano: Aspectos Metropolitanos e Locais*, ed. J. Barat, Série Monográfica, n.º 22 (Rio de Janeiro: IPEA/INPES, 1976), pp. 23-85.

techniques, who argue for measures to raise the employment/capital ratio. Since the pros and cons of these suggestions have already been discussed at some length,[16] we shall leave them aside and consider the alternative of direct public employment.

An important question refers to the source of funds required to finance increased public employment. If the government were to provide minimum-wage jobs for all the critically underemployed in urban areas (currently estimated at approximately two million persons),[17] it would have to disburse about Cr$ 20 billion per year, or roughly 10% of the federal budget revenue estimated for 1977. It would clearly not be advisable for the government to furnish direct employment for all two million workers. Nor would it be necessary, since increased public employment would have a multiplier effect, expanding the internal market and raising the demand for the goods and services produced by the private sector.

Public employment, especially in the provision of urban services in small and medium-size centers, could serve as a key instrument for stemming or controlling migratory flows. In Brazil, migration apparently occurs in stages, with moves being first from rural areas to small centers, then to medium-size cities, the regional capital, and finally the principal metropolitan areas. However, this process is not accompanied by government efforts to help migrants adapt to the cultural standards and labor markets of the large urban centers. In this respect, public employment in the small and medium-size cities could retard the flow of migrants to the large cities, and/or prepare them — socially and professionally — for life in their final destination centers. This would also benefit the metropolitan areas, for if migrants were readily assimilated, the cost of improving environmental conditions would be lower.

Job opportunities in small cities could be created in two ways. First, *municipios* might be encouraged to use more labor-intensive techniques in services such as public cleaning and road maintenance, which are often mechanized to an unnecessary degree. Second, if less emphasis were placed on capital expenditures, social services (e. g. education and medical care) could absorb more labor, which is, after all, the major factor of production, and the one responsible for quality, in these areas. Better social services, by raising the standard of living, would keep some potential migrants from shifting to the larger cities.

[16] With regard to proposals for the further utilization of labor-intensive techniques, see Bacha *et al.*

[17] Tolosa, p. 44. "Critical underemployment" is defined as less than 14 hours of work per week.

The creation of public-employment opportunities in small and medium-size centers would also improve the regional distribution of income by transferring funds from the more developed to the less developed areas. Moreover, the indirect effects could be as important as the direct effects. Whereas transfers to individuals (e.g. social security benefits or negative income tax) only influence the local economy via increased family expenditures, additional public employment could lead to the expansion and improvement of urban services throughout the region.

Likewise, public employment could exert a positive impact on the distribution of personal income. Since there is a shortage of jobs for unskilled labor, the government might make more use of untrained manpower in urban services. Even now, among the government activities, it is urban services that employ the greatest number of unskilled workers, especially in the smaller *municipios*.

In the urban centers, the focus should be on raising the standard of living. Given the size of these centers, it may not be feasible to utilize labor-intensive techniques in the provision of some urban services. However, social services, which are often claimed to be deteriorating, could be improved through the expansion of public employment. In fact, there is a growing pressure to set up programs not only in the fields of education and medical care, but also in basic sanitation, environmental control, social assistance, cultural activities and recreation. Significantly, the quality of these services depends on manpower.

Inflation and the balance of payments in a dependent economy

Celso L. Martone *

1 — Introduction

This paper analyzes the monetary adjustment of a dependent economy in a model which brings together the quantity theory of inflation and the monetary approach to the balance of payments. An open dependent economy is understood to be one that is relatively small, and that responds passively to events in the international markets for goods and capital.

Since the article attempts to explain how a dependent economy makes the monetary transition from a closed to an open system, it deals with short-term situations covered neither by the traditional theory of inflation nor by the monetary theory of the balance of payments.[1]

The two-sector model employed distinguishes an international (tradable) sector and a domestic (nontraded) sector, with the two being differentiated mainly in terms of the price-formation mechanism. This division of the market for goods is essential to the dynamics of inflation in an open economy, because it implies the coexistence and interaction of two price levels within the system. A credit market is also explicitly introduced into the model in a form close to the loanable-funds theory, and particularly to the

* Economic research institute, Universidade de São Paulo.

1 To cite only the best-known works, one should mention M. Friedman, *The Optimum Quantity of Money and Other Essays* (Chicago: Aldine, 1969), chap. 1, on the traditional theory of inflation, and H. G. Johnson, *Further Essays in Monetary Economics* (Cambridge: Harvard University Press, 1973), chap. 9, on the monetary theory of the balance of payments.

"cumulative process"[2] recently incorporated into the so-called Keynes-Wicksell growth models. This allows that the monetary approach be used to examine adjustments in the goods and capital accounts separately.[3]

The model is applied to basic problems of monetary adjustment such as modifications in domestic credit policy, foreign-exchange devaluations, and changes in international interest and inflation rates. Each of these problems is examined in the context of different institutional frameworks. Thus, in section 3, the economy is characterized by a fixed exchange rate and a total lack of capital mobility. In section 4, the exchange rate is still fixed, but capital is mobile. Finally, section 5 allows for an automatically indexed exchange rate, based on the purchasing-power parity theory.

2 — The analytical framework

The economy to be analyzed is small relative to the rest of the world, and its price structure is flexible enough to guarantee continuous full employment. There is no government in the usual sense of the term, only a monetary authority that sets the exchange rate and conducts monetary transactions with the rest of the banking system and directly with the public.

In any discussion of price dynamics in an open economy, the first precaution, and one of the aims of this study, is to define the general price level. For a closed economy, one-sector models in which the price level equals the "price of national income" may be useful; but for an open economy, such models lose much of their predictive power due to the existence of one price-formation mechanism for international goods and another for domestic goods. The prices of tradable goods are determined on the international market, in which the individual economy presumably plays a negligible role. In contrast, the prices of domestic goods are set by local supply and demand. Since we are concerned with a dependent economy, tradables will be exchanged on a "fixprice" market in which adjustments are basically made through changes in the quantities traded or purchased at a given price level, while nontradables will be exchanged on a "flexprice" market in which

2 K. Wicksell, *Interest and Prices*, Reprints of Economics Classics (New York: A. M. Kelley, 1965).

3 See, for example, J. Stein, "Monetary Growth Theory in Perspective," *American Economic Review* 60 (March 1970): 85-106; S. Fischer, "Keynes-Wicksell and Neoclassical Models of Money and Growth," *American Economic Review* 62 (December 1972): 880-90; D. Laidler, *Essays on Money and Inflation* (Chicago: University of Chicago Press, 1975), chap. 5.

price adjustments are responsive to a disequilibrium between supply and demand.[4] Explicit consideration of these two types of goods improves the forecasting ability of the model, as well as facilitating analysis of the dynamics of inflation in an open economy.

In defining nontradables, consideration must be given not only to natural barriers and transportation costs, but also to the tariff structure. Since a prohibitive tariff can transform importable or exportable goods into nontradables, the latter are those products and services that do not enter international trade flows should there be a slight change in their relative prices. This class of goods, so defined, has received due attention from only a few writers.[5]

Disregarding tariffs and subsidies, and letting e stand for the exchange rate and $\overline{P}_T$ for the price index of tradables, free trade will guarantee that

$$P_T = e\,\overline{P}_T \tag{1}$$

In the type of economy under discussion, P_T is determined independent of the domestic market. Hence, assuming a fixed exchange rate, the rate of variation in the prices of tradables will be given exogenously by

$$\pi_T = \pi_T \tag{1a}$$

The general price level is defined as a weighted average of the price indices for the two classes of goods:

$$P = aP_N + (1 - a)\,P_T \tag{2}$$

where a is the share of the nontradable good N in total expenditure of the economy.

[4] The terminology is that used by J. R. Hicks, *Capital and Growth* (London: Oxford University Press, 1965), chaps. 6 and 7; *idem*, *The Crisis in Keynesian Economics* (London: Oxford University Press, 1974).

[5] Among whom might be cited the following, in chronological order: I. F. Pierce, "The Problem of the Balance of Payments," *International Economic Review* 2 (January 1961): 1-28; R. A. Mundell, *Monetary Theory, Inflation, Interest and Growth in the World Economy* (Pacific Palisades: Goodyear, 1971), chap. 9; R. Dornbush, "Devaluation, Money and Nontraded Goods," *American Economic Review* 63 (December 1973): 871-80; A. Krueger, "The Role of Home Goods and Money in Exchange Rate Adjustments," in *International Trade and Finance*, ed. W. Sellekaerts (New York: International Arts and Sciences Press, 1974), pp. 141-61; M. Bruno, "The Two-Sector Open Economy and Real Exchange Rate," *American Economic Review* 66 (September 1976): 566-77; M. Connolly and D. Taylor, "Adjustment to Devaluation with Money and Nontraded Goods," *Journal of International Economics* 6 (1976): 289-98.

Supposing that $P_N = P_T = 1$ initially, the domestic rate of inflation will be

$$\pi = a\pi_N + (1 - a)\,\pi_T \tag{2a}$$

Excess real expenditure Z is the sum of excess expenditure in each market:

$$Z = \frac{P_N}{P} Z_N + \frac{P_T}{P} Z_T \tag{3}$$

Given a fixed exchange rate, the balance of trade will be equal to the variation in one of the components of the monetary base $\dot{B}_1^F$ and to the excess supply of tradables:

$$\dot{B}_1^F = - P_T Z_T \tag{4}$$

Likewise, the net inflow of capital K will be equal to the variation in another component of the monetary base, $\dot{B}_2^F$:

$$\dot{B}_2^F = e\,k \tag{5}$$

The variation in the economy's stock of international reserves (by definition equal to the balance-of-payments surplus) will therefore be the sum of (4) and (5):

$$\dot{B}^F = \dot{B}_1^F + \dot{B}_2^F = - P_T Z_T + e\,K \tag{6}$$

The monetary base is comprised of two elements: a domestic component B^D, the result of monetary and credit operations between the authorities and the domestic sectors (banks and the public); and an external component B^F, defined in (6) as the outcome of operations between the authorities and international agents, whether on current account or in the form of capital transactions:

$$\dot{B} = \dot{B}^D + \dot{B}^F \tag{7}$$

Whereas the domestic component is under the control of the monetary authorities, who can use it to attend to their policy goals, the two external components are determined in the markets for credit and for goods and services.

The money supply is the product of the banking multiplier s, assumed constant for simplicity, and the monetary base:

$$\dot{M}^S = s\,\dot{B} \tag{8}$$

There are three markets relevant to the analysis: (1) the market for goods and services, divided into the international sector and the domestic sector, which roughly determine the balance of trade and the domestic rate of inflation, (2) the credit market, which basically determines the nominal rate of interest or the capital balance (depending on the hypothesis concerning international capital mobility), and (3) the money market, which will be explicitly eliminated. The credit market is defined, in the tradition of Brunner and Meltzer,[6] as the market for the securities that comprise the assets of the banking system. An important aspect of the model is the absence of explicit portfolio behavior of the private sector. This intentional omission is due to the interest in examining income, expenditure, money and credit flows in the spirit of the loanable-funds theory, in which the reader will recognize the dynamics of the Wicksellian "cumulative process".

Excess global absorption of goods and services is a function of the expected real rate of interest:

$$Z = Z(r - E),\ Z' < 0 \tag{9}$$

where r is the nominal rate of interest, E the expected rate of inflation, and

$$Z(n) = 0 \tag{10}$$

Rate n, assumed to be constant, can be identified with the Wicksellian "natural interest rate", that is, the real interest rate which, given the expectations concerning the future of the economy, will balance savings and investment flows and consequently the market for goods. Under these conditions, and assuming that r is the prevailing rate in the credit market, to the extent that the expected real rate $r-E$ is lower (higher) than the natural rate n, there will be excess (insufficient) absorption in the market for goods. Through analogy with the natural rate, the expected real rate is associated with the Wicksellian "monetary interest rate", or the expected real rate that brings the credit market to short-run equilibrium.

The real absorption of money, or the real flow demand for money by the private sector, is formed of two elements: a replacement

[6] K. Brunner and A. H. Meltzer, "Monetary and Fiscal Policy in Open, Interdependent Economies with Fixed Exchange Rates," paper presented at the 3d Paris-Dauphine Conference, March 1974; *idem*, "An Aggregative Theory for a Closed Economy," paper presented at the Conference on Monetary Economics, Brown University, November 1974.

demand for expected capital losses due to inflation; and a demand for increases in the real stock, owing to disequilibria between the desired and the actual real stock:

$$\frac{\dot{M}^d}{P} = E\, m + \Theta\ (m^d - m) \tag{11}$$

$$m^d = f(r),\ f' < 0 \tag{12}$$

with m being the real money stock and θ a coefficient of stock adjustment.

The excess demand for each class of goods is a function of excess absorption Z and of the relative price between the goods $W = P_N/P_T$:

$$Z_N = Z_N\,(Z, W),\ Z_{N1} > 0,\ Z_{N2} < 0 \tag{13}$$

$$Z_T = Z_T\ (Z, W),\ Z_{T1} > 0,\ Z_{T2} > 0 \tag{14}$$

The expected rate of inflation is a function of the economy's past inflation:

$$E_N = E_N[\pi(t)] \tag{15}$$

Any change in the international rate of inflation π_T is immediately perceived by the private sector, which does not foresee additional changes. Hence, aggregate expectations concerning inflation are expressed by

$$E = aE_N + (1 - a)\,\pi_T \tag{16}$$

On the domestic market, prices adjust to excess demand and expected inflation:

$$\pi_N = E_N + \lambda\ Z_N\ (Z, W) \tag{17}$$

where λ is a coefficient of ajustment. On the foreign market (balance of trade), adjustments are ultimately of quantities in relation to

an exogenous rate of change in prices, as shown in equations (1a) and (4).[7]

This difference between the two price equations is a basic characteristic of the model, and makes it possible to render the theory of inflation for a closed economy compatible with the monetary theory of the balance of payments. How this becomes feasible may be summarized as follows: in the short run, an open economy behaves in line with the traditional macroeconomic models for closed economies; but in the long run, it acts according to the monetary pattern traced by writers such as Johnson.[8] The most convincing way to link the two models and to allow for the transition from a closed to an open economy is therefore to divide goods into two classes, each with its own dynamics of adjustment.

3 — Monetary adjustment in the absence of capital mobility

In order to study the process of adjustment in the absence of capital mobility, we shall further hypothesize that, in terms of velocity, the credit market adjusts infinitely faster than the market for goods. Thus, the credit market will always be in equilibrium at nominal interest rate r. Underlying this hypothesis is the assumption that the transaction costs and time required to effect changes in the interest rate are infinitely lower than those needed to bring about changes in prices. The loanable-funds equation, which is the condition for credit-market equilibrium, will therefore be constantly met:

$$Z(r - E) + Em + \Theta(m^d - m) = s\frac{\dot{B}^D}{P} + (s - 1)\frac{\dot{B}^F}{P} \qquad (18)$$

bearing in mind that at this point $eK = \dot{B}_g^F = 0$.

The above equation may be rewritten as follows:

$$\frac{P_N}{P} Z_N(Z, W) + Em + \Theta(m^d - m) = s\frac{\dot{B}}{P} \qquad (18a)$$

7 Alternatively, it could be hypothesized that expectations are also formed in relation to the international rate of inflation, as suggested by D. Laidler and R. Cross, "Inflation, Excess Demand and Expectations in Fixed Exchange Rate Open Economies," in *Inflation in the World Economy*, ed. M. Parkin and G. Zis (Manchester: Manchester University Press, 1976), pp. 221-55. However, maintaining different expectations for the domestic and world markets requires that exchange speculation be introduced into the model. Besides complicating the model itself, in the long run this makes it unfeasible to sustain the hypothesis of a fixed exchange rate.

8 Johnson, chap. 9.

As (18a) indicates, the hypothesis regarding the velocities of adjustment implies that the excess demand for nontradables will equal the flow excess supply of money in real terms.

In this system, equilibrium can be defined in two ways. First, steady-state equilibrium can be determined for a given rate of change in the price of tradables and a given rate of world inflation, as seen from the standpoint of the country itself. This state will be reached when the economy satisfies three requirements: (1) the market for nontradables is in equilibrium, (2) the actual and expected rates of change in prices equal the international inflation rate, and (3) the rate of expansion of domestic credit (on the part of the monetary authorities), μ^D, is also the same as the international rate of inflation. That is,

$$\begin{aligned} & Z_N(Z, W) = 0 \\ & \pi_N = E = \pi = \pi_T \\ & \mu^D = \pi^T \end{aligned} \tag{19}$$

Under these conditions, the nominal interest rate will be $r = n + \pi$, and the real stock of reserves will be constant over time. In other words, the balance-of-payments surplus will be exactly the amount required to hold constant the real value of the international reserves, depreciated by world inflation. The community will therefore be paying an inflation tax, equal to the balance-of-payments surplus, in the form of a transfer to the world reserve centers.

Second, a sort of quasi-equilibrium can be distinguished. This allows for analysis of situations which, while not indefinitely sustainable, may occur at certain points. Such quasi-equilibrium arises when

$$\begin{aligned} & Z_N(Z, W) = 0 \\ & \pi_N = E = \pi = \pi_T \end{aligned} \tag{20}$$

with μ^D unrestricted. In this article, the emphasis will be on these situations. Clearly, any rate of expansion of domestic credit other than π_T will eventually imply a change in the real stock of reserves. Thus, at quasi-equilibrium, the model restates a basic proposition of the monetary approach, i.e. the existence of perfect substitution between domestic and international assets within the monetary base.

Some types of disequilibria can also be understood through the conditions for quasi-equilibrium described in (20), definition equations (1a), (2a) and (16), and adjustment equations (4), (15) and (17). Thus, equation 20) can be used as the reference point:

for defining two kinds of disequilibria which can occur in the model. In the first case, the domestic market may be in equilibrium ($Z_N = 0$) with the flow excess supply of money therefore nil, at the same time the trade balance is registering a deficit equal to excess global absorption:

$$\frac{P^T}{P} Z_T = Z = -\frac{\dot{B}_1^F}{P} \tag{21}$$

This condition has often been used by monetarists to make forecasts concerning the balance of payments. The market for nontradables is assumed to remain in equilibrium, and all monetary adjustments to occur via international reserve flows.

The second type of disequilibrium arises when the trade balance is in equilibrium ($Z_T = 0$), but there is an excess demand for nontradables equal to excess global absorption and to the flow disequilibrium in the money market:

$$\frac{P_N}{P} Z_N = Z = s\frac{\dot{B}^D}{P} - \frac{\dot{M}^d}{P} \tag{22}$$

This is the case which comes nearest to a closed economy. It may occur in the short run if the trade balance adjusts at very low velocity, so that the credit-market disequilibrium shows up in the market for nontradables.

These two forms of disequilibria can only be observed in a two-sector model which includes, in addition to the absorption effect typical of a closed economy, a substitution effect that operates through the relative price W. Although the existence of a substitution effect has been proposed by several economists,[9] its connection to inflation dynamics in an open economy has not been explored.[10]

We shall now consider the three types of exogenous shocks that comprise the usual monetary problems of the balance of payments: a change in the rate of expansion of domestic credit, a devaluation, and a change in the international rate of inflation.

[9] Such as Pierce, Mundell (chap. 9), and Dornbush.

[10] I considered this problem in an earlier article, but was unable to cover adjustment under persistent inflation due to the static framework adopted. See C. L. Martone, "O Enfoque Monetário à Teoria do Balanço de Pagamentos: Algumas Implicações," *Revista de Estudos Econômicos* 4 (May 1974): 193-220. The work that comes closest to the model presented here (with neither an explicit credit market nor a formal adjustment model, however), is that by M. I. Lejer, "Money, Prices and the Balance of Payments: The Case of Mexico" (Ph.D. dissertation, University of Chicago, 1975).

Let us assume, to start with, that the quasi-equilibrium conditions described in (18) and (20) are met at a rate of expansion of domestic credit exactly sufficient to maintain the trade balance at equilibrium ($Z_T = 0$),

$$\mu_0^D = \frac{\pi_T}{1 - C} \tag{23}$$

with $c = B^F/B$ being the share of reserves in the monetary base. As of time t_0, the monetary authorities increase the rate of growth of domestic credit to μ_1^D and hold it at the new level by, for example, increasing rediscounts or direct credit to the private sector. It is known that, at the new quasi-equilibrium the economy will eventually attain, the only lasting effect of the change in monetary policy will be a current-account deficit that will allow the private sector to eliminate the excess money being created by the authorities.

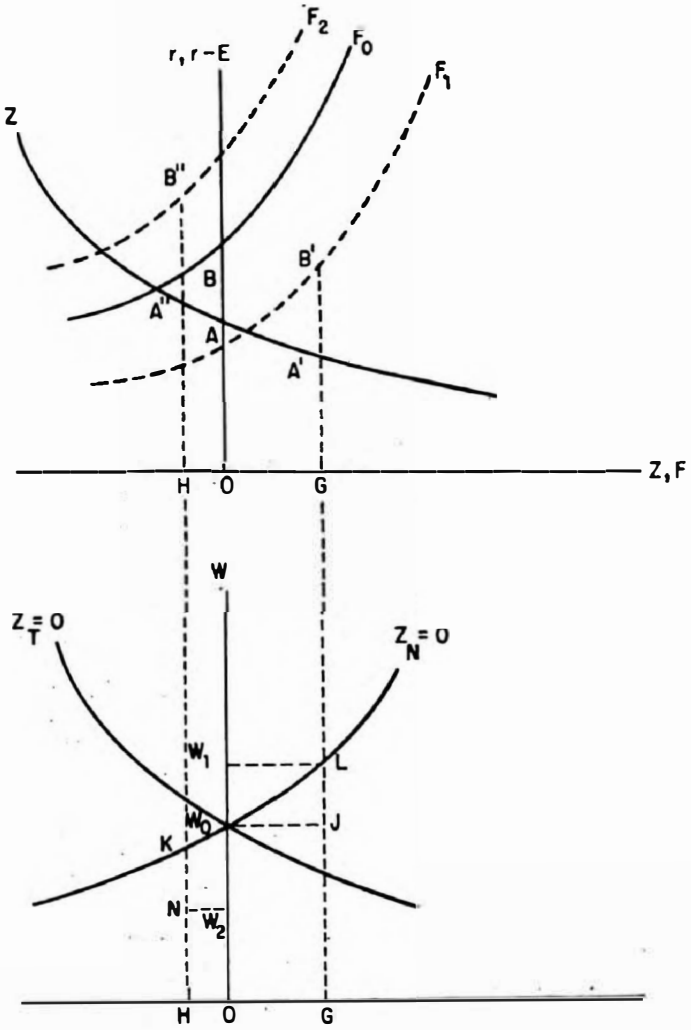
However, during the adjustment process, the economy initially behaves as if it were closed, and only gradually reaches the situation predicted by the monetary approach. At first, the acceleration in the expansion of credit pushes down the nominal interest rate and directly encourages spending. The subsequent excess absorption is manifested in the excess demand for both nontradables and tradables, so that $Z_N > 0$ and $Z_T > 0$. On the domestic market, this excess demand results, via equation (17), in prices rising at a faster rate. On the foreign market, it leads to larger purchases at the same rate of change in prices (equations [1a] and [4]), and hence creates a trade deficit equal to the excess demand.

At this point, a double Wicksellian process begins on the credit market. On the one hand, there will be an internal drain due to the real value of the credit flow decreasing as the average rate of inflation increases. This effect will be passing, however, since in the long run the domestic rate of inflation will return to the level prevailing in the rest of the world. On the other hand, an external drain will be occasioned by the excess demand for tradables leading to a loss of foreign assets and slowing down the growth of the money supply. The fact that both effects are stabilizing means that the cumulative process set in motion by monetary policy will eventually run down and the system will come to a new quasi-equilibrium.

In the meantime, the relative price W will rise, generating a substitution effect against nontradables, so that the absorption effect will worsen the trade balance and hasten the loss of reserves. The market for nontradables will eventually reach equilibrium at a higher relative price W but at the original rate of inflation. In this way, the economy will attain the situation described by the monetary

Figure 1

ADJUSTMENT OF AN OPEN ECONOMY TO A MONETARY SHOCK



approach and represented by equation (21), in which the excess absorption of goods (continually renewed by the expansion of domestic credit) is equal to the trade deficit.

The expansionary process brought about by the overly rapid growth of domestic credit cannot be sustained in the long run, because the level of reserves sets the upper limit on the economy's deficit. Once reserves have been depleted, monetary policy will have to be revised. It is in this sense that the situation is one of quasi-equilibrium.

Figure 1 illustrates this case. In the upper part, curve Z stands for excess absorption and falls with the expected real rate of interest. Curve F, defined by the flow excess supply of money in real terms, slopes upward due to the fact that an increase in the nominal rate of interest will lower the real demand for money and accelerate the flow excess supply:

$$F = s\frac{\dot{D}}{P} - Em - \Theta(m^d - m) \tag{24}$$

In the lower part of the figure, curves $Z_N = 0$ and $Z_T = 0$ are depicted on plane (Z, W). The initial quasi-equilibrium occurs at nominal interest rate OB, inflation $\pi_T = AB$, and relative price W_0. As soon as the monetary authorities increase the rate of expansion of domestic credit, the economy will start to move toward excess absorption OG (at the outset, $AB = A'B'$). This excess absorption will be reflected in an excess demand for both nontradables and tradables, thereby characterizing point J as the effect of monetary expansion.

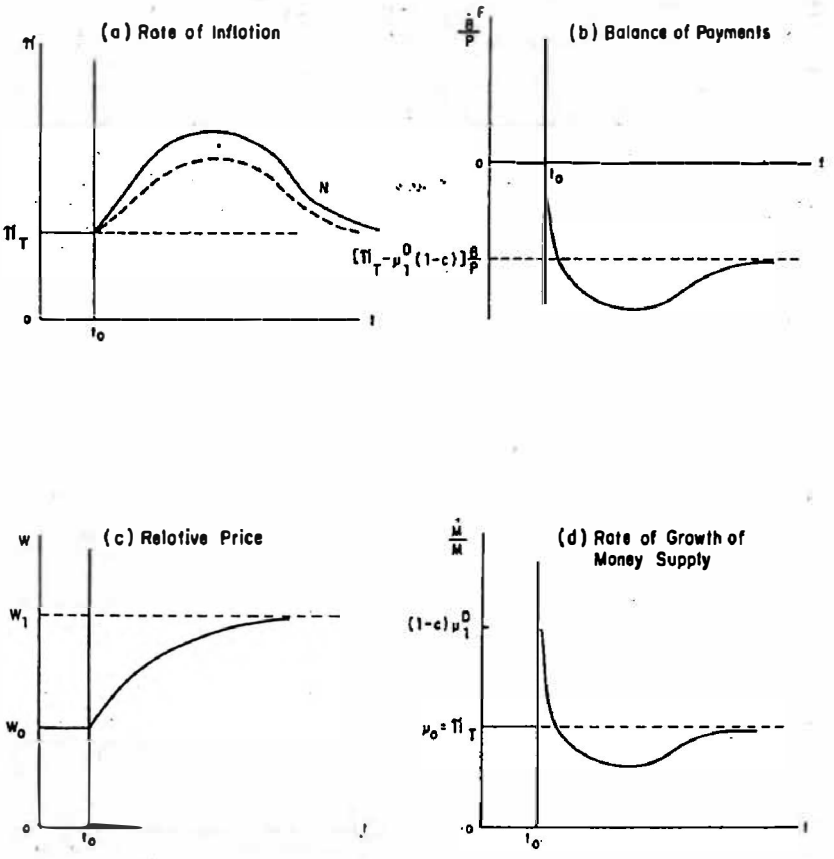
At point J, the inflationary pressure on nontradables will be such that their prices will change at a faster rate and W will rise above its original level. During some time, π_N will be over its long-range level, thus causing the forementioned internal drain, which will cease when the domestic market achieves its new equilibrium at point L. The situation at point L corresponds to the proposition presented in equation (17), which shows excess absorption equaling the trade balance.[11]

Some possible time paths for the main variables are traced in figure 2. In section (a), the domestic rate of inflation temporarily differs from the world rate in order to allow the relative price W

[11] The path from J to L will be nonlinear, since the expected rate of inflation will rise for some time, thereby lowering the real interest rate and encouraging additional spending. It should be observed that expected inflation has a transitional effect, influencing the credit market only while the domestic market is in disequilibrium. Although this effect was not included in the figure, it may be responsible for oscillations in the path of the variables. In our case, expectations were relegated to a secondary role owing to the fixed-exchange-rate hypothesis and to the fact that long-range domestic inflation is inextricably linked to the rate of change in the prices of tradables.

Figure 2

TIME PATHS OF THE MAJOR VARIABLES AS OF AN UPWARD TURN IN THE RATE OF MONETARY EXPANSION



to increase and the domestic market to reach equilibrium. This relative price approaches the new equilibrium (section [c] of the figure) asymptotically.

The behavior of the balance of payments, as well as that of the money supply (sections [b] and [c]), is the direct result of the absorption effect and the substitution effect:

$$\dot{Z}_T = Y_{T.1}\,\dot{Z} + Z_{T.2}\,\dot{W} \tag{25}$$

Since Z immediately shifts to OG, the balance of payments becomes negative and the deficit will continue to increase as long as Z and W are rising. This explains the possibility of the real trade balance surpassing its new equilibrium value, and then stabilizing at $OG = [\mu_1^D \ (1-c) \ -\pi_T] \ B/P$. The growth rate of the money supply will behave in accord with the balance of payments, since the expansion of domestic credit is constant throughout the process.

Let us now suppose, starting from the same position of quasi-equilibrium, that the monetary authorities devalue the exchange rate from e_0 to e_1. Whereas an increase in the rate of expansion of domestic credit was seen to affect the rates of change of the variables, a devaluation initially alters the levels of the variables and only later influences their rates of change.

Devaluation immediately exerts two effects. First, the relative price W drops in direct proportion to the devaluation, thus generating a substitution effect in favor of nontradables on the part of consumers, and in favor of tradables on the part of producers. In other words, there is a shift in spending, to the advantage of nontradables, for a given level of global absorption. Second, the devaluation instantly raises the general price level by $(1 - a)$ $(e_1/e_0 - 1)$ percent, thereby reducing the real money supply and causing monetary disequilibrium. It therefore operates as a tax or represents capital losses to money-holders, and induces the private sector to temporarily acquire more money at the cost of a lower absorption of goods. This is possible due to an increase in the nominal interest rate, which leads to an excess supply of goods $(Z < 0)$ and a negative absorption effect in the market for goods.

Both the absorption effect and the substitution effect cause trade surpluses and accumulation of reserves. On the domestic market, however, the two effects oppose each other. Thus, in order to achieve stability, it is necessary to create an excess demand for nontradables that will bring the relative price W back to its original level. In this context, devaluation tends to restore equilibrium. On the one hand, the surplus increases the growth rate of the monetary base, thereby allowing the private sector to re-establish its real money supply. On the other hand, the higher rate of inflation on domestic goods contributes to returning relative prices to their initial level. It is for this reason that devaluation is said to have a transitional effect, and to be incapable of permanently influencing relative prices or real magnitudes. In other words, as various economists have suggested, unless domestic credit policy is altered at the same time, devaluation merely serves to increase the reserves of the monetary authorities.

The adjustment that follows a devaluation is shown in figure 1. At the outset, $r_0 = OB$, $\pi = BA$, and $W = W_0$. The devaluation shifts curve F_0 to F_1 (generating an excess supply of goods OH), and lowers the relative price from W_0 to W_1 (creating an excess demand for domestic goods and an excess supply of foreign goods, represented by point N). Over time, the surplus will cause F_1 to return to F_0, and the excess demand for domestic goods will lead to an increase in π_N, once again raising the relative price W_1 to W_0. This second quasi-equilibrium will coincide with the first, that is, with the situation prior to the devaluation.

The effects of sterilization of the monetary impact of the balance of payments can also be analyzed with the aid of figure 1. Thus, let us assume that the authorities adopt the following procedure:

$$\mu^D = \frac{1}{1-c}\left[\pi^T - c\frac{\dot{B}^F}{B^F}\right] \tag{26}$$

so that, regardless of what happens to the stock of reserves, the base grows at a constant rate π_T. In this case, curve F_1 will not return to F_0, for the attempt of the public to acquire money via the trade balance will be frustrated by the countervailing transactions of the authorities. "Equilibrium" will be maintained at point K, where there will be insufficient aggregate demand ($Z < 0$), equilibrium in the market for nontradables ($Z_N = 0$), and a balance-of payments surplus ($Z_T < 0$) exactly equal to the difference between income and absorption. The supply of reserves will therefore grow indefinitely.

Finally, let us consider the case in which the rate of inflation of the price of tradable goods rises from π_T^0 to π_T^1 and remains at the new level. Assuming that the increase in the international rate of inflation is perfectly perceived, the domestic rate will rise by $d\pi = dE = (1 - a)\,d\pi_T$. This increase will have an uncertain impact on the money market, for even if the replacement demand rises, the demand for the real stock will fall. Thus, the flow of private money purchases may either increase or diminish.

As time passes, however, two things should occur. In the first place, the relative price W should start to drop, causing a substitution effect in favor of nontradables which will, in turn, generate excess demand in this market. In the second place, the new path of prices P will impose larger capital losses on moneyholders and lead to a higher absorption of money at the cost of a lower absorption of goods. Eventually, the excess demand for nontradables will cause W to turn upwards, thus indicating how the rate of inflation on this class of goods adjusts to the new world rate of inflation. At

the end of the process, the economy will be characterized by insufficient absorption OH, with $A''B'' = AB + d\pi_T$, and by a point such as K (in the lower part of figure 1) standing for equilibrium in the market for nontradables at a relative price lower than the original, and a balance-of-payments surplus equal to OH.

The model therefore recognizes two means of transferring inflation from country to country. One is a relative price mechanism that generates excess demand for domestic goods and pushes inflation upwards; the other, a monetary mechanism, imposes capital losses on those who demand money, while creating a surplus in the balance of payments and an increase in the quantity of money. Whereas the first operates through relative prices, the second is purely monetary.

From all the cases examined in this section, it is clear that a dependent economy with a fixed exchange rate can neither avoid "importing" inflation from the rest of the world, nor "export" domestically-generated inflation abroad. Thus, in the long run, the domestic rate of inflation will converge toward the world rate, as represented by the rate of change in the price of tradables. Nonetheless, when there are disequilibria, the rates of inflation in various economies with fixed exchange rates may differ substantially. A question therefore arises as to the duration of long-range adjustment. With respect to the period it takes for domestic prices to adjust to international prices, two a priori arguments come to mind. First, the degree of openness of the economy (measured, for instance, by the share of the tradable sector in total income) determines the velocity of adjustment: the higher the degree of openness, *ceteris paribus,* the shorter the adjustment period. This is due to the fact that the greater the share of the tradable sector in total expenditures, the sooner an exogenous disturbance will affect the balance of payments and the domestic rate of inflation.

Second, any change in the relative price W involves reallocating expenditures and resources between the two sectors, and it is precisely this reallocation that allows for the new equilibrium. The greater the possibility of substitution in demand between the two classes of goods, and the greater the factor mobility between the two sectors, the faster the economy will readjust. If, for example, a devaluation is immediately followed by a shift in demand to nontradables and a rapid transfer of factors to the production of tradables, the upward pressure on the domestic rate of inflation will be greater and a trade surplus will be created more quickly. On the contrary, at the same degree of openness, the more rigid the final demand and the less mobile the factors of production, the longer the period of adjustment and hence the time required to bring the balance of payments and the domestic market back to equilibrium. Since short-range

supply-and-demand substitution is limited, inflation rates in countries with fixed exchange rates may differ considerably for prolonged periods if the level of reserves or the chances of external financing are conducive.[12]

4 — The effects of international capital mobility

In the previous section, it was hypothesized that all the securities issued and exchanged within the economy were nontradable, just as some goods and services. However, for dependent economies that have reached a certain stage of financial development and integration into the international capital market, it would be more realistic to follow the pattern established for goods and divide credit instruments into two classes — those that can be traded internationally and those that are strictly domestic, with interest rates being set by the respective markets. This explains why there are different rates of interest on different instruments, and, since these rates are set via separate mechanisms, why international rates are not likely to become uniform even in the long run.[13]

Nonetheless, at this point our analysis will be limited to those instruments that can be exchanged on the world market. Since the hypothesis of a small country will be retained, the domestic and international nominal rates of interest will be the same. We will therefore be examining the macroeconomic adjustment of an open economy from the other extreme. Whereas it was previously assumed that all securities were nontradable, it will now be held that all instruments are fully mobile on the world market.

In discussing international capital transactions, the monetary approach has given little attention to the fact that monetary adjustment is effected not only through the goods and services market, but also via credit operations, and without changes in expenditure flows. We will now focus on this question, together with the efficacy of domestic monetary policy, which may be offset by private capital flows.

12 This problem is to an extent structural, since some economies are better prepared to mobilize resources and transfer them from one sector to another. In an economy with prohibitive tariffs and a limited import-export list, for instance, the possibility of short-and medium-term substitution is virtually zero. A recent example is the petroleum crisis. The European economies, with their greater degree of openness and relative facility for mobilizing internal resources, have managed to adjust their trade balances more quickly than the dependent economies of the Third World, including Brazil.

13 These characteristics are taken into account by R. Dornbush in "A Portfolio Balance Model of the Open Economy," *Journal of Monetary Economics* 1 (March 1975): 3-20.

The general structure of the model has already been described, but the credit-equilibrium equation may now be rewritten as

$$Z(r - E) - F = s \frac{\dot{B}_s^F}{P} = s \frac{c}{P} K \tag{27}$$

$$F = s \frac{\dot{B}^D}{P} + (s - 1) \frac{\dot{B}_l^F}{P} - Em - \Theta (m^d - m) \tag{28}$$

The demand for foreign credit, or the net inflow of financial capital, is clearly an excess demand not satisfied by internal sources. For this reason, the widely-held hypothesis that capital movements are basically speculative and effected through portfolio changes that reflect expected interest-rate differentials is rejected. Moreover, since long-term domestic and international inflation rates tend to converge, and since these rates are necessarily incorporated into nominal interest rates, the factor that determines net capital inflow will be the difference between natural interest rates, or, in the last instance, between expected rates of return on capital. Depending on these differences, the country will or will not be a net foreign borrower in the long run (in the short run, its relative position may be changed or even reversed due to monetary disturbances).[14]

To complete the model, it is necessary to specify the supply of foreign funds available to the country. Although the economy is too small to modify international interest rates, it will not be able to draw on a perfectly elastic supply of funds at the prevailing world rate. Owing to the risk of lending to the country, the marginal interest rate will be higher than the average, with the difference being explained mainly by the leverage of the economy. The risk to the lender will be a rising function of the net volume of indebtedness, such that the supply function will be

$$K = K\left(\bar{r},\ D - \frac{B_s^F}{c}\right) \tag{29}$$

where $\bar{r}$ is the nominal international interest rate, and D stands for the gross indebtedness of the economy. Alternatively, the interest rate paid by the country at a given moment, r^F, can be said to

[14] On this point, see the important study by Metzler, who bases his model on the hypothesis that one country is backward relative to the other, i.e. that the rate of return on capital is higher in the former than in the latter. L. A. Metzler, "The Process of International Adjustment under Conditions of Full-Employment: A Keynesian View," in *Readings in International Economics*, ed. R. E. Caves and H. G. Johnson (Homewood: Richard D. Irwin, Inc. 1968), pp. 465-86.

be comprised of two elements: a rate that does not take risk into account (r), and a risk premium associated with the net volume of indebtedness (v). That is,

$$r^F = \bar{r} + v\left(D - \frac{B_2^F}{e}\right), \qquad v' > 0 \tag{30}$$

This supply curve can eventually limit the indebtedness of the economy, for r^F will rise with D until $r^F - \pi \geqslant n$ and the excess absorption that is one of the factors responsible for the debt disappears. However, instead of considering what brings the economy to this limit, we will assume that r^F remains constant over the period covered by the analysis; in other words, we will deal with short-run situations only.

Visually, the operation of the system in the short run can be depicted through adaptation of the previous figures. Thus, in the upper part of figure 3, curves Z and F reappear (the latter redefined in [28]), together with an additional line r_0^F, which stands for a given level of net indebtedness. According to one possibility, the credit market will be in equilibrium at the international interest rate r_0^F (equal to the nominal domestic interest rate), excess absorption of goods OG, and net capital inflow HG, for a given rate of growth of domestic credit,[15] which determines the position of curve F.

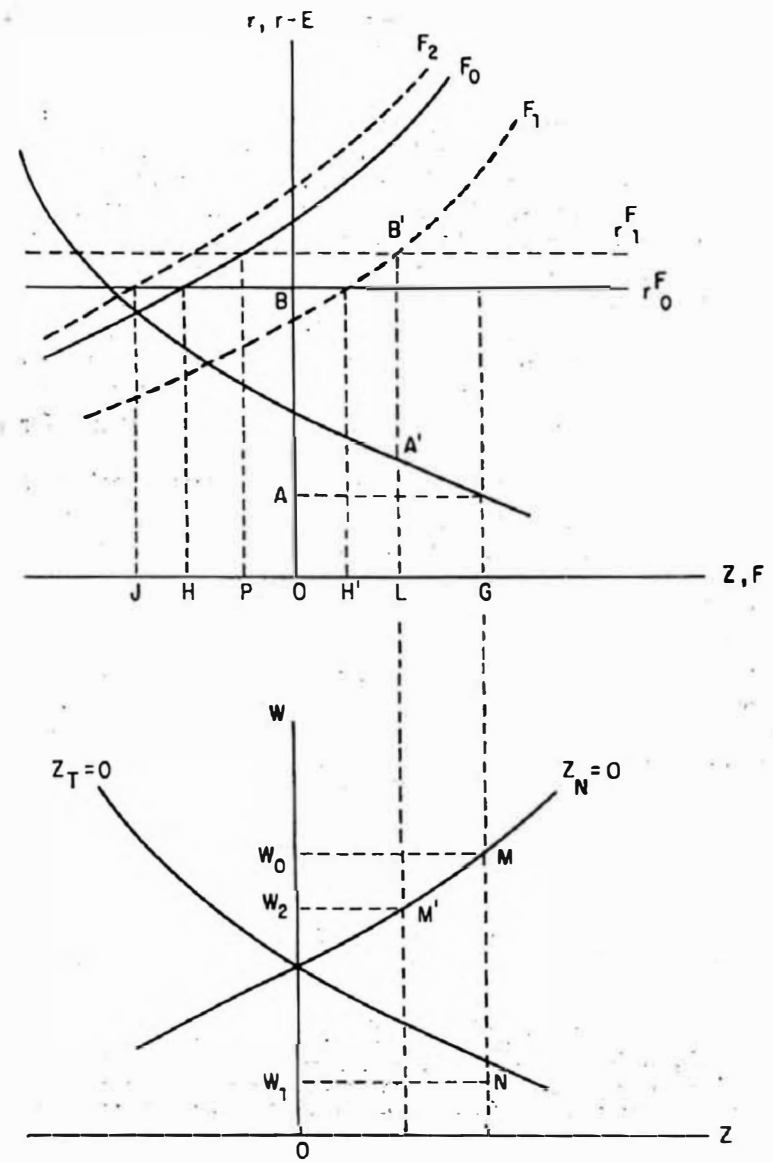
Quasi-equilibrium is defined as follows. At the interest rate r_0^F and inflation (international and domestic) $\pi_e = AB$, excess absorption will be OG, because the real rate of interest on the credit market, OA, will be lower than the natural rate n, which is supposedly constant. Excess absorption OG may be financed either by net domestic credit (curve F) or by net capital inflows (HG for a given rate of expansion of domestic credit). Foreign funds will be allocated partly to financing excess expeditures OG, and partly to fulfilling the replacement demand for money by the private sector (HO).

Domestic equilibrium is at point M in the lower part of the same figure. Under these circumstances, the relative price W_e will

15 So as to avoid any problems due to the scale of the figure, it will be assumed that $s = 1$ throughout the rest of this section; thus, the measure in the upper part of the figure will directly indicate the net capital inflow (HG). The multiplier s has no bearing on the conclusions drawn. When discussing the balance of payments, the service account will not be taken into consideration.

Figure 3

INTERNAL AND EXTERNAL EQUILIBRIUM UNDER FIXED EXCHANGE RATES AND CAPITAL MOBILITY



lead to a trade deficit equal to excess absorption OG. The economy will have an overall balance-of-payments surplus and a rate of accumulation of reserves HO, a trade deficit OG, and a net capital inflow HG. The balance-of-payments surplus is the outcome of an excess demand (flow) for money which is not fully met by internal credit. It is for this reason that Johnson contends that the monetarist approach is a "money account" or reserve-flow theory, in which what happens "above the line" in the balance of payments does not matter.[16]

In this case, unlike that considered in the preceding section, it cannot be argued that equilibrium depends on the balance-of-payments surplus or its components being nil, since these are now goals of economic policy. For the balance-of-payments surplus to be nil, the monetary authorities would have to control the rate of expansion of domestic credit in such a way that curve F passed through point B, where the net capital inflow would exactly cover the trade deficit. For each of the items in the balance of payments to be nil, the authorities would have to free the domestic interest rate from the international rate. Given perfect capital mobility, this could only be done through interest-equalization taxes or other rationing measures. In other words, to attain equilibrium in the balance of payments, it would be necessary to meet the flow demand for money via domestic credit only.

The adjustment problems examined in the previous section can also be viewed in the light of capital mobility. To recapitulate, these problems are a change in the rate of expansion of domestic credit, a foreign-exchange devaluation, a change in the international interest rate, and a change in the world rate of inflation as it affects the prices of tradables.

An increase in the rate of expansion of domestic credit, starting from the equilibrium situation shown in figure 3, will have simple repercussions. Curve F_0 will shift to the right (to F_1 for example) and raise the amount of internal credit available. Since the nominal interest rate is fixed at r_J^F, there will be substitution between the internal and external credit channeled to the private sector, nd consequently a reduction in the net capital inflow. This may occur almost instantaneously, such that credit sources change while hardly affecting total credit flows. This type of substitution, emphasized

[16] H. G. Johnson, "The Monetary Theory of Balance of Payments Policies," in *The Monetary Theory of the Balance of Payments*, ed. J. Frenkel and H. G. Johnson (London: G. Allen & Unwin, 1976), pp. 262-84.

by Sjaastad,[17] has not been contemplated by the monetary approach, which assumes that changes in the balance of payments are inevitably preceded by disequilibria in the money market. In our model, a change in domestic credit policy will not necessarily lead to disequilibrium between the supply and demand for money, as long as there is sufficient capital mobility or reasonable integration of the domestic and international credit markets. Instead, the sole impact of the policy change will be to influence the credit market directly by means of substitution between the credit sources, and without altering the total flow of available funds.

Regardless of the specification of aggregate demand, this substitution effect will explain the whole process, and there will be no way for monetary policy to affect expenditure, income, or balance-of-trade flows. The adjustment will be realized through the recomposition of credit sources alone, without creating any disequilibria in stocks. Despite retention of the proposition that domestic credit policy can only modify the composition of the monetary base between domestic and international assets, the mechanism through which assets are reallocated is quite different from that described by the monetary approach.

In the situation depicted in figure 3, the trade deficit will still be OG, but since the net capital inflow will be only $H'G$, the overall balance-of-payments deficit will be OH'.

Turning to the case of a foreign-exchange devaluation, let us assume initial equilibrium at interest r_0^F, a trade deficit OG, a balance-of-payments surplus HO, and a relative price W_0 (figure 3). A devaluation will immediately cause disequilibrium in the money market, shifting curve F_0 to F_1 and raising capital inflows from HG to JG in order to meet the excess demand for money. At the same time, W_0 will fall (to W_1 for instance), thereby inducing a substitution effect to the advantage of nontradables that may be favorable enough to turn the trade deficit into a surplus, as at point N. Under these circumstances, the devaluation will improve the balance-of-payments position from the commercial as well as the financial side.

The stock disequilibrium in the money market may be offset at once by capital inflows. Thus, F_0 will shift to F_1 then quickly return to its former position. Due to the drop in the relative price W, the markets for the two classes of goods will remain in disequilibrium, and a temporary increase in the domestic rate of inflation will therefore be required.

17 L. A. Sjaastad, "On the Monetary Theory of the Balance of Payments: An Extension," paper presented at the 6th Konstanz Seminar on Monetary Theory and Policy, Konstanz, June 1975.

The basic difference between these results and those in the previous section is that the private sector will no longer need to reduce its expenditure flow in order to acquire the additional money it desires to bring the real value of its stock back to the pre-devaluation level. Since the nominal interest rate remains constant, the real money stock can be adjusted through importing foreign capital. With respect to the market for goods, though there will be no absorption effect, there will be a substitution effect in favor of nontradables that will shift expenditures in their direction. Given capital mobility, a devaluation should therefore exert less impact on the balance of trade, but perhaps more impact on the balance of payments as a whole. If the devaluation is meant to correct trade disequilibria, capital mobility will limit its range to a potential substitution — not always possible at short term between the two classes of goods.

The effects described will clearly be temporary, for as capital flows increase and as the trade deficit is reduced, the excess demand for money will disappear, shifting F_1 back to F_0. Moreover, due to excess domestic demand, the rate of inflation on nontradables will begin to rise, and W_1 will return to W_0. That the effect of devaluation is transitory is therefore inevitable in a small open economy. It should be remembered, however, that the monetary authorities are enabled to increase their reserves during the adjustment process.

At this point, it is easy to foresee the results of a change in international credit terms. Let us assume that the nominal interest rate goes from r_0^F to r_1^F, owing either to a change in the world situation or to a modification in the net indebtedness of the economy due to former capital surpluses. At the outset, the rate of inflation is AB, the real interest rate OA, and the relative price W_0 (figure 3). The higher interest rate will lower the demand for money and reduce expenditures, so that credit flows will be brought into equilibrium with less capital entering from abroad (PL). Simultaneously, the reduction in expenditures will shrink the trade deficit and create an excess supply in the domestic market, thus forcing down the relative price. The deflationary pressure will exist long enough to re-establish equilibrium in the market for goods at point M', where excess absorption OL will again equal the trade deficit. Hence, a rise in the international interest rate will ultimately reduce both balance-of-payments accounts, lower the surplus to OP, and reduce the relative price between the classes of goods.[18]

[18] The same results could be produced by means of a tax on capital movements that raised the interest rate on private borrowing from abroad.

The process by which the economy imports inflation is quite similar to that examined in the earlier section, as can be easily deduced by the reader.

5 — The indexed exchange rate

Various economists have suggested that indexation (or "monetary correction", as it is known in Brazil) may be capable of eliminating the risks associated with unexpected changes in price levels or in the purchasing power of a given currency, and therefore of minimizing the allocative and distributive effects of inflation. The hypotheses underlying indexation are classic. On the one hand, since money is considered neutral in the long run, indexation is also held to be neutral and unable to affect the real equilibrium of the economy due to the fact that it involves purely nominal adjustments. On the other hand, changes in relative prices are assumed to be incapable of generating inflation, which is viewed as the result of monetary policy. Quite to the contrary, in the absence of indexation, inflation may cause undesirable changes in relative prices that could be avoided or at least reduced were monetary correction introduced in certain strategic markets.

Although indexation can hasten or reduce the costs of adjustment in given markets or in the economy as a whole,[19] we will limit our analysis to monetary correction of the exchange rate. The assumption that no other markets are subject to such correction will allow us to focus more clearly on the case of an indexed exchange market.

Fixed exchange rates are an international regime par excellence, since all the economies that participate in the world market are exposed to the same monetary and real changes that occur in it. In the long run, each country should register a rate of inflation close to the world rate, and, if financial capital is mobile, maintain a nominal interest rate that is stable relative to the international rate. Given the automatic adjustment mechanism inherent in the system, fixed exchange rates therefore impose a sort of monetary discipline on each economy in accord with its balance-of-payments position and its international liquidity over time. It has been widely recognized, especially since the work of Mundell[20] and others,

19 Since they are beyond the scope of this paper, the various advantages of indexation will not be considered. For a discussion of these benefits, see M. Friedman, *Monetary Correction*, Occasional Paper, n.º 41 (London: Institute of Economic Affairs, 1974).

20 R. A. Mundell, *International Economics* (New York: MacMillan, 1968), part 3; *idem*, *Monetary Theory*, part 2.

that monetary policy cannot be manipulated independent of the balance of payments, i.e. that it cannot be used to achieve internal equilibrium.

Failure to obey the rules leads to occasional devaluations or to the choice of another system. It may be that fixed exchange rates have gradually been abandoned since World War II due not so much to the fundamental disequilibrium admitted by the International Monetary Fund, as to monetary authorities using policy instruments to attain internal equilibrium. In fact, the majority of governments have not allowed the automatic adjustment mechanism — summed up in the phrase that a deficit economy should contract and a surplus economy should expand — to operate. Since contraction can mean unemployment and expansion may signify inflation (in virtue of lags and the rigid response of nominal magnitudes, mainly in wages and certain prices), governments tend to allow external disequilibria to exist for fairly long periods so as not to upset internal equilibrium or stability. This eventually forces them to forsake the fixed exchange rate and temporarily realign their rates.

It is even more difficult to adhere to the rules of this regime in dependent economies, where the aim of domestic equilibrium or stability is coupled with the goal of economic growth. To finance growth, the monetary authorities generally accept a rate of inflation that is higher than the world rate, and this acceptance eventually leads to abandonment of the fixed exchange rate. In this section, it will be assumed that this is the case — in other words, that the initial pressure to relinquish the system comes from within the economy, and is only secondarily reinforced by pressure from without (e.g. imported inflation). This implies that the direction of causality is essentially from domestic inflation to the exchange rate, rather than from the exchange rate or international prices to domestic inflation.

Once the economy has decided to accept a rate of inflation that is higher than the world rate, it may realign domestic and international prices via either a flexible exchange rate or foreign-exchange indexation. In a way, automatic indexation under persistent domestic inflation parallels a fixed exchange rate given internal price stability. Indexation brings together characteristics of flexible and fixed rates. Like the former, it breaks the tie between domestic and world inflation; and like the latter, it maintains the link between the balance of payments and the monetary base. An indexed rate may thus be considered a special fixed rate. Monetarily speaking, an indexed exchange rate, unlike a flexible rate and a free exchange market, keeps the economy within the international community. Hence, dependent economies may have

opted for indexed rates instead of free exchange rates so as not to lose a control and planning instrument, as well as to avoid erratic fluctuations that would be detrimental to trade and capital flows.

It must be remembered, however, that no foreign-exchange system can bypass or eliminate adjustments related to real changes, such as variations in the terms of trade, shifts in international demand, changes in productivity or income levels, etc. What it can do is ease or accelerate the adjustment process and diminish its costs. As already pointed out, there are *prime facie* arguments in favor of indexation in dependent economies. All that remains is to determine whether, under specific conditions, this type of indexation reduces the costs of adjustment.

As Friedman points out, it would be preferable to eliminate the reasons for indexation since it is never cost free.[21] This being impossible, it is best to accept the system and strive to minimize any undesirable effects. Moreover, there is an implicit trade-off involved: the economy either pays more at present for adjusting to the current international monetary situation and experiences lower inflation in the future, or pays less now at the cost of higher inflation later on. Exchange indexation, like indexation in general, should be viewed from this perspective. If it succeeds in limiting the repercussions of policy-induced monetary changes — expressed in terms of alterations in the relative prices between "flexprice" (nontradables) and "fixprice" (tradables) markets — it will contribute to the real stability of the economy.

Given the point to which theory has developed, there is little to be gained from setting criteria for an indexing system that includes the capital account, because there is no way to safely predict the impact of indexation on interest rates and capital flows. This is not to say that indexation does not affect the capital account, but merely to say that its impact on this account will be left outside the model and will be offset by other policy instruments, if desirable. The question of indexation will therefore be examined on the assumption that international capital is immobile, in which case the balance of payments and the balance of trade are identical.

Formally, exchange indexation is based on a function that links the rate of change in exchange rate g to a set of observable variables. Once determined, this function serves as an automatic guideline for the monetary authorities in their attempts to minimize the effects of monetary disturbances and adjust the economy as quickly as possible.

[21] Friedman, *Monetary Correction*.

In Brazil, two indexing criteria have been widely discussed. One is based on the argument that the rate of devaluation should maintain the balance of trade in equilibrium at all times. This approach simulates a flexible exchange rate as set on a free market without government intervention. Its success depends on accurate knowledge of the general equilibrium system that determines the exchange rate. Since such knowledge is in fact unobtainable, if this criterion is chosen, it would be preferable to openly adopt a flexible exchange rate, which would achieve the desired end more efficiently than the monetary authorities.

This criterion having been rejected due to its being nonoperational because of the difficulty of estimating the structural model that describes the economy, a second alternative arises: the purchasing-power parity theory developed by Cassell.[22] According to the absolute version of the parity theory, in the absence of real changes in the economy, the long-term equilibrium exchange rate will move toward the ratio between internal and external general price levels; according to the relative version, starting from initial equilibrium, the rate will vary with the relative change in the two price levels. Under this criterion, the forementioned function is specified as

$$g = \pi - \bar{\pi}$$

where π is the rate of change in the general price level in the country, and $\bar{\pi}$ is the equivalent rate in the rest of the world or in the representative sample of countries engaged in mutual trade.

The general price indices include both tradables and nontradables. Since the economy under analysis is small, the parity theory will be simplified by replacing $\bar{\pi}$ with $\bar{\pi}_T$, the rate of change in the prices of tradables in the rest of the world, so that

$$g = \pi - \bar{\pi}_T \tag{32}$$

[22] For an extensive theoretical and empirical evaluation of the parity theory, consult L. H. Officer, "The Purchasing-Power Parity Theory of Exchange Rates: A Review Article," *IMF Staff Papers* 23 (March 1976): 1-60. Another useful reference is L. B. Yeager, *International Financial Relations* (New York: Harper and Row, 1976). The problems concerning the choice of a representative price index will not be covered here.

An interesting exercise in simulating an equilibrium exchange rate for Brazil (though in the context of partial equilibrium) was performed in order to assess the parity theory; see A. C. Pastore, J. R. M. de Barros and D. Kadota, "A Teoria da Paridade do Poder de Compra, Minidesvalorizações e o Equilíbrio na Balança Comercial Brasileira," *Pesquisa e Planejamento Econômico* 6 (August 1976): 287-312. In an earlier article, several of the points mentioned above were considered in an analysis of the Brazilian minidevaluation policy; see A. M. Silva, "Bases da Política de Minidesvalorização," *Estudos Econômicos* 6 (May 1976): 97-112.

The basic difference between criteria (31) and (32) is that whereas (32) guarantees that the relative price between nontradables and tradables ($W = P_N/P_T$) will remain constant, (31) allows for the possibility of changes in this relative price. To confirm this, let

$$\frac{\dot{W}}{W} = \pi_N - \pi_T = \pi_N - \bar{\pi}_T - g \tag{33}$$

since, under free-trade conditions, $\pi_T = \bar{\pi}_T + g$. Substituting (32) into (33), we find that $\pi = \pi_n$, and thus verify that the relative price W will be stabilized. With respect to (31), let

$$\begin{aligned} \pi &= a\pi_N + (1 - a)\ \pi_T \\ \bar{\pi} &= a\pi_N + (1 - \bar{a})\ \bar{\pi}_T \end{aligned} \tag{34}$$

where a and $\bar{a}$ stand for the share of nontradables in total expenditures of the country and of the rest of the world, respectively. Substituting in (31) and then (33) yields

$$\frac{\dot{W}}{W} = \frac{\bar{a}}{a}\ (\bar{\pi}_N - \bar{\pi}_T) \tag{35}$$

Due to the indexing procedure, the relative price will be biased upward or downward, depending on whether the domestic rate of inflation in the rest of the world is higher or lower than the rate of inflation on tradables. In the first case, the indexing criterion will underestimate the exchange adjustment required, and in the second case will overestimate it. To avoid such biases, we shall reject (31) in favor of (32), which stabilizes the relative price and keeps monetary shocks (whether internal or external) from distorting the allocation of expenditure and real resources between the domestic and international sectors.

Once it is assured that there will be no changes in relative price W, a single general price level and a single rate of inflation can be defined for the economy:

$$\pi = \pi_N = \pi_T = \bar{\pi}_T + g \tag{36}$$

The complete model, on the hypothesis W W_0, may now be presented as follows.

Credit market:

$$Z\,(r - E) + E\,m + \Theta\,[f\,(r) - m] = s\,\mu^{D}\,\frac{B^{D}}{P} - \frac{(s-1)}{1 + a\,(W_0 - 1)}\,Z_T\,(Z,\,W_0) \tag{18a}$$

Market for goods:

$$Z_N\,(Z,\,W_0) = 0 \tag{37}$$

Expected inflation:

$$E = E\,[\pi\,(t)] \tag{15}$$

Price and balance-of-payments adjustments:

$$\pi = E + \lambda Z_N\,(Z,\,W_0) \tag{17a}$$

$$\frac{\dot{B}^{F}}{P} = -\frac{1}{1 + a\,(W_0 - 1)}\,Z_T\,(Z,\,W_0) \tag{4a}$$

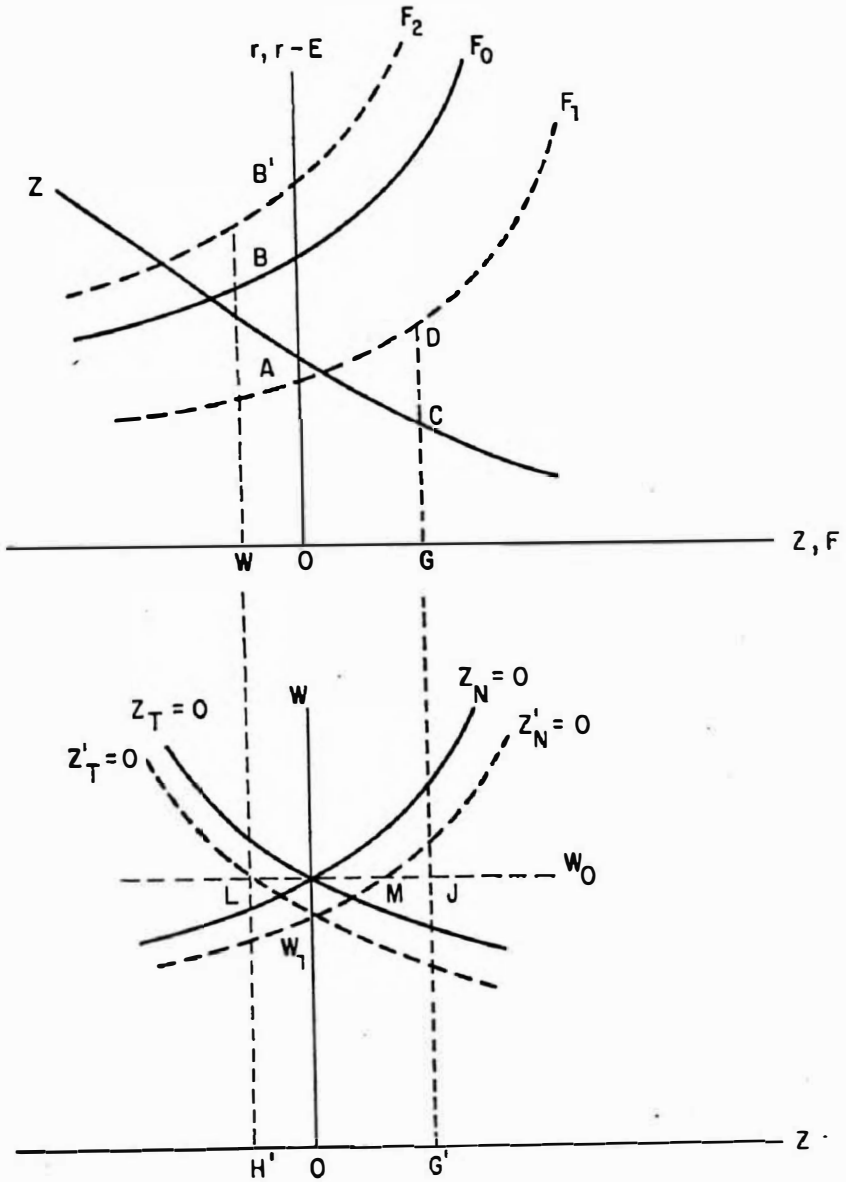
Assuming that relative price W_0 has been stabilized at the level that brings the markets for goods to equilibrium when excess absorption $Z = 0$, the operation of the system can be visualized with the aid of figure 4. Quasi-equilibrium, for a rate of international inflation π_T and a rate of expansion of domestic credit μ^D, will occur when curve F crosses the vertical axis at point B, thereby determining a real interest rate OA, a nominal rate $OB = OA + AB$, and a real and expected rate of inflation AB. Under these conditions, $Z = 0$, and the relative price between the two classes of goods (and hence the balance of payments) will be at equilibrium. In order to sustain a rate of inflation $\pi_0 = AB$, the rate of growth of domestic credit will have to be

$$\mu_0^{D} = \frac{\pi_0}{1 - c} \tag{38}$$

Starting from the situation described, an increase in the rate of monetary expansion will temporarily shift F_0 to F_1, and thus lower the expected real interest rate to point J (figure 4), which corresponds to excess absorption OG. The resulting excess demand for both classes of goods will lead to a higher domestic rate of inflation and a larger trade deficit. However, since relative price

Figure 4

ADJUSTMENT OF THE ECONOMY TO AN INDEXED EXCHANGE RATE



W_0 will be unable to adjust due to the fact that any increase in the domestic rate of inflation is automatically offset by a stronger devaluation, neither expenditures nor factors will be transferred from one sector to the other. Thus, the only way for the system to return to equilibrium will be to eliminate excess absorption OG.

The adjustment of the economy can be explained as a Wicksellian cumulative process. Initially, the rise in the rate of monetary expansion will lower the expected real rate of interest. This, in turn, will generate excess expenditures on goods and services which will result in a higher rate of inflation and a balance-of-payments deficit. Excess absorption OG will be gradually eliminated by two phenomena that will reverse the cumulative process. On the one hand, the increase in the domestic rate of inflation will comprise an internal drain on the credit market, pushing the real flow supply of credit down for the same rate of change in the nominal supply. On the other hand, there will be an external drain via the balance of payments, which, by absorbing money, will reduce the rate of growth of the monetary base to below the rate of expansion of domestic credit at the same rate of inflation. This phenomenon is illustrated by the equation describing the change in the real money supply over time:

$$\frac{\dot{m}}{m} (1 - c)\ \mu_D - c\frac{P_T Z^T}{B^F} - \pi \qquad (39)$$

As figure 5 shows, the rate of change in the real money supply will immediately shift from zero to the new level set by the monetary authorities. Then begin the internal and external drains. The rate of inflation starts to climb and raises expected inflation. The excess demand for tradables leads to a trade deficit and a loss of reserves. In response, curve F_1 moves in the direction of F_0 and eventually passes this point (figure 4).

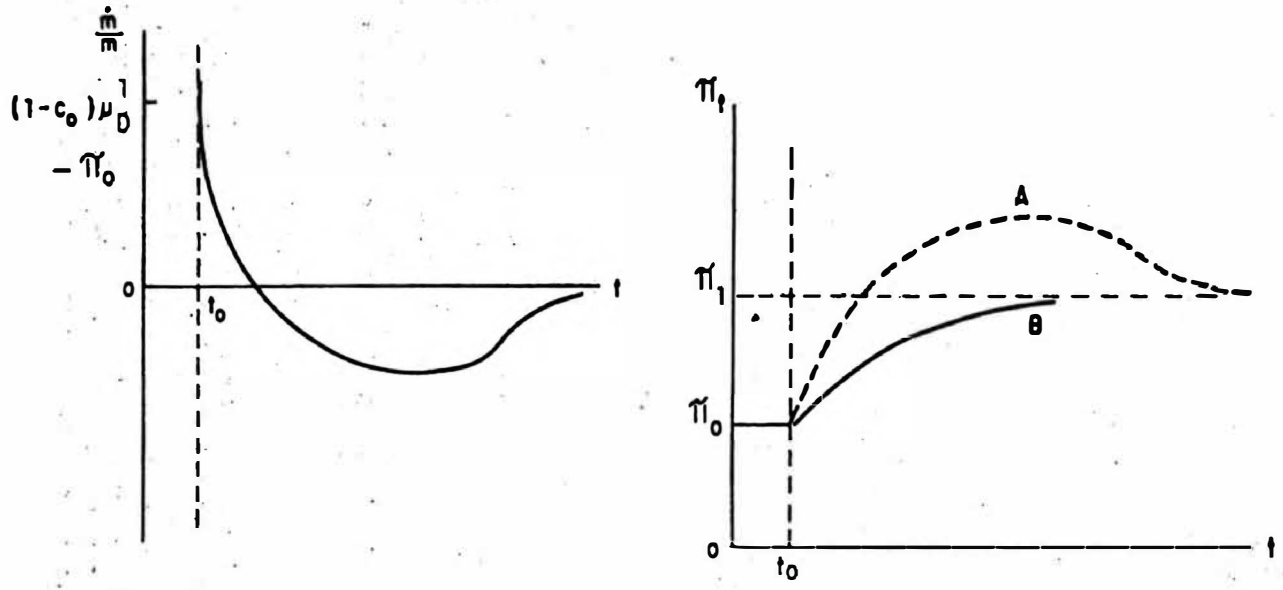
At the new equilibrium, the economy will experience a higher rate of inflation (AB' for instance) in accord with the new rate of monetary expansion:

$$\pi_1 = \mu_D^1\ (1_1 - c),\ c_1 < c_0 \qquad (40)$$

In the meantime, in virtue of the external drain via the balance of payments, there will be a substitution of domestic assets for reserves in the monetary base, so that the domestic rate of inflation will rise less than the rate of expansion of domestic credit. Turning back to equation (39) and thinking in terms of restoring the economy to equilibrium, one is led to ask about the relative importance of adjusting the rate of inflation (internal drain) and the

Figure 5

TIME PATHS OF THE RATE OF CHANGE IN THE REAL MONEY SUPPLY AND IN THE RATE OF INFLATION



deficit (external drain). In other words, to what point should the real money supply be adjusted through the nominal supply, and to what point through the price level? To answer this question, three factors must be taken into account: the degree of openness of the economy, the velocity of adjustment of the domestic rate of inflation to excess demand, and the velocity of adjustment of expectations concerning inflation.

The degree of openness of the economy, measured by the share of tradables in total expenditures, is important because it shows how the excess demand generated by domestic credit policy will be divided between tradables and nontradables. The higher the degree of openness, the greater the extent to which the adjustment will be directed to the balance of payments and the nominal quantity of money. The velocity of adjustment of the rate of inflation, represented by λ, indicates how soon excess domestic demand will raise inflation and subsequently erode the real flow of credit and money. The higher this velocity, the greater the degree to which the adjustment should affect the price level. Finally, since the rate of inflation is partly conditioned by expectations, equation (17a) makes it clear that the extent to which the adjustment affects the price level will also depend on the pace at which these expectations change.

An important implication of the above is that, unlike the quantitative theory models for closed economies, our model for an open economy does not require that the path of inflation overshoot its new equilibrium level during the adjustment period. This will depend, instead, on the relative importance of the balance-of-payments adjustment. In addition, since the private sector has another means of adjusting its real money supply, an open economy may reach its new quasi-equilibrium in less time. Two typical paths of inflation are depicted in figure 5. Path A is what ought to be expected in a closed economy: the larger part of the curve is over the new equilibrium level π_1.[28] This could but will not necessarily be the case in an open economy. If the external drain is sufficient, the path could be B or anywhere between A and B, with the rate of inflation approaching the new equilibrium asymptotically. This second equilibrium will be lower than that corresponding to the new rate of expansion of domestic credit, i.e. $(1 - c_\theta)\mu_D'$. The only restriction is that the real money supply will now be smaller, as indicated by the area beneath curve $\dot{m}/m$ being larger than the part above it.

[28] This is being disregarded due to the fact that in a closed economy the equilibrium rate of inflation is over π_1.

Another salient feature of an indexed exchange rate is that it shelters the domestic economy from monetary disturbances from abroad which are reflected in changes in the international rate of inflation. Any alteration in this rate (represented here by the rate of inflation on tradables) is automatically countered by a change of the same magnitude in the rate of devaluation. Thus, the indexing procedure analyzed protects the economy from monetary shocks of external origin, as well as from internally-generated disturbances and the accompanying distortions in the allocation of factors and resources between the domestic and international sectors.

If an economy refuses to accept the discipline of a fixed exchange rate and align its rate of inflation with the world rate, an indexed exchange system may be preferable to a fixed or flexible system. Nonetheless, given the limitations of indexing itself, it should be borne in mind that the model presented is purely monetary. Nothing can keep real changes, whether independent or induced, from altering the equilibrium relative price (W_0). Even if the monetary authorities initially set this price at the equilibrium level, how will they respond to a real change that permanently modifies it? This appears to be one of the major problems in any indexing scheme — how to reconcile it to relative price changes that are essentially equilibrating.

To illustrate this problem, suppose that demand turns in favor of tradables, so that curves Z_N and Z_T shift downward to establish a new equilibrium relative price $W_1 < W_0$ (figure 4). This will generate a balance-of-payments deficit and deflationary pressure on the domestic market. However, since indexation will prevent W_0 from changing, the disequilibrium will tend to persist. As a result, curve F will keep shifting to the left as long as the deficit exists, an the adjustment will be made at a higher real interest rate but with insufficient global absorption. Eventually, the balance of trade will momentarily reach equilibrium at point L, but since the deflationary trend will continue in the system, point L will not be maintained. Whereas deflationary pressure will force the economy to the left of L, the surplus will push it to the right. Thus, in the absence of other measures, the economy will fluctuate around L or move to the left indefinitely.[24]

[24] On the one hand, if the monetary authorities try to halt the deflationary trend by expanding credit, the economy may shift to point M, but with a deficit in the balance of payments. There will continue to be an inconsistency between internal and external equilibrium, and a stop-go monetary policy may be adopted. On the other hand, it may prove possible to solve the problem via trade policy (forcing the relative price down by erecting a trade barrier, for instance). However, this involves instruments that have no relation to indexation.

Unless indexation is temporarily abandoned in order to allow for the downward adjustment of W, it will be impossible to simultaneously achieve balance-of-payments and domestic equilibrium. This inconsistency underlines the fact that, in a general-equilibrium system, setting a relative price out of equilibrium prevents the system as a whole from reaching equilibrium. Whereas a fixed exchange rate stabilizes the price of tradables (the exchange rate) and leaves that of nontradables free to adjust, a flexible rate sets the price of nontradables and leaves that of tradables (the exchange rate) free to adjust. Hence, both systems stabilize an absolute price (or its rate of change) which is a nominal magnitude. In contrast, an indexed rate sets a relative price, and is therefore subject to problems in case of real changes.

In sum, the analysis suggests that an indexed exchange rate not only protects the system from both internal and external monetary shocks, but that it probably reduces the time and costs of adjustment. It is characterized, however, by a rigidity in relative prices that may prevent complete equilibrium from being reached in the event of real changes. The subsequent policy impasse arising from the impossibility of using monetary instruments to attend to both internal and external disequilibria may have to be overcome via relinquishment of the indexed rate.

This shortcoming does not invalidate exchange indexation for the type of economy described. If the most important and most frequent disturbances to which the economy is subject are monetary in nature, and if the real changes are linked to structural factors (a hypothesis that appears to underlie many quantitative theory arguments), an indexed rate may be the most efficient. However, should real changes be frequent and significant and therefore necessitate constant changes in relative prices, the benefits of indexation become questionable.

Unplanned settlement in the Amazon region *

Jean Hébette **
Rosa E. Acevedo Marin **

1 — Introduction

This article examines certain aspects of the government policy related to the settlement of the Amazon. It is based on a year and a half of field work on settlement, generally unplanned, along the Belém—Brasília highway. The study covers 52 *municípios* in the following microregions: Bragantina (Pará), Imperatriz (Maranhão), far northern Goiás, the lower Araguaia (Goiás), the mid-Tocantins-Araguaia (Goiás), and the *município* of Ceres (Goiás), taking into account unplanned and occasional planned colonization in an area totaling 206 thousand square kilometers. The conclusions rest mainly on personal observation and first-hand material. Even so, coupled with direct and indirect (though at times more superficial) knowledge of other parts of the Amazon, they allow for hypothetical generalizations that might serve as the basis for further debate.

From the empirical standpoint, maximum care was taken to utilize the most complete data possible and to avoid drawing a priori conclusions. In this respect, the major obstacle encountered in studies such as ours is limited access to certain sources of information. As a result of our efforts, we hope to have overcome some errors or gaps in information and to have corrected some previous misinterpretations.

* This study was sponsored by the Ministério da Educação e Cultura; the Universidade Federal do Pará, IPEA and the Ford Foundation.

** Universidade Federal do Pará.

2 — Settlement and land distribution

The issue under scrutiny is government policy concerning the settlement of the Amazon, specifically the so-called "spontaneous" or "unplanned" colonization along the Belém—Brasília highway. In fact, planned and unplanned settlement are not thouroughly distinct, with one characterized by outside intervention (government or private) and the other free thereof. The difference refers to the timing and intensity of government action. Recent experience along the Transamazon and Belém—Brasília highways furnishes examples of the two types, respectively. Planned settlement, like that observed along the Transamazon, is marked by direct government initiative. Such initiative is generally taken in order to open a new region, and involves choices regarding the area to be occupied (the location, size, surveying and distribution of plots, as well as the location of housing), the individuals or groups to be encouraged to move to the region, the type of economic activites to be conducted, etc.

In the case of unplanned settlement, these decisions are left to the colonists. Outside influence becomes only gradually apparent and tends to be sporadic and less organized. Even so, unplanned settlement is subject to overall regional and national integration policies. Land statutes, for instance, affect the settler from the time he first migrates until he opts to remain at his new destination or is pushed farther onward. Transport and communications networks also exert an indirect impact. Other policies (e. g. land-distribution, agricultural, and credit policies) have a direct influence which becomes increasingly evident as the new frontiers are integrated into the national economy. In this study, the focus will be on the impact of some of these policies on the migrant population.

Land distribution is one of the key factors, though hardly the single cause, underlying the problems that afflict the agricultural sector. When analyzed in terms of its effects over time, it is an excellent indicator of the trends and eventual impact of the policies adopted at different points in history.

The land-distribution patterns in the 53 *municípios* studied in the area of the Belém—Brasília highway vary widely from region to region and with socio-economic background. These patterns have been grouped into four classes designated as "highly concentrated", "concentrated", "fragmented" and "highly fragmented". The characteristics of each class are indicated in tables 1 to 5, which give the land-distribution figures and the corresponding Lorenz curves for five representative *municípios*.

Table 1

Rural Holdings by Size Class: Município *of Araguatins, Goiás, 1976*

Size Class (Hectares)	Number of Holdings		Area in Size Class (Hectares)		
	Total	%	Total	Average per Holding	%
< 2	3	.19	3	1	.00
2 ⊢ 5	18	1.12	64	3	.01
5 ⊢ 10	18	1.12	140	7	.02
10 ⊢ 25	181	11.27	3 661	20	.41
25 ⊢ 50	373	23.23	15 482	41	1.73
50 ⊢ 100	357	22.23	28 709	80	3.21
100 ⊢ 200	265	16.50	37 054	139	4.14
200 ⊢ 500	244	15.19	80 716	330	9.03
500 ⊢ 1 000	53	3.30	40.583	765	4.54
1 000 ⊢ 2 000	37	2.30	53 220	1 438	5.95
2 000 ⊢ 5 000	41	2.55	121 224	2 956	13.56
5 000 ⊢ 10 000	4	.25	31 222	7 805	3.49
≥ 10 000	12	.75	481 912	40 159	53.91
Total	1 606	100.00	893 990	556	100.00

Source: Instituto Nacional de Colonização e Reforma Agrária (INCRA).

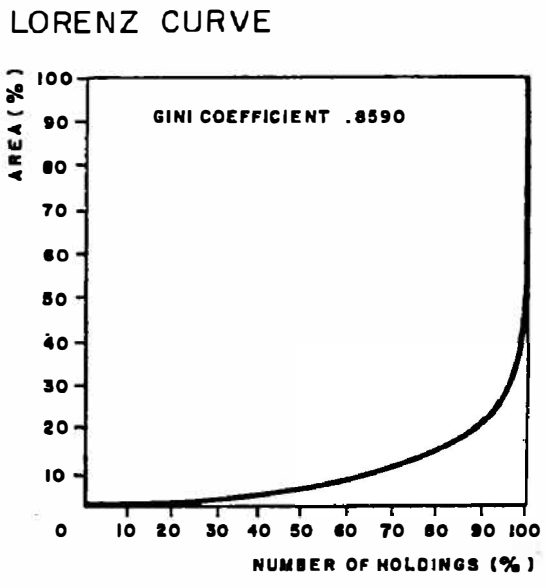


Table 2

Rural Holdings by Size Class: Município *of Guaraí, Goiás, 1976*

Size Class (Hectares)	Number of Holdings		Area in Size Class (Hectares)		
	Total	%	Total	Average per Holding	%
< 2	0	.00	0	0	.00
2 ⊢ 5	0	.00	0	0	.00
5 ⊢ 10	4	.47	30	7	.01
10 ⊢ 25	19	2.24	371	19	.11
25 ⊢ 50	82	9.68	2 884	35	.85
50 ⊢ 100	126	14.88	8 858	70	2.61
100 ⊢ 200	235	27.74	32 727	139	9.65
200 ⊢ 500	232	27.39	69 844	301	20.60
500 ⊢ 1 000	81	9.56	55 990	691	16.51
1 000 ⊢ 2 000	46	5.43	61 705	1 341	18.20
2 000 ⊢ 5 000	17	2.01	55 065	3 239	16.24
5 000 ⊢ 10 000	4	.47	33 247	8 311	9.80
≥ 10 000	1	.12	18 392	18 392	5.42
Total	847	100.00	339 113	400	100.00

Source: INCRA.

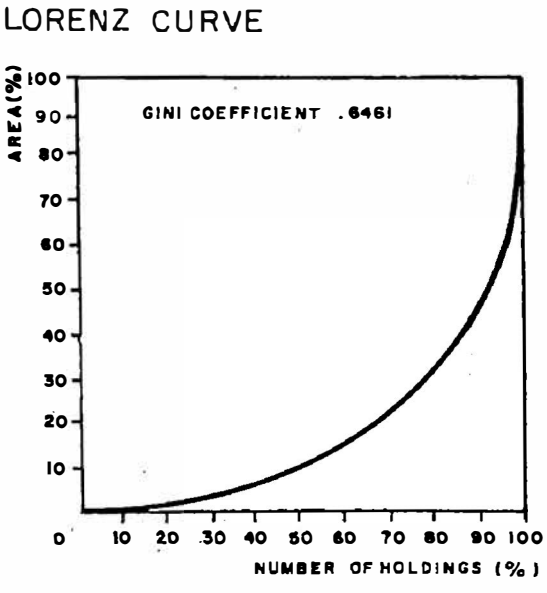


Table 3

Rural Holdings by Size Class: Município *of Santa Izabel do Pará, Pará, 1976*

Size Class (Hectares)	Number of Holdings		Area in Size Class (Hectares)		
	Total	%	Total	Average per Holding	%
< 2	13	1.27	13	1	.02
2 ⊢ 5	56	5.46	159	2	.29
5 ⊢ 10	42	4.10	276	6	.50
10 ⊢ 25	444	43.32	8 301	18	15.00
25 ⊢ 50	283	27.61	8 506	30	15.37
50 ⊢ 100	96	9.37	6 362	66	11.49
100 ⊢ 200	46	4.49	6 312	137	11.41
200 ⊢ 500	28	2.73	8 820	315	15.94
500 ⊢ 1 000	12	1.17	8 084	673	14.61
1 000 ⊢ 2 000	3	.29	3 327	1 109	6.01
2 000 ⊢ 5 000	2	.19	9 178	2 589	9.36
5 000 ⊢ 10 000	0	.00	0	0	.00
≥ 10 000	0	.00	0	0	.00
Total	1 025	100.00	55 338	53	100.00

Source: INCRA.

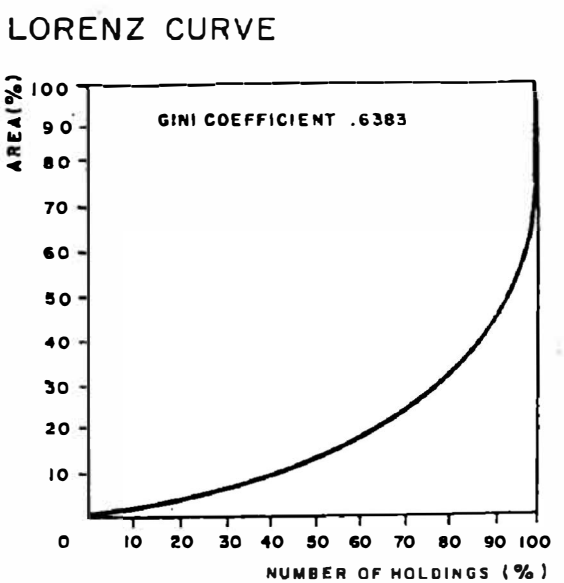


Table 4

Rural Holdings by Size Class: Município *of Igarapé-Açu, Pará, 1976*

Size Class (Hectares)	Number of Holdings		Area in Size Class (Hectares)		
	Total	%	Total	Average per Holding	%
< 2	4	.37	3	0	.00
2 ⊢ 5	7	.65	14	2	.03
5 ⊢ 10	9	.83	59	6	.13
10 ⊢ 25	85	7.84	1 701	20	3.71
25 ⊢ 50	703	64.85	18 028	25	39.36
50 ⊢ 100	204	18.82	11 993	58	26.18
100 ⊢ 200	56	5.17	6 943	123	15.16
200 ⊢ 500	13	1.20	3 682	283	8.04
500 ⊢ 1 000	2	.18	1 446	723	3.16
1 000 ⊢ 2 000	1	.09	1 936	1 936	4.23
2 000 ⊢ 5 000	0	.00	0	0	.00
5 000 ⊢ 10 000	0	.00	0	0	.00
≥ 10 000	0	.00	0	0	.00
Total	1 804	100.00	45 805	42	100.00

Source: INCRA.

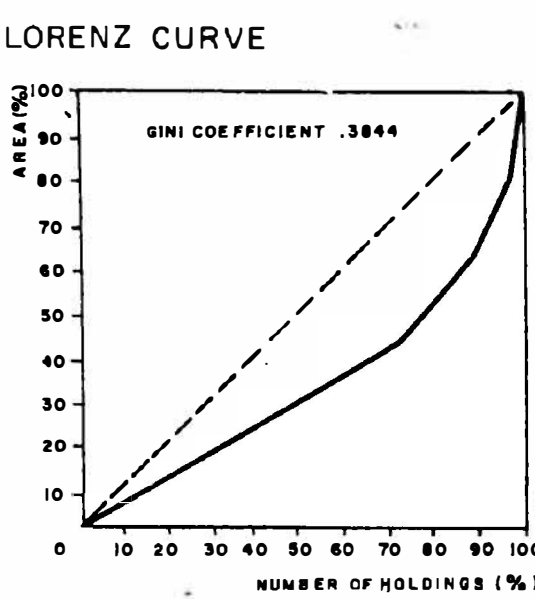
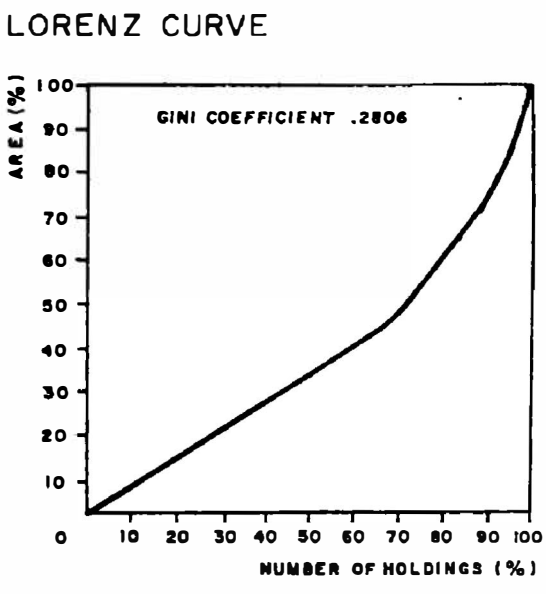


Table 5

Rural Holdings by Size Class: Município *of Santa Maria do Pará, Pará, 1976*

Size Class (Hectares)	Number of Holdings		Area in Size Class (Hectares)		
	Total	%	Total	Average per Holding	%
< 2	0	.00	0	0	.00
2 ⊢ 5	7	.96	17	2	.06
5 ⊢ 10	4	.55	23	5	.09
10 ⊢ 25	20	2.74	373	18	1.40
25 ⊢ 50	503	69.00	12 758	25	47.89
50 ⊢ 100	160	21.95	8 995	56	33.76
100 ⊢ 200	32	4.39	3 740	116	14.04
200 ⊢ 500	3	.41	734	244	2.76
500 ⊢ 1 000	0	.00	0	0	.00
1 000 ⊢ 2 000	0	.00	0	0	.00
2 000 ⊢ 5 000	0	.00	0	0	.00
5 000 ⊢ 10 000	0	.00	0	0	.00
≥ 10 000	0	.00	0	0	.00
Total	729	100.00	26.640	36	100.00

Source: INCRA.



A *município* falls into the highly concentrated class if 50% of the land occupied belongs to holdings having one thousand or more hectares and to less than 20% of the property-owners. Holdings of over 10 thousand hectares exist, and the average is 400 hectares (table 1). In the concentrated class are those *municípios* in which holdings of more than one thousand hectares account for 30 to 50% of the area settled, and the average size ranges from 100 to 400 hectares (table 2). The fragmented class includes *municípios* where 75% of the landowners retain 30 to 50% of the area occupied in the form of plots having 10 to 100 hectares. Less than 20% of the land belongs to holdings of over one thousand hectares, and the average size is 50 to 75 hectares (table 3). In the highly fragmented class, over 50% of the land settled is distributed among more than 85% of the property-holders in plots of 10 to 100 hectares. Less than 10% forms part of holdings with more than one thousand hectares, and the average size is approximately 50 hectares (tables 4 and 5).

The differences in the land-distribution patterns reveal that unplanned settlement favors the latifundio, whereas planned colonization is partial to the minifundio. Nearly all the *municípios* in Goiás and Maranhão, where unplanned settlement dominated, are in the highly concentrated or concentrated classes. Moreover, there is a marked trend toward even larger holdings not only in the *municípios* characterized by latifundia traditionally engaged in livestock activities (which comprise the majority), but also in those that were practically unsettled until the sizable in-migrations of the last two decades. In contrast, most of the *municípios* in Bragantina (Pará), as well as the *município* of Ceres (Goiás), belong to the fragmented or highly fragmented classes. Significantly, these were settled under government programs.

Despite the fact that the settlement of Bragantina goes back a century, the land-distribution pattern of the region has changed little; and unlike the trend observed in Goiás and Maranhão, until recently the tendency has been to divide properties among heirs rather than to group them into larger holdings. In only two *municípios* — Bragança and São Miguel do Guamá, whose agrarian structures date from colonial times and are therefore the outcome of the *sesmarias* — are holdings of a thousand, or even 500, hectares common.

The land-distribution patterns along the Belém—Brasília continue to reflect historico-institutional forces. Small family farms are still the norm in areas settled under colonization schemes, and large holdings still the standard in regions occupied in colonial times or recently without specific orientation.

3 — Settlement and employment

Where planned colonization has prevailed, the small plots (25-50 hectares) have proven able to support the population only up to the second or third generation, when the working-age youth move to the urban centers. Plots of this size can do no more than offer new families a subsistence living. Knowing that large families are the rule in the rural environment (over five children per family in our sample), and assuming that traditional techniques continue to be used, the plot becomes too small in the space of a generation.

The *municipio* of Ceres, which corresponds to an agricultural colony established in 1941 in the midst of near virgin forestland, had 29.5 thousand inhabitants in 1950 and 42.8 thousand in 1960. However, by 1970 the figure had dropped to 39.5 thousand, with the overall decrease being 7.8% and the farm-population decrease 21.4% during the intercensal period. With regard to Bragantina, the 1940 census points to such a strong dwindling of the population that the reliability of the information is placed in doubt. While there was certainly marked rural out-migration, return migration to the Northeast must also be taken into account. Between 1960 and 1970, the rural population grew only 15.6%. Since the natural growth rate was considerably higher, net rural out-migration was clearly significant. At least half of the microregional growth favored the urban center of Castanhal, whose population increased 152% over this 10-year period. The primary sector continued to engage the same number of workers in absolute terms, but its share of total employment diminished roughly 10% despite the fact that relatively few new urban job opportunities were created. This indicates that the land is simply unable to support population increases under existing technological conditions, and that many therefore abandon farming. This conclusion is confirmed by a questionnaire distributed in 1976 in the rural and suburban areas of the *municipio* of Castanhal. Of the 83 persons interviewed, 45 had first worked in agriculture as landowners, squatters or hired hands. Of these, in turn, only 32 had remained in the agricultural sector. Since four had shifted from other activities to farming, the net sectoral outflow was around 20%.

It must be stressed that most of the workers in the region merely survive at an extremely low economic level. Despite government goals, the standard of living has not been raised, and sufficient health care, education and technical training have not been provided. Moreover, in both Ceres and Bragantina the plots have proven

too small to hold the population to the land. Relevantly, the family units along the Transamazon highway have been raised to 100 hectares.

Where unplanned settlement was the norm, the rural exodus has resulted from the concentration of land and the expansion of latifundia. The intensive in-migration of the late fifties and early sixties has given way to out-migration. Between 1960 and 1970, the populations of Imperatriz and the lower Araguaia more than doubled, while those of far northern Goiás and the mid-Tocantins—Araguaia increased by just over 60%. Initially, free access to previously uncultivated land allowed these migrants to maintain subsistence farms. However, the lack of effective guarantees soon led to these farms being overtaken by latifundia.

The *municipio* of Imperatriz, for example, was essentially comprised of uninhabited virgin forest until the opening of the Belém—Brasília highway. In 1972, the private property registered by INCRA[1] totaled 1.4 million hectares — more than the actual area of the *municipio* (13.4 thousand square kilometers). Of the 3.3 thousand holdings listed, the 23 that had five thousand or more hectares accounted for 488.6 thousand hectares and generally belonged to individuals from outside the region. Although the deeds had still not been recorded at the time, these holdings are closed to third parties. In the absence of adequate measures, such concentration is sometimes effected against the law and even against the Constitution. For example, by means of false deeds, areas of over three thousand hectares are appropriated by foreigners without the required Senate approval. Another practice is for a single owner to register a series of holdings of just under three thousand hectares each in the names of other family members.

Besides making the land inaccessible to others, the process of concentration frequently leads to systematic eviction of the previous occupants, in most cases the original settlers. The only way to avoid the resulting conflicts, insecurity and deaths would be to enact and immediately enforce preventive measures. Although these events are public knowledge, it is difficult to attenuate the situation to the extent that the policies are influenced by those who would be prejudiced by changes in the current practice. As long as 15 years ago, SPVEA[2] alerted policymakers to the need for guidelines and control measures. Thus, the above mentioned occurrences cannot be imputed to surprise or want of foresight, but to lack of planning.

[1] Instituto Nacional de Colonização e Reforma Agrária (INCRA).

[2] Superintendência para a Valorização Econômica da Amazônia (SPVEA).

Whether violent or pacific, the expulsion of small-scale farmers creates problems for them and for the region. Those forced to move either (1) migrate to other rural areas, (2) seek wage work on large holdings, generally as contracted hands, or (3) move to urban centers in the same or other regions.

Rural-rural migration is only a solution when there is a chance for permanent settlement at the new destination. Otherwise it leads to chronic moves along migratory routes such as the one from the Itapecuru, Mearim and Pindaré valleys to the Tocantins valley (Imperatriz and far northern Goiás) and hence to the left bank of the Araguaia and the Transamazon highway. Moreover, the current settlers are often displaced by the new arrivals. For all concerned, migration means loss of land, abandonment of shelter, sale of livestock, payment of transport, and a period without income.

For the hired hand, wage work on a cattle ranch signifies exploitation and an insecure future. He is only needed during the year or two that it takes to clear the land and plant rice, corn and hay. Once the artificial pasture has been formed, the ranch requires no more than two or three herdsmen per thousand hectares, so the migrant is pushed his nomadic way. The agricultural worker employed on a crop farm, a pepper plantation, or an experimental rubber plantation in Bragantina is in a better position. There is a possibility of his staying in the same place and of familiarizing himself with crops and methods that he may eventually turn to his own account, as observed among the pepper workers in the area of Tomé-Açu. This alternative would be even more attractive were wages to correspond to the high profits derived from these undertakings.

Migration to a medium-size or large urban center is often the only way to escape a nomadic life and obtain the minimum advantages of modernization, such as schooling, medical care and public transportation. This explains the swelling of the towns along the highways. The urban population of Imperatriz, for example, climbed from nine thousand inhabitants in 1960 to 34.7 thousand in 1970, and this figure has now more than doubled. Over the same period, the population of Araguaína rose from 2.4 to 17.5 thousand, that of Gurupi from 12.0 to 24.6 thousand, and that of Castanhal from 10.2 to 25.8 thousand inhabitants. Throughout the region, urban population growth has far surpassed total population growth. It is questionable, however, whether these numerous new arrivals are truly "urban" residents. They live in the outskirts where infrastructure is minimal, social services limited, and job opportuni-

ties insufficient (except for a few towns in Bragantina, mainly Castanhal and Capanema, where industry is better prepared to receive those expelled from the agricultural sector).

In the locational decisions of firms, the financial factor generally outweighs the raw-material and labor factors. The raw material receives minimum processing in the region. Most of the rice grown along the Belém—Brasília, for instance, is exported unhulled to Anápolis and the Center-South; timber is simply cut and shipped in the form of logs; part of the mallows raised in Bragantina leave the region completely uncured. There are exceptions. The location of the cement plant in Capanema was determined by the proximity of limestone deposits, and the meat-processing plant in Araguaína makes use of local manpower. Since the rice-processing facilities in Gurupi utilize capital-intensive techniques, relatively few job opportunities exist; in order to benefit from tax exemption and massive SUDAM fiscal incentives (Cr$ 86.5 million), the firm had to be located in Amazônia Legal, but the owners chose the southernmost site, nearest to Goiânia, where they reside and where the administrative activities are conducted.

The process of urbanization opens new employment opportunities, but due to lack of skills and formal education, the rural migrant is barred entry. Hence, the administrators, bankers, and members of the liberal professions come from the capitals, and their assistants from among youth with high-school educations. Typically urban commercial establishments (supermarkets, department stores, appliance and furniture shops) are staffed by clerks who have completed their primary schooling. Only public works (federal, state or municipal) simultaneously provide opportunities for skilled, semi-skilled and unskilled labor. Unfortunately, even governments are apt to hire outside manpower for the construction of roads, urban infra-structure, and public buildings. This is the case of the private firm contracted to maintain the Belém—Brasília highway; since its headquarters are located outside the region, it gives preference to workers from other areas.

The outcome is a low-wage, low-productivity craft and service sector comprised of self-employed workers with neither fixed capital nor financial resources. Even when they earn wages, they do not enjoy social-security benefits because they do not have signed working papers. They find their customers mainly among their own in the suburbs and rural areas, and secondarily among the rising urban middle class until it organizes higher-quality capitalized service establishments. This transitional phase having been passed, only household servants continue to find regular work in the urban centers proper. Besides personal services, these low-wage tertiary workers

are engaged in residential construction (masons, plasterers, painters, carpenters, cabinetmakers), crafts (clothing and furnituremaking), or small-scale merchandising (stalls that sell food staples, kerosene, items for personal hygiene).

This is illustrated by the answers to the questionnaires distributed to 600 heads-of-household in Castanhal, Imperatriz and Araguaína. Three hundred and twenty-four first worked in agriculture: 73 as landowners, 190 as squatters and 61 as hired hands, with the proportion of landowners being higher in Castanhal than in the other *municípios*. At present, only 229 are engaged in agriculture — 204 of the original 324 plus 25 who were initially employed in other activities. Thus, 120 persons left the sector while only 25 entered. This represents an overall net loss of 37%, distributed as follows: 29% in Castanhal, 40% in Imperatriz and 35% in Araguaína. The percentages are high, and indications are that the trend is rising.

The present occupations of those pushed out of agriculture are revealing. Fifty were engaged in commercial activities: 41 stallkeepers, five traveling salesmen, and one canteen owner (only three have achieved the higher status of operators of medium-size establishments). Sixteen entered construction as masons, carpenters or assistants. Five render personal services: a cook, a washerwoman, a guard, a shoemaker and a tailor. Five were chauffeurs. Only three of those interviewed were specialized and perhaps employed in industry: a lathe operator, a mechanic and a machinist. At the time of the survey, eight were unemployed and four were retired.[3]

Thus, the majority of these onetime agricultural workers transferred from specialized (albeit low-productivity) activities to unskilled employment in an artificial, already saturated service sector. Under these circumstances, rural-urban migration benefits neither the individual nor society at large. The solution may lie in strengthening the agricultural sector technically and economically so as to discourage rural out-migration and balance the development of the urban tertiary sector in the settlement regions. At the same time, industrialization via small and medium-size firms should be fostered.

It is important to emphasize that the employment situation described applies not only to those who were originally farmers, but to all who were interviewed. Only time will tell whether the

[3] In addition to requiring no specific skills, the majority of the occupations listed allow the individual to retain his independence and are in accord with the poor health common among those who have worked on farms for 20 to 30 years (malaria, rheumatism, physical exhaustion).

phase is passing and at the sacrifice of a single generation, or a lasting structural feature likely to affect the professional opportunities of the children of the migrant settlers.

4 — The feasibility of small-scale agriculture

Closely linked to the social questions discussed above are the economic problems of productivity, profitability and soil protection. In the Amazon region, the productivity of farming and livestock activities is known to be very low at both the micro- and macro-economic levels, for the individual producer and the sector as a whole. As already pointed out until recently, most of the holdings in the area crossed by the Belém—Brasília highway were either minifundia (subsistence firms) or latifundia (cattle ranches), and nonproductive in either case. Over the last few years, regional agricultural policy has therefore aimed to promote medium-size undertakings. This has been the thrust of the fiscal-incentive programs.

There are, however, circumstances under which the small-scale farm is feasible. This is proven by the success of the Japanese immigrants, and of those who followed their steps, in planting pepper. Another example is provided by the truck farms located around the large urban centers. Gurupi, a small private settlement in the *municipio* of Imperatriz, manages to send one thousand crates of tomatoes to Belém per week. One of the factors that makes pepper highly profitable is the price it commands on the international market. In contrast, subsistence items are sold on the domestic market at prices that do not even keep pace with inflation. As a farmer from Bragantina observed, one trades a sack of manioc flour for a kilo of coffee (or three kilos of pepper).

The relative profitability of different crops is usually calculated on the basis of terms of trade expressed in prices. However, this micro-economic indicator does not reflect the social importance of individual products, and, as economic history attests, prices can be manipulated. Paradoxically, the unprofitability of subsistence crops is due to their forming the dietary basis of the rural population and the greater part of the urban population. Manioc, rice and fruit in Belém — corn, rice and fruit in Goiás — all are retailed at low prices, with the producers receiving very little and only the more important intermediaries enjoying reasonable financial returns. In turn, the low prices of subsistence items exert downward pressure on wages in industry, commerce, administration and other services.

The rising importance of small-scale agriculture, at the regional as well as the national level, is indicated by the data gathered by Ribeiro. Whereas the share of Maranhão and Goiás in domestic rice production has increased, that of the states in the South and Center-South has decreased. At the same time, the quantity grown n the Northeast, with the exception of Bahia, has diminished. The share of Maranhão in domestic rice output rose from 5.7% in 1960 to 8.9% in 1970, despite the fact that the state population dropped from 3.5 to 3.3% of the national population. Goiás, which produced 15.1% of the total in 1960, answered for 16.1% by the end of the decade. Significantly, it is the colonization areas, specifically the small farms, that are responsible for this growth in output. The owners of the larger holdings generally prefer cattle ranching, which is more financially attractive. Over the same period, the contribution of Minas Gerais to total rice production fell from 19.6 to 15.4%. and that of São Paulo (where the population grew twice as fast as rice production) from 19.2 to 14.0%. The farmers of Rio Grande do Sul and Paraná have increasingly come to prefer soy and wheat, the former because of favorable international prices and the latter owing to government subsidies aimed at reducing wheat imports and stemming foreign-exchange outflows. From a macro-economic standpoint, the importance of subsistence crops is obvious, especially in an economy that is developing via accelerated industrialization.

Thus, until it is proven that entrepreneurs are truly interested in cultivating basic crops in the colonization areas, it might be best to support small-scale agriculture for two reasons. Firstly, what would happen if the increase in food production failed to accompany demographic growth during a world food crisis? Secondly, small-scale farmers and their families comprise an important part of the Brazilian population, and from 60 to 80% in the settlement regions (567 thousand persons in 1970 in the five areas studied).

An argument against the minifundia, besides their low productivity, is that they are responsible for the soil depletion so detrimental to agricultural activities. Bragantina is often cited in this respect. However, it is not the slash-and-burn method per se that is prejudicial, but its being used in an area where the population pressure makes it unfeasible. This brings us back to the size of the plots, and calls for the reminder that the adequacy of the plot depends on the crop. Even in Bragantina, where traditional crops are considered unviable, it is profitable to plant pepper on plots the same size as those distributed when the area was first settled.

Economic feasibility and environmental protection therefore depend not on a single factor, but on a series of interrelated factors — the crop itself, the technology used, the capital available.

In the Amazon, the liability for soil and resource depletion ultimately rests with the outside interests that make speculative investments rather than with the small-scale farmers. These interests could well afford to conduct agricultural research, but are hardly likely to do so given its delayed results and low returns relative to industrial research. Thus, the ranchers in Paragominas, one of the *municípios* in Pará most benefited by SUDAM, comment that local livestock activities are a failure and that the pastures do not protect the soil — this in a region that was virgin forest 20 years ago.

The importance of agricultural research is stressed by Alvim, who points to the need for methods proper to tropical ecosystems (and therefore free of excessive dependence on fertilizers and other inputs) and the manufacture of equipment designed to overcome regional problems. He also suggests that studies be made concerning the cultivation of certain native species adapted to the poor soils of the region: Brazil-nuts, *guaraná*, palms, *pupunha*, etc. At this point, the success of the Japanese colony in Bragantina should be recalled, together with the fact that Japan has a research and training tradition. In Brazil, likewise, the necessary steps will have to be taken by the government. Certain federal [4] and state agencies have moved in the right direction, but the measures adopted so far have been too weak to offset the systematic exploitation that goes back more than a century, or to establish an agricultural sector capable of holding the population to the land. Among the instruments used to attend to these goals are fiscal incentives and credit, to which we shall now turn our discussion.

5 — Fiscal incentives and agricultural credit

Any appraisal of the impact of fiscal incentives on farming and livestock activities along the Belém—Brasília highway is necessarily discouraging. To begin with, the projects presented to SUDAM for approval tend to be poorly planned and to lack the information required for evaluation in legal and practical terms. Although the stated aim is often cattle ranching, the actual end is usually timbering or land speculation. This may explain the decadence of the ranches in Paragominas and the disinterest of their owners.

[4] Including IPEAN (Instituto de Pesquisa Agropecuária do Norte), now EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária), and INPA (Instituto Nacional de Pesquisas da Amazônia)

By way of example, one project approved in 1972 was to create 21 job opportunities — including only one for a herdsman — in an area of over seven thousand hectares. Although there are data available on 26 projects covering 619.7 thousand hectares in the area studied, it was impossible to construct precise indices of their economic and social profitability due to the above mentioned vagueness of purpose. Nevertheless, the figures are suggestive. Eleven projects representing a total of 216.7 thousand hectares were to create 275 jobs for managers, skilled workers and manual laborers, that is, one job per 788 hectares. As to the expected size of the herd (usually larger than the real size), 16 projects were to place 204.1 thousand head of cattle on 485.8 thousand hectares, for an average 2.4 head per hectare (artificial pastures can support one head, and natural pastures .2 head, per hectare per year). The ratio of workers to head of cattle is therefore 1:331, which is more or less what is observed in practice. While other jobs are created in the clearing and pasture-formation phases, these have the disadvantage of being seasonal and short-term. It should also be remembered that, besides providing little stable employment for the local or regional population, the large ranches generally evict the squatters who previously tilled the land.

The speculative nature of certain projects is further revealed by the large number of investors with no farming or ranching background (engineers, lawyers, doctors, civil servants), as well as by the frequent and repeated sale of holdings.

While it is common to attribute these distortions to SUDAM [5] and IBDF [6] bureaucracy, the real issue is not red tape but the fitness of the policies that these agencies are in charge of carrying out. In other words, is the offering of subsidies and incentives to large enterprises (especially those from outside the region) the best way to promote the socio-economic development of the Amazon? The data indicate that the undertakings financed have introduced little capital and technology and have created few job opportunities. Moreover, their capital is still comprised mainly of land, which has been appreciated principally through speculation. The incentives, instead of being used to create one new job per 788 hectares, could be directed to settling, in an equivalent area, eight farm-operator families employing traditional crop rotation, or 15 to 20 families utilizing methods based on modern inputs. In addition, technical assistance to local farmers could be provided.

[5] SUDAM (Superintendência do Desenvolvimento da Amazônia) is the regional development agency.

[6] Instituto Brasileiro de Desenvolvimento Florestal (IBDF).

There are various credit programs in effect in the Amazon. The most important is PROTERRA,[7] established in 1971 and responsible for investing an initial Cr$ 4.0 billion. In this region, the financial agents of the program are the Banco da Amazônia S. A. (BASA), the Caixa Econômica Federal, and the Banco do Brasil (the principal institution for carrying out the financial policies related to the agricultural sector throughout the country). Although private entities can act as intermediaries between the banks and the clients, it is usually EMATER (ex-ABCAR) [8] that performs this role, accompanying the projects from the planning stage and supervising the use of funds.

PROTERRA offers credit on exceptionally attractive terms — so attractive that many investors prefer this program to the SUDAM fiscal incentives. At 7% per year, interest rates are practically negative, and some investments (e.g. in techniques to raise productivity) are interest-free. The loans can be repaid in five to 12 years, with a two-to-three year grace period. Credit can be obtained for ends as diverse as the initial purchase of land, the enlargement of small farms, the acquisition of heavy machinery, and the modernization of rural holdings. Although the maximum is 15 thousand times the highest monthly minimum wage in the country, the needs of the small and medium-scale farmer are also foreseen. Thus, collateral is required only when the loan is over 50 times the highest minimum wage, and repayment commences only after the activity has begun to yield returns. Unfortunately, few small and medium-scale producers are benefited in practice. This is confirmed by the data relating to two centers located on the Belém—Brasília highway.

In the first (table 6), information was gathered on 49 cattle-ranching projects which had been granted financing valued at Cr$ 18.0 million by February 1975. Only eight of the holdings had less than 100 hectares, and together accounted for only 2% of the credit. Twenty-five were properties of more than 500 hectares; of these, 15 had areas of over one thousand hectares and monopolized 87% of the credit, worth Cr$ 15.7 million and averaging more than Cr$ 1.0 million per holding. This amount was requested mainly for the purchase of cattle (15.9%) and construction (39.2%). It is common to use loans for construction since this is one way of raising the value of land.

7 Programa de Redistribuição de Terras e de Estímulos à Agroindústria do Norte e do Nordeste (PROTERRA).

8 Empresa Brasileira de Assistência Técnica e Expansão Rural (EMATER); Associação Brasileira de Crédito e Assistência Rural (ABCAR).

The data collected in the second center (tables 7 and 8) allowed for analysis of the distribution of PROTERRA funds by the Banco do Brasil up to early 1975, as well as identification of the groups benefited. The list covers 85 farming projects and 63 cattle-ranching projects, which received, respectively, Cr$ 2.3 million and Cr$ 8.4 million. The funds granted to cattle ranchers were for the formation of 4.7 thousand hectares of pastures, the purchase of 3.5 thousand head of cattle, two pieces of hydraulic equipment and various other types of machinery and equipment, and the construction of 25 kilometers of roads, 155 kilometers of fencing, 26 houses, 17 corrals, 146 troughs, 28 gates, and six dams. Clearly, an appreciable part of the financing is directed to fixed investments.

It was possible to examine 65 of the farming projects and 56 of the ranching projects (76% and 90% of the respective totals) in greater detail. Of the borrowers, 14 farmers and three stockmen worked plots of less than 100 hectares. Seven of the 14 farmers were members of a cooperative organized by the private settlement in which they lived; six received loans for planting tomatoes, a crop that was being promoted in the settlement at the time. The credit granted to landowners in the 100-or-less hectare class corresponded to only 6% of the total conceded for farming and 1% of the total for ranching. In the 500-and-over hectare category, there were 49

Table 6

Sample 1: Distribution of Credit to Farming and Livestock Activities by Size Class, February 1975

Size Class (Hectares)	Number of Projects	Total Area of Holdings in Class (Hectares)	Amount of Credit (Cr$)		
			Total	Average per Project	Average per Hectare
0 |— 99	8	493	362 076	45 259	734
100 |— 299	10	1 829	671 285	67 128	367
300 |— 499	6	2 072	428 905	71 484	207
500 |— 999	6	4 061	876 750	146 125	215
≥ 1 000	19	117 940	15 695 320	826 069	133
Total	49	126 395	18 034 336	368 047	142

Source: NAEA/UFPa, case study of unplanned settlement along the Belém-Brasília highway, 1975

projects, but these concerned only 36 producers, since 13 simultaneously benefited from farming and ranching credit. Moreover, half went to the 15 cattlemen with holdings of one thousand or more hectares. One-third of all the loans granted by the local branch of the Banco do Brasil went to 18 producers who were members of four families and owners of 25 thousand hectares in the *municipio*. As is to be expected in a competitive system, the concentration of credit parallels the concentration of land, and therefore of income.

While the bargaining power of the dominant groups is in itself sufficient to explain this fact, concentration is also facilitated by the conditions on which credit is made available. BASA, for instance, requires that the applicant present the title deed. In the absence of the deed, the Banco do Brasil, showing a flexibility in accord with the actual situation in the colonization areas, will accept a letter of approval from INCRA. However, few such letters are submitted to the bank due to the fact that they are given only following an investigation which confirms possession of the land and lack of conflict — a prerequisite that is rarely verified.

Another factor that makes it difficult for producers without capital to take advantage of credit facilities is the condition on which the parts of the loans are released. Each part is granted only

Table 7

Sample 2: Distribution of Credit to Farming Activities by Size Class, 1974-1975

Size Class (Hectares)	Number of Projects	Total Area Cultivated in Class (Hectares)	Amount of Credit (Cr$)		
			Total	Average per Project	Average per Hectare Cultivated
0 ⊢ 99	8	163	74 895	9 302	459
100 ⊢ 299	16	713	221 684	14 779	311
300 ⊢ 499	9	347	185 880	20 654	535
500 ⊢ 999	15	1 335	640 580	42 705	479
≥ 1 000	11	1 444	685 560	62 323	474
No Information	7	302	125 800	17 971	416
Total	65	4 304	1 934 399	29 760	449

Source: Same as table 6.

after it has been ascertained that the previous part has been properly used. This presupposes regular checks, but in the colonization areas the distances and difficulty of access, especially during the rainy season, plus the shortage of inspectors, often hinder such checks. The resulting delays in the release of funds are particularly detrimental to the small-scale farmers who, unlike the ranchers who can sometimes offer even air transport to the inspectors, have no way to speed up the process.

The stronger selective mechanisms are, however, those that administratively link the interest of the financial intermediaries to those of the wealthiest clients. The intermediaries receive a set rate relative to the value of the loans: 2% for technical assistance, and 4 to 5% to cover costs and risks. Since costs rise with the number but only slightly with the size of projects, intermediaries give preference to larger undertakings. As one of the managers of BASA explained, the bank is a commercial as well as a development institution, and thus owes satisfaction to its stockholders. For this reason, it rarely finances farming activities because they incur losses. Or, in the words of an ex-ABCAR agronomist: "We are self-financing. If we plan a small project valued at 10 thousand *cruzeiros*, we receive 200; if we design one worth a million, we charge 20 thousand. Thus, the costs are practically the same." Due to this

Table 8

Sample 2: Distribution of Credit to Livestock Activities by Size Class, December 1974

Size Class (Hectares)	Number of Projects	Total Area in Use in Class (Hectares)	Amount of Credit (Cr$)		
			Total	Average per Project	Average per Hectare in Use
0 ⊢ 99	3	111	88 340	29 443	795
100 ⊢ 299	13	2 502	581 650	44 742	232
300 ⊢ 499	17	6 820	1 596 100	93 888	234
500 ⊢ 999	8	4 807	1 084 042	135 505	225
≥ 1 000	15	31 857	4 143 900	276 260	130
No Information	7	—	943 550	134 793	—
Total	63	46 097	8 437 582	133 931	—

Source: Same as table 6.

profit-maximization approach, the preference (and search) for large projects is even further accentuated when the official intermediaries have to compete with private agents. One way to avoid such distortions would be to oblige intermediaries to allocate a certain percentage of available funds to small projects, another to medium-size undertakings, and another to large enterprises. This might contribute to guaranteeing that loans be used as stipulated, rather than deflected to other ends.

6 — Conclusions

Two years of research among the migrant settlers along the Belém—Brasília highway served to clarify the role and underline the social importance of these peasant farmers. In the area studied, 60 to 80% of the population depends on the primary sector. After the initial phase of colonization via in-migration, this proportion decreases while certain urban centers grow unchecked. The proportion then stabilizes. Thus, even in Bragantina, approximately 65% of the population still depends on the sector. It goes without saying that so appreciable a part of the society cannot be permanently overlooked.

Despite substantial rural-urban migration, the overall employment structure has changed only slightly. Traditional agriculture continues to support the rural population and to meet the greater part of the food requirements of the urban poor at low prices. It thereby allows the rest of the primary sector to engage in activities better able to compete on the national and international markets, and provides industry and urban services with a cheap labor supply. For these reasons, the permanent settlement of small-scale farmers appears, to date, to be the best way to utilize capital in the form of land. Of late, however, large-scale cattle ranching has been fostered despite its offering economic advantages neither in terms of the preservation and rational use of natural resources nor in terms of income and the introduction of modern techniques (with a few exceptions).

Apparently underlying the stimulus to cattle ranching is the assumption that the Amazon is an immense virgin territory that can be settled and economically developed within the desired time only through the offering of incentives to large private undertakings. In following this approach, the policymakers have either been unaware of (or neglected) the volume and social significance of migration, of the agricultural manpower forced to move on by the progress of the latifundia. Part of this population pushes further

on to open new lands that will also eventually be occupied by latifundia. Another part, tired of this unprofitable historical role, move to the nearest urban centers. Here, unable to enter the regular labor market, they find their living in the low-wage service sector.

This marginalization of a substantial portion of the population is an indirect reflection of sectoral and regional national priorities. Since regional interests have been secondary under this approach, it is time that the government adopt measures specifically designed to improve the socio-economic conditions of the Amazon itself. Private outside groups are hardly likely to identify their interests with those of the region. For them the Amazon is an area of the future, a vast reserve in which they strive to guarantee their rights over a still unknown potential. Since their investments are mainly speculative, they feel no pressure to make them truly productive.

It is difficult to set guidelines for regional settlement and development, but given recent experience in the Amazon, those proposed by SPVEA over 15 years ago take on renewed force. Some of the SPVEA suggestions refer to agriculture and permanent settlement. Others have to do with creating opportunities in the urban centers.

With respect to permanent settlement, measures ought to be directed to land distribution and the support of small and medium-size undertakings. The land-distribution policy should guarantee the settler a plot suited to conducting his activities with available technology. He need not be granted ownership, since colonists are often led to sell their land for personal or other reasons. According to SPVEA, it is better to concede renewable long-term or life leases. This system also has the advantage of giving the government permanent control over the land, thereby allowing it to protect natural resources, redistribute plots as technical advances render smaller farms economically feasible, and simplify the legal formalities in cases of abandonment of the land or death of the beneficiaries.

The economic viability of small farms depends on research, experiment and technical assistance, all of which have often been halted due to lack of funds. Hence, direct credit, which is only available to those with sufficient capital and collateral, should be complemented by free services such as surveying, technical assistance and regular follow-up, pest and disease control, and the provision of seeds and seedlings. Some services — health and education, for instance — can be highly profitable in that they raise the productivity of the labor force. In the Amazon, an excellent example is the campaign to eradicate malaria.

Steps also need to be taken to raise the earnings of the small-scale producers of basic food commodities. Prices can be supported either directly or indirectly through increasing the monetary value of the products (e.g. warehousing) or decreasing costs (transportation, middlemen, etc.).

Colonization policies should be complemented by urban employment policies. As pointed out in the SPVEA study, it is essential to establish industries for processing farm and livestock products. These might take the form of large mixed enterprises in order that agro-industrial activities be vertically integrated. In addition, more public works should be undertaken: local roads, urban infrastructure, schools, hospitals and health centers. Even when the construction is in private hands, preference should be given to locally produced material and equipment and locally available manpower, both often disdained by outside entrepreneurs.

References

Alvim, P.T. *Agricultura nos Trópicos Úmidos: Potencialidade e Limitações.* Ilhéus: Centro de Pesquisa do Cacau, 1975.

Falesi, I.C., *et al.* "Sugestões ao Soerguimento Econômico do Setor Primário no Estado do Pará." Belém: Empresa Brasileira de Pesquisa Agropecuária, 1974 (Mimeographed).

Ribeiro, S.W. *Desempenho do Setor Agrícola, Década 1960/70.* Série Estudos para o Planejamento, n.º 6. Brasília: IPEA, 1973.

Silva, D. "Colonização e Zonas Pioneiras." *Anais da Associação dos Geógrafos* 18 (1973).

Superintendência do Desenvolvimento da Amazônia. "Colonização na Amazônia." Belém, n.d. (Mimeographed).

Superintendência para a Valorização Econômica da Amazônia. *Política de Desenvolvimento da Amazônia.* 2 vols. Rio de Janeiro, 1954-1960.

Passenger transportation in metropolitan areas

Josef Barat •

1 — Introduction: general characteristics of private and public transportation systems

Recent urban transportation studies, as well as urban plans in general, point to significant interaction between available passenger services, which explain the degree of mobility of city dwellers and current land-use patterns. In turn, the spatial allocation of urban activities determines which zones will be predominantly residential. As a rule, the settlement and development of these zones is a function of existing public services, especially transportation. By integrating places of work and places of residence within the urban space, transportation contributes to the overall urban economy by allowing for decentralization of the labor market.

In fact, the basic role of transportation is to integrate the urban and metropolitan areas not only in terms of space, but also in terms of urban activities (economic, social, residential and recreational). It also serves to consolidate the markets for the factors of production. Thus, the daily pendular movements to and from work which constitute the majority of trips in metropolitan areas are to a great extent conditioned by land-use patterns. Likewise, since residential choices are based on existing transportation facilities, these pendular movements can influence land use and future metropolitan development.[1]

The modes of metropolitan transport may be grouped into the following systems: (1) mass transportation or rapid transit — metros, commuter railways, streetcars and buses, with the last two being so

• Adviser to the president of the national bank for economic development (Banco Nacional do Desenvolvimento Econômico — BNDE).

1 J. Barat and M. N. Batista, "Transporte Público e Programas Habitacionais," *Pesquisa e Planejamento Econômico* 3 (June 1973): 375-88.

classed when operating at high capacity on preferential lanes or with rights-of-way; (2) public transport or local transit systems to complement the preceding, comprised of buses and minibuses; (3) individual transport, including private cars and taxis. Transportation planning should strive to coordinate and link these systems so that each is used at the stage for which it is best suited, that is, for urban, suburban or intercity trips. However, what is often observed in present-day conurbations is competition between public and private transportation, or even between the different modes of public transport.[2]

An efficient public transportation system offers its users regular integrated services and provides them with alternative schedules for the different stages of urban, suburban and intercity trips within the metropolitan area.

Public transportation systems bring together mass transport (or rapid transit) modes that operate on exclusive tracks or lanes with rights-of-way, and complementary modes (or local transit) that use the regular urban road network to pick up and distribute both rapid-transit users and strictly local passengers.[3]

Since mass transportation systems require heavy investment in the initial phase, they are economically and financially feasible only where a great number of users can be channeled along determined routes or "corridors". When the permanent facilities are used at capacity, they allow for gains in scale and appreciable reductions in the per unit costs of services. Thus, up to a certain limit, the passenger/kilometer cost decreases as the number of users and/or the distance covered increases. This is especially true of urban or suburban rail transport. It should also be recalled that by reaching high average speeds and making few stops, mass transport modes allow for significant flows at short regular intervals. In addition, they offer high standards of comfort and safety.

However, only the metropolitan conurbations — already saturated by road traffic and its accompanying pollution and congestion, and with dense enough populations to guarantee the intra- and inter-urban lines a sufficient number of users — can bear the financial burdens required to construct metros, pre-metros and commuter railways.

[2] J. Barat, "Planejamento de Transportes nas Áreas Metropolitanas," *Revista de Administração Municipal* 19 (September-October 1972): 5-28.

[3] J. W. Dykeman, "Transportation in Cities," *Scientific American*, September 1965, pp. 163-74.

In small and medium-size urban centers, therefore, it may be preferable to use high-capacity bus systems with the vehicles operating at ground level on reserved or right-of-way lanes. The advantage of such systems is that fixed costs are lower, operating procedures are simpler, and the reserved lanes can be used when necessary by other types of vehicles at off-peak hours or on special occasions. Many European and North American cities have, in turn, introduced rapid-transit characteristics into their streetcar systems by raising the passenger capacity of the rolling stock, eliminating stops, and running lines on reserved lanes or even underground as a step toward metros (pre-metro systems).

Rapid transit can be complemented by local buses and minibuses, which present the following advantages in transporting passengers to or from the centers where movement is concentrated: (1) flexibility in serving a large number of origins and destinations; (2) rapid adaptation to changes in the urban road network; (3) lower capital expenditures in comparison to those required by rail modes; (4) better operating conditions for low- and medium-density flows.

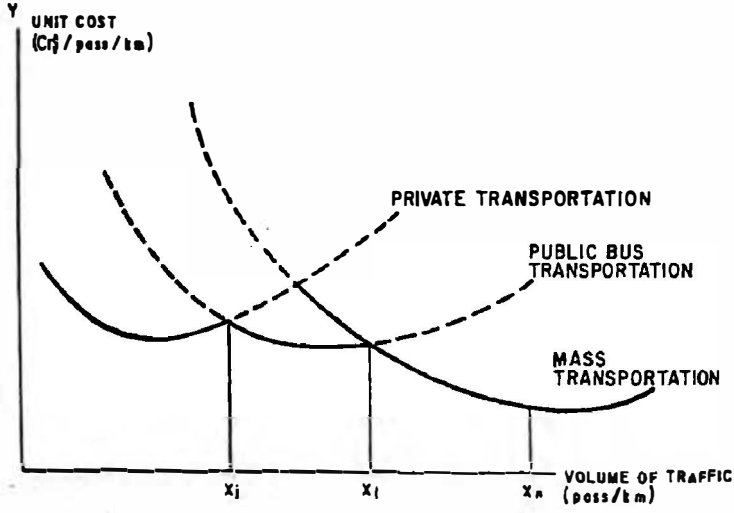
Since the cost of operating buses and minibuses rises with distance and congestion — which ultimately results in more current and potential users — this mode becomes too costly once a certain passenger/kilometer level has been reached. For this reason, it should complement mass transport by serving passengers on less dense, shorter routes. Nonetheless, if the passengers are willing to pay higher fares, it may also be used to link the metropolitan nucleus to the outlying towns.

That the fixed costs of buses and minibuses are lower than those of rail systems does not, however, justify transforming these vehicles, as conventionally operated, into the principal mode of transportation in the major metropolitan areas. As a rule, they move at low speed and irregular schedules on the congested urban roads. This leads to overuse, higher operating and maintenance costs, and, finally, to shorter duration of the vehicles themselves. There is also the social cost of gas fumes and noise levels.

Even so, it must be stressed, once again, that although modern mass transport modes can carry great numbers of passengers at substantially reduced costs, they require massive investments in infrastructure facilities and rolling stock. Hence, even on high-density routes, buses and minibuses should be transformed into a com-

FIGURE 1

COMPARATIVE UNIT COSTS OF METROPOLITAN TRANSPORTATION MODES



plementary mode only after road congestion has raised their operating costs more than proportionately to those of the rapid-transit alternative.[4]

Finally, the mode that presents the greatest divergency between the costs to the user and the cost to society is individual transport. Each person who opts for individual transport may consider it worth bearing the subsequent operating and maintenance costs; but a series of such decisions may impose high costs on all those who have elected this mode and on society as a whole.

The unit costs of the three above mentioned metropolitan transportation modes are compared in figure 1. These costs, measured per passenger/kilometer, include fixed costs related to in-

[4] On high-density routes, hourly capacity can be raised by replacing conventional buses with buses that operate on reserved lanes with rights-of-way. The latter comprise a mass transportation mode well-suited to the relatively scarce finances of less developed countries or regions.

frastructure, terminals, vehicles and other equipment, plus variable maintenance and operating costs. Depreciation and costs to the user in terms of time are also taken into account.

Note that mass transportation, due to high initial fixed costs, is more advantageous than conventional bus transportation only from point X_i up to capacity overload at point X_n. Furthermore, adequate utilization of the scales offered by mass transportation systems (especially rail modes) depends on high traffic densities, on a fair degree of homogeneity in the distribution of trips along the various routes, and on better use of existing facilities whenever replacement costs are within limits that make this economically feasible.[5]

The use of private modes depends, in turn, on the level and distribution of urban income. Even if the standard of living and the distribution of income in a given metropolitan area allow for at least one car per family — which is far from the case in developing countries — the congestion associated with the extensive use of cars makes this mode less attractive than rapid transit, and at times even less enticing than the conventional bus. Due to the negative externalities associated with its consumption in densely populated urban areas, the private automobile may not be considered a "free" service.

In practice, since the various modes coexist, changes do not occur in the absolute manner suggested by the curves in figure 1. Rather, shares of the total demand come to be met more adequately by alternative systems. New systems are constructed, existing capacity expanded, and intermodal possibilities explored when the rising volume of passengers per kilometer so requires.

Having considered the supply of metropolitan transportation services, the demand for such services should be examined. Demand is understood to be the outcome of the spatial distribution of urban activities (work, home, shopping, recreation, school, etc.) and the corresponding need for trips from one place to another. Given the complexity of the "trip" concept, when analyzing demand, several aspects must be investigated: who or what is transported, the purpose of the movement, the mode used, the origin and destination of the trip, and the time at which it is made (whether at peak or off-peak hours). Moreover, the same purposes and/or the same pairs

5 J. R. Meyer, J. F. Kain and M. Wohl, "Urban Transport Cost," in *The Urban Transportation Problem* (Cambridge: Harvard University Press, 1966), pp. 299-306.

of origins and destinations may be served by different types of transportation. This must be borne in mind when coordinating and integrating the several modes so as to render the overall metropolitan transportation system as efficient as possible.[6]

Total demand should be estimated in such a way as to allow for determination of the above mentioned aspects of urban movements. As a rule, field research is conducted to identify origins and destinations and the distribution of trips first at the traffic zone level, and then at the city or metropolitan level. Once the modal split of the flows has been ascertained and the trips have been grouped by purpose, the passenger transportation study is complete.[7] The frequency and direction of trips clearly depends on the characteristics of the specific metropolitan area: its total and active population, income level and distribution, land in use, transportation supply, and traffic density. In general terms, global demand can be expressed by a function having multiple variables:

$$Y = K + A_0 X_0 + A_1 X_1 + A_2 X_2 + \ldots \ldots$$
$$\ldots + A_{n-1} X_{n-1} + A_n X_n$$

where X_0 to X_n are independent variables related to population, socio-economic conditions, distances, etc. For example,

X_0 = total population

X_1 = active population

. . .

X_{n-2} = income per inhabitant

X_{n-1} = total land in use in metropolitan area

X_n = fleet of vehicles

A_0 to A_n are the coefficients of the respective independent variables, with K being a constant that stands for the share of the value of Y not explained by the variables considered.

[6] M. Barbier and F. Mellet, "Determination of Elasticities of Demand for the Various Modes of Urban Passenger Transport," working paper prepared for a meeting at the Economic Research Center of the European Conference of Transportation Ministers, Paris, 1971, pp. 7-8.

[7] Our exclusive concern with passenger movements is not to underrate the importance of freight movements in metropolitan areas. Despite their significance in the coordination of urban and long-distance networks, as well as in the location of warehouses and supply terminals, cargo movements have often been overlooked in transportation studies.

2 — The problem of congestion: the competition for urban road space

When determining how to best utilize existing metropolitan ways, the major problem is congestion, and especially that due to competition between the various surface modes. In transportation economics, traffic congestion is said to occur at the point at which the social costs arising from the use of ways and facilities by additional users exceed the private costs incurred by these same users. Although this definition is applicable to all modes of urban transportation, most of the studies conducted in the developed countries refer to private cars. After all, this is the dominant mode in the urban areas of many of these countries, mainly the United States. Despite the fact that the pattern is somewhat different in the developing countries, it is worth examining how automobiles lead to congestion by way of understanding how they interfere with other surface modes, and particularly with buses.[8]

The trips made by car along a certain section of urban roadway over a given period of time (V) may be considered a function of the density of cars along the section (Z) and of their average speed (W).

$$V = V(W, Z)$$

Assuming the existence of an average speed limit (W_0) and a maximum capacity of the road section in the form of a maximum density reached at peak periods (Z_{m+1}), the traffic flow for a given value of $V = V_m$ will attain its optimum then rapidly fall until the load capacity of the section is reached (V_{m+1}). Thus, the following relationship can be established between the trip flow and the independent variables:

$$V_0 = W_0 \quad Z_0$$
$$V_1 = W_1 \quad Z_1$$

[8] The theory that follows is based on the concepts presented by A. A. Walters, "The Theory and Measurement of Private and Social Cost of Highway Congestion," *Econometrica* 29 (October 1961): 676-99; M. Wohl and B. Martin, *Traffic System Analysis for Engineers and Planners* (New York: McGraw-Hill Book Company, 1967), chap. 10; A. F. Schreiber, P. K. Gatons and R. B. Clemmer, *Economics of Urban Problems, An Introduction* (Boston: Houghton Mifflin Company, 1971), vol. 2.

$$V_{n-h} = W_{n-h} \cdot Z_{n-h}$$

$$V_n = W_n \cdot Z_n$$

$$V_{n+1} = W_{n+1} \cdot Z_{n+1}, \quad \text{where}$$

$$W_0 > W_1 > \ldots > W_{n-h} > \ldots > W_n > \ldots$$

$$\ldots > W_{n+1} \quad \text{and}$$

$$Z_0 < Z_1 < \ldots < Z_{n-h} < \ldots < Z_n < \ldots Z_{n+1}$$

For a given value of $V = V_n$, the trip flow reaches its maximum, then quickly diminishes until the road is overloaded (V_{n+1}) due to the increase in vehicle density and the decrease in average speed. The capacity of the road is therefore defined by $V = V_n$.

In fact, the value of V, which determines the capacity of the road, corresponds to the maximum point of the function $V = V(W, Z)$, since at this point

$$\frac{\delta v}{\delta w} = \frac{\delta Z}{\delta z} = 0, \quad \text{and}$$

$$dv = \frac{\delta v}{\delta w} dw + \frac{\delta v}{\delta z} dz = 0 \quad \text{and}$$

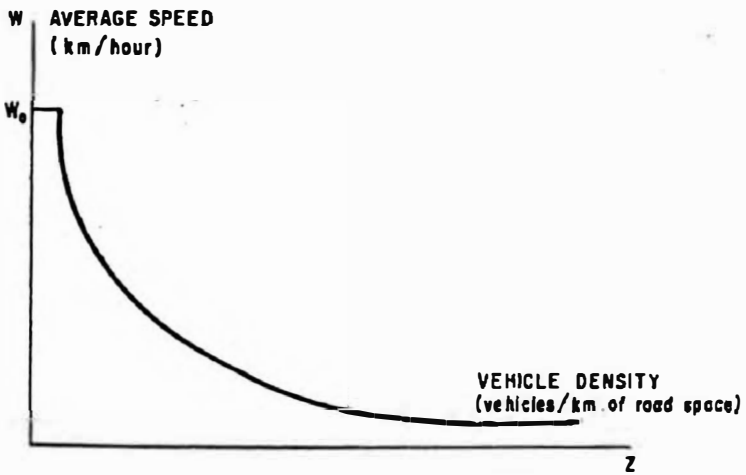
$$d^2 v < 0, \quad \text{since} \quad \frac{\delta^2 v}{\delta_w{}^2} < 0, \frac{\delta^2 v}{\delta_z{}^2} < 0$$

As the vehicle density rises, the average speed falls, tending to zero at congestion when starting from a value established as the speed limit. The relationship between vehicle density and average speed is shown in figure 2, where it is assumed that density adjusts instantaneously due to its being equal along the route under consideration.

Furthermore, as the average speed drops to zero, travel time — which is its reciprocal — tends toward infinity. Assuming that

FIGURE 2

RELATIONSHIP BETWEEN AVERAGE SPEED AND VEHICLE DENSITY



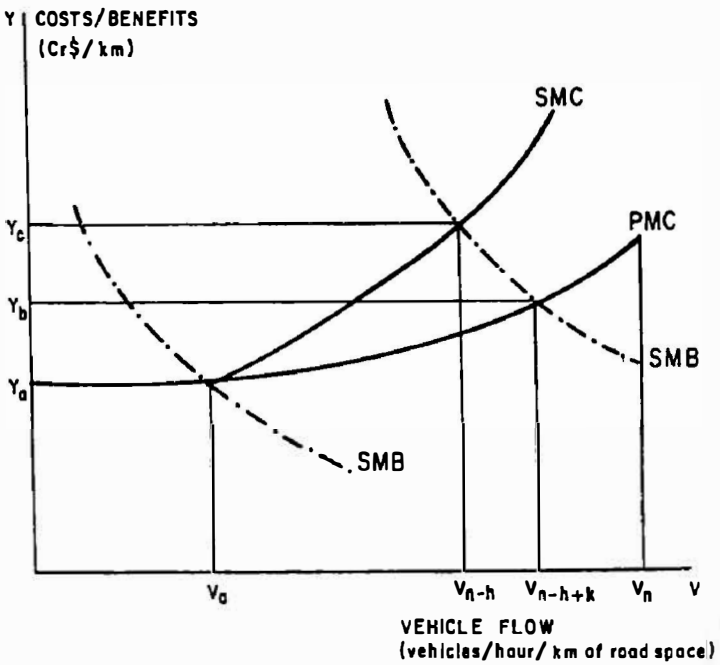
travel costs are directly proportional to travel time, it is possible to determine a relationship between unit costs and trip flow levels, and subsequently establish demand curves for the use of the route that can be used in examining the problem of congestion.[9]

The costs and benefits (private and social) associated with the use of urban roads at different times of the day are depicted in figure 3. The costs, measured in *cruzeiros* per kilometer, cover vehicle operating and maintenance expenses (including depreciation) as well as the value attributed to the travel time of the users. No congestion is observed up to point V_a because there is not yet a difference between private and social marginal costs (*PMC* and *SMC*, respectively), which both remain at level Y_a, and because cars are still able to travel at the speed limit permissible (W_0). As of a frow higher than V_a, however, the additional vehicles on the road slow down traffic. Moreover, since these additional cars oblige all

[9] Decreases in average speeds and increases in travel-time costs give rise to the transportation gradients that determine land-use patterns (imperfections in routes, connections, etc.).

FIGURE 3

RELATIONSHIP BETWEEN COSTS / BENEFITS AND VEHICLE FLOW

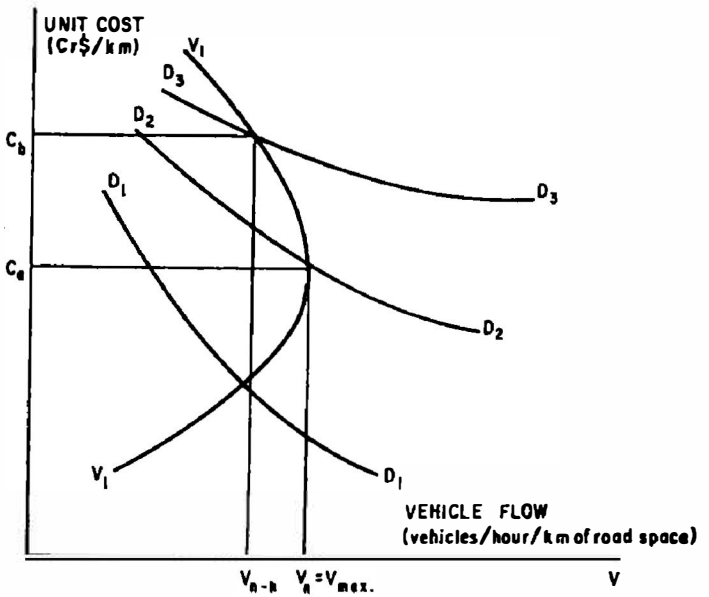


drivers to reduce their speed, the social marginal costs come to exceed the private marginal costs, due to the fact that the former rise more than proportionally to the latter. Even so, the vehicle flow (trips per hour and per kilometer) will continue to increase until capacity V_n. Once this point is reached, however, the number of trips will begin to decline in absolute terms, since, as already seen, lower average speeds result in higher densities.

In the same figure, the total benefits derived by users at two distinct periods are indicated by the curves SMB_1 and SMB_2: (1) the peak hours, when the concentration of pendular home/work/home movements represents the highest social benefits (SMB_1), and (2)

FIGURE 4

RELATIONSHIP BETWEEN HIGHER TRIP DEMAND AND LOWER VEHICLE FLOW

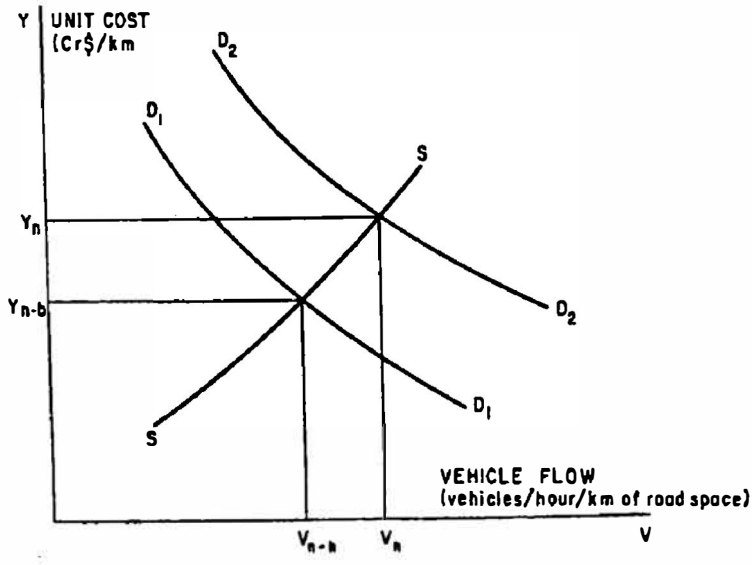


the off-peak hours, when shopping trips, leisure trips, etc. are made by fewer users at lower travel-time costs (SMB_s). Although there is a capacity-utilization level at which the private and social marginal benefits are the same (V_{n-k+k}), these benefits are surpassed by the social marginal costs. This means that the net social benefits are not being maximized — in other words, from the collective point of view the road is not being efficiently used. Efficiency would be reached at point $V = V_{n-k}$; thereafter, the social marginal costs will always exceed the social marginal benefits.

The effect of an increase in trip demand on unit costs (including the value ascribed to travel time) and on vehicle flows is illustrated in figure 4. Considering that lower average speeds mean higher densities, curve VV' may be said to show the equilibrium relationship between unit costs and trip flows (number of vehicles).

FIGURE 5

THE ADJUSTMENT OF TRIP DEMAND TO AVAILABLE ROAD SPACE



Maximum equilibrium will be achieved at $V = V_n$, which defines the capacity of the road, that is, the highest flow level that can be maintained for long periods of time. Since this level has a single demand curve, any increase in trip demand will diminish the flow and consequently raise unit costs $(C_b - C_a)$.

Demand can be adjusted to supply through charges and/or subsidies. In other words, in a market system, demand can be restricted, or the availability of goods or services decreased or increased, by controlling the final price to the consumer. In the case of transportation services, prices below the equilibrium price create excess demand and consequently result in congestion. As indicated in figure 5, for a shift in demand from D_1D_1 to D_2D_2, the vehicle flow will rise from V_{n-b} to V_n despite the higher costs to the users. The willingness of drivers to face congestion means that the price level is still low.

From the standpoint of the community as a whole, the marginal costs and benefits will be aligned only if SMC and SMB_1 are equalized (point V_{n-1} in figure 3). This implies reducing the number of trips through a specific user charge. Such a charge serves to ration the use of the road when necessary (above point V_a) so as to assure its efficient utilization. The user charge should vary depending on the time of the day and the degree of congestion, and should be equal to the difference between the social and private marginal costs.

Nevertheless, practical problems arise when attempting to restrict automobile traffic through user charges. While it is relatively easy to charge tolls on expressways, it is quite difficult to do so on regular urban streets. In addition, if drivers are discouraged from traveling on expressways at peak hours, additional users may seek other routes and cause congestion at other points. Moreover, the congestion at new points may not be sufficient to stimulate users to transfer to public transportation alternatives. For this reason, some developed countries are now studying the possibility of discouraging individual transport by directly subsidizing bus train and metro fares.[10]

The problem of capacity overload — the result of competition between private cars, taxis and buses for existing road space — merits special attention, since it is the major hindrance to the mobility of city dwellers in the metropolitan areas of the developing countries.

For a trip between a given pair of origins and destinations (i, j), there is a demand curve DD, as shown in figure 6. The cost of this trip is indicated at C_1, where the demand curve crosses the supply curve for urban road space S_1S_1.[11] At the cost C_1, the number of trips is represented by V_1. If traffic congestion lowers the supply of available road space to S_2S_2, trips will be reduced in number (to V_2) and made at a higher cost. The increase in travel costs is represented by the area C_1ABC_2, which is shaded in figure 6.

This rise in costs, which means a loss to the community in the form of fewer trips, can be expressed as follows:

$$\Delta C = (C_2 - C_1)\ V_1 + 1/2\ (C_2 - C_1)\ (V_2 - V_1)$$

$$\Delta C = 1/2\ (C_2 - C_1)\ (V_1 + V_2)$$

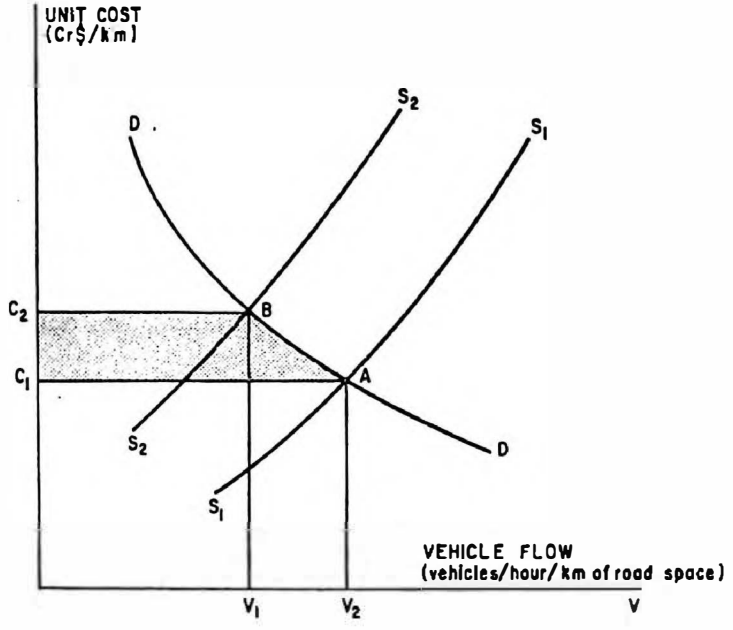
where AB is assumed to be a straight line.

10 See R. Sherman, "Subsidies to Relieve Urban Congestion," *Journal of Transport Economics and Policy* 6, n.º 1 (1972): 22-28.

11 It should be recalled that this cost includes the monetary value attributed to travel time.

FIGURE 8

COST INCREASES DUE TO ROAD SPACE CONSTRAINTS



This expression, which refers to a particular set of origins and destinations (i, j) and a specific route, can be generalized for the overall urban road network:

$$\sum_{i,j} \Delta C = 1/2 \sum_{i,j} (C_2 - C_1)\ (V_1 + V_2)$$

At this point, it is important to stress that congestion leads to heavy costs in terms of travel time not only for the businessmen and professionals who drive their own cars, but, above all, for the majority of the metropolitan population, who are obliged to ride buses. As the roads become overloaded, buses operate at higher costs and lower speeds. The resulting decrease in the number of trips is detrimental to the urban economy as a whole. After all,

though buses require less road space than private cars and taxis, they obviously transport more passengers per vehicle. Thus, each additional automobile raises the social marginal costs not only for other users of individual modes, but also — and more importantly — for users of public transportation.[12] In fact, congestion comprises the outstanding limitation on the improvement of bus services.[13]

Also pertinent is the fact that private and public modes compete mainly during the peak hours when most of the pendular home/work/home movements occur. Since shopping and leisure trips are generally made at off-peak hours and present a far greater variety of origins and destinations and route options, they rarely lead to congestion.

Thus, it is at the peak periods, when traffic is strongly concentrated along the principal routes of access to the central business district, that car owners derive inherent price advantages. In fact, in a situation in which all pay fuel and road-use taxes, the lack of a specific charge on pendular movements in private cars acts as an implicit subsidy to such movements.[14] Moreover, where individual modes have been encouraged and rapid transit alternatives bypassed, car owners have further benefited from an intersectoral transfer, since the transportation infrastructure is built and maintained with funds from the public in general, including those who do not own cars.[15]

It is now time to consider precisely how cars and buses compete for road space in the absence of underground rapid transit or right-of-way mass transportation.[16] In examining the way this com-

12 The feasibility study for the Rio de Janeiro metro revealed that in 1968 only 9% of the road space was utilized by buses, compared to 56% and 37% for private cars and taxis, respectively. Furthermore, whereas each bus transported an average 29.8 passengers, cars carried an average 1.7 and taxis only 1.1.

13 W. L. Henderson and L. C. Ledebur, *Urban Economics: Processes and Problems* (New York: John Wiley and Sons, Inc., 1972). In the industrialized countries, the measures taken to relieve the congestion that hinders the improvement of bus services include: (1) exclusive lanes for buses on expressways and highways, and (2) rights-of-way for buses on the main arteries of metropolitan areas (guaranteed by signals at major intersections).

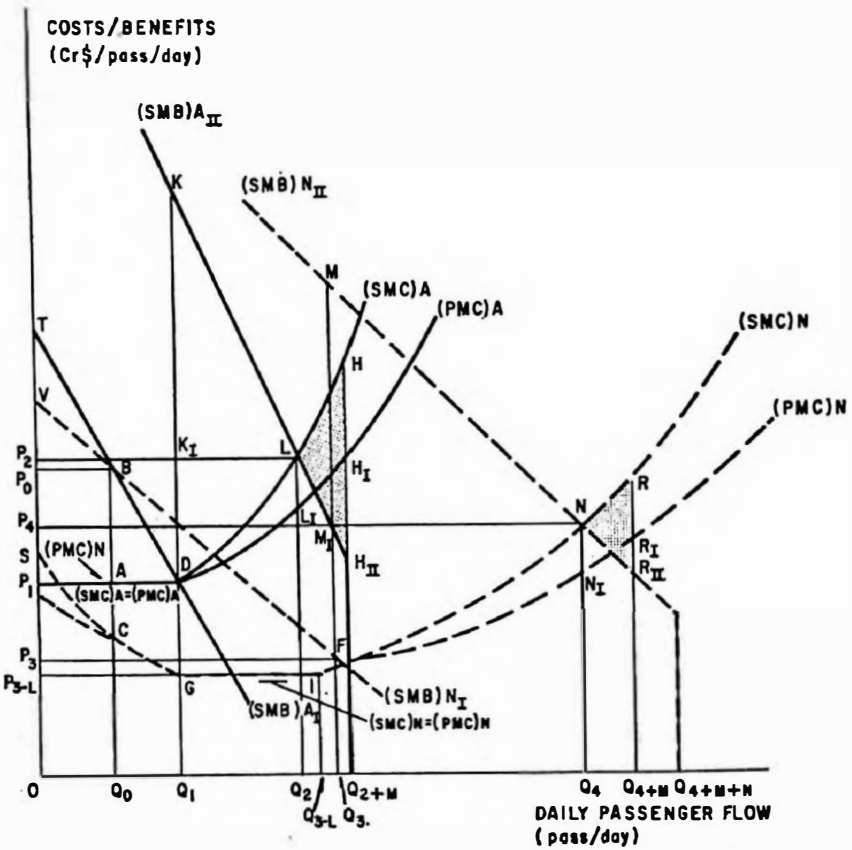
14 D. Netzer, *Economics and Urban Problems: Diagnoses and Prescriptions* (New York: Basic Books, Inc., 1970), pp. 141-44.

15 An example of such transfers is provided by the fact that buses are not allowed to use tunnels, elevated roads or expressways in Rio de Janeiro. See J. Barat, "Transporte e Ecologia," *Revista de Administração Municipal*, n.º 119 (July-August 1975): 19-34.

16 In Brazil, there are only two examples of underground mass transportation systems. Only São Paulo has a metro line in operation, and in Rio de Janeiro the priority line is now being constructed.

FIGURE 7

EQUILIBRIUM SITUATIONS: THE PRIVATE CAR VERSUS THE PUBLIC BUS



petition affects the spatial organization of metropolitan areas, it should be borne in mind that, in developing countries, public transportation has long been neglected as an instrument for fostering rational urban growth and land-use patterns.

Bus franchises are granted in a laissez-faire atmosphere that is hardly conducive to social efficiency, and services come to be provided at private costs (incurred by the owners of the lines and

theoretically covered by passenger fares) that differ from the social costs. The latter take the form of substantial negative externalities: excessive noise, air pollution, inadequate use of road and terminal space. Furthermore, the rising trip demand is met through overutilization of vehicle capacity or through fare and tax devices that are actually disguised subsidies to the franchise holders (high depreciation rates, faulty maintenance inspection, lack of speed-limit enforcement, etc.).

Thus, it is from the standpoint of social efficiency that the equilibrium situations presented in figure 7 should be compared. These situations refer to the meeting of passenger demand by private cars[17] and public buses at peak (pendular movements) and off-peak (non-pendular movements) periods along given sections of principal urban routes on which the two types of vehicles circulate simultaneously. The suppositions that underlie this figure are the same as those implicit in figure 3: (1) as of a certain passenger volume, market imperfections lead to differences between private and social costs; (2) the benefit curves reflect the marginal value of services to their users; (3) prices are in equilibrium when social marginal costs equal social marginal benefits;[18] (4) disequilibria mean socially inefficient provision of services, and should be corrected through government intervention.

Nonetheless, it must be emphasized that, in this analysis, equilibrium situations based on social marginal costs and benefits are meant to serve as no more than a point of reference for designing fare and charge systems aimed at the better use of funds and road infrastructure.[19] Examination of price policy alternatives, or of the merits of policies based on marginal costs, are therefore beyond the scope of this study, especially considering the complexity of the issue when it includes intermodal complementarity and substitution, to say nothing of aspects related to the redistribution of income.[20]

17 For the sake of simplification, taxis are classified as "private cars" despite the fact that they comprise a public transportation mode. It is assumed that they do not contribute to congestion off the main road network.

18 Externalities are generated as of the volume of services consumed when the prices of these services correspond to their value to the community.

19 W. Vickrey, "Some Implications of Marginal Cost Pricing for Public Utilities," *American Economic Review, Supplement* 45, n.º 2 (1955): 605-620.

20 Many English and American writers (including Vickrey, p. 606) contend that charges based on marginal costs to the user of urban roads would be regressive if viewed as taxes. While the argument may be valid in countries where most surface trips are by car, imposing marginal charges on drivers would have progressive effects in Brazil.

In figure 7, the passenger flow (in millions of passengers/day, for example) is represented on the horizontal axis, and the costs or prices (*cruzeiros*/passenger/day) are shown on the vertical axis. With respect to non-pendular movements, equilibrium occurs for cars and buses at points D and F, respectively. Buses obviously carry more passengers than cars ($Q_3 > Q_1$). Curves $(SMB)\,N_1$ and $(SMC)\,N$ intersect at a price level lower than curves $(SMB)\,A_1$ and $(SMC)\,A$; hence, $P_3 < P_1$.

There are certain volumes of passengers for which buses, like cars, have the same private and social marginal costs (between Q_0 and Q_3). This means that returns to scale vary for this mode of public transportation.

In absolute terms, the social value of bus services is higher than that of individual transport,[21] and marginal decreases are less abrupt. Nonetheless, up to flow level Q_0 the private costs of bus services exceed their social costs for two reasons. First, it is not in the interest of franchise holders to operate lines with few passengers. Second, in the absence of congestion, individual transport is more attractive to the users.[22]

Thus, from the standpoint of the community, for non-pendular movements, equilibrium F (buses) is preferable to equilibrium D (cars) because a larger volume of passengers is transported at a lower social cost — which means lower prices for the users. Assuming that buses attend non-pendular demand only as of point Q_0, the net social benefits resulting from their operation will be represented by area $BCGIF$, while those associated with cars will be indicated by the triangle BAD ($S_{BCGIF} > S_{BAD}$).[23] When more nonpendular passengers are transported by bus than by car ($\overline{Q_0Q_3} > 0\overline{Q_1}$),[24] therefore, the former mode offers relatively greater social advantages than the latter ($P_3 < P_1$).[25] Moreover, provided there is no con-

21 By way of simplification, it is assumed that the social marginal benefit curve coincides with the demand curve. This implies that buses can serve more passengers, and that their peak capacity will be reached more slowly.

22 This is the case in the smaller American cities, where buses are left to students and the elderly. Although the same pattern is to an extent observed in the Center-South region of Brazil, the marked socio-economic differentials generally make the bus the principal mode of transportation for those in the lower income brackets.

23 It is obviously being assumed that the car is superior to the bus for satisfying demand $0Q_0$. Indeed, the curve $(SMB)\,A_1$ is higher than $(SMB)\,N_1$ up to point B. In terms of net benefits, the area TP_1AB is larger than $VSCB$.

24 This is, and will long continue to be the case in the developing countries, where the absolute levels and the concentration in the distribution of urban income preclude car ownership on the American or even the European scale.

25 Note that for a volume below $Q_3\,(Q_{3-1})$ and a range Q_3Q_{3-1}, both private and social costs drop to level P_{3-1}. This widens the difference between the costs associated with buses and cars, because $P_1 - (P_{3-1}) < P_1 - P_3$.

gestion on the main road network, the circulation of cars will not restrict the mobility of bus passengers, and the mere difference in private marginal costs will assure rational utilization of the principal routes.

It is for pendular movements, however, that buses offer truly marked social advantages. The same type of analysis made for points D and F can now be made for L and N, which represent the equilibrium situations for cars and buses, respectively, in the transportation of pendular passengers at peak periods. It should be noted from the outset that the position at which curve $(SMB)N_{II}$ crosses $(SMC)N$, compared to that at which curve $(SMB)A_{II}$ crosses $(SMC)A$, underlines the above mentioned disparity between the social marginal costs of using cars and buses for flows from Q_1 to Q_3, since $P_2 - P_4 > P_1 - P_3$.

Assuming that the principal routes become congested at a daily passenger flow Q_1Q_2 for cars and at Q_3Q_4 for buses, the net benefits of the two modes of transportation will be indicated by the areas $KK'DL$ and $MM'FN$, respectively, with the latter being greater.[26]

Furthermore, since the difference between private and social marginal costs is greater for cars than for buses, and since the social value of the services provided by the former diminishes more quickly than that of the services rendered by the latter, an increase m (beyond equilibrium Q_2 and Q_3) in the volume of passengers will involve higher social losses if these additional users are transported by car than if they are conveyed by bus.[27] Hence, the shaded area LHH'' is larger than the shaded area NRR''.

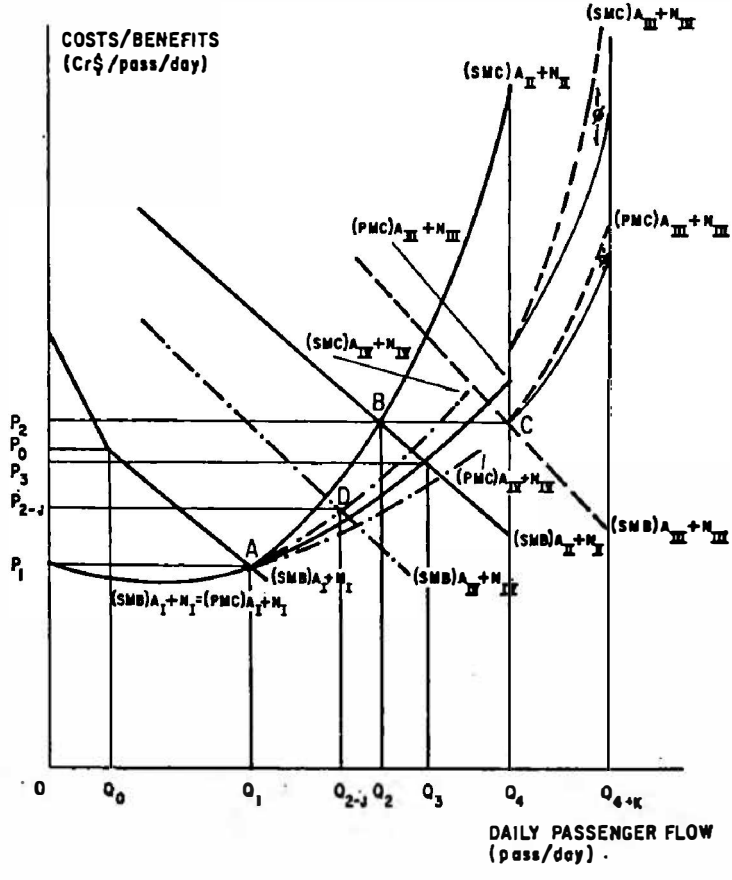
Because the cost differential widens faster for cars ($\frac{HH'}{LL'} > \frac{RR'}{NN'}$), if more passengers are transported by this mode, the main road network will be overloaded sooner. Hence, if an additional demand m is satisfied by cars, the flow capacity will be reached at Q_{2+m}. In contrast, if an additional demand $m + n$ is served by buses, flow capacity may not be attained until, let us say, Q_{4+m+n}. This is explained by the fact that, since the average number of passengers per car is lower than the average number per bus, the

[26] Transformation of these figures into triangles via linearization of the intervals DL and FN of the curves $(SMC)A$ and $(SMC)N$ is based on the hypothesis that the height of triangle MFN $(M''N)$ more than offsets the base of triangle KDL (KD). In other words, the increase in the volume of passengers transported by bus Q_3Q_4, which is larger than the increase in the volume carried by car Q_1Q_2, more than overweighs the high initial benefits derived from the use of cars at Q_1.

[27] Note that two basic premises underlie the subsequent reasoning: compared to bus transportation, there is a greater difference between the private and social marginal costs of driving cars, and the social value of the latter drops more sharply due to congestion.

FIGURE 8

EQUILIBRIUM SITUATIONS : CONSOLIDATED ROAD SYSTEM



use of cars for pendular movements results in higher traffic densities (vehicles/km) and higher total flows (vehicles/day/km).

Finally, when comparing the net benefits arising from the utilization of cars and buses, the high degree of interdependence between these benefits cannot be overlooked. For example, better bus services (higher safety and comfort standards, regular service at frequent intervals, preferential lanes and rights-of-way) can make a significant

contribution to relieving congestion by attracting some car users to public transportation and consequently increasing the benefits for those who still choose to travel by automobile.

Having analyzed the costs and benefits of car and bus services separately, let us now examine those of a consolidated road transportation system, together with the advantages and disadvantages of such a system in comparasion with a rapid transit system that does not operate on the main surface routes.

The equilibrium situations for a consolidated car-bus system, relative to non-pendular movements (situation *I*, point *A*) and pendular movements (situation *II*, point *B*), are shown in figure 8. On the horizontal axis, section $0Q_1$ corresponds to the passenger flow on the main road network at off-peak hours, while Q_1Q_2 represents the increase in the flow at peak hours. Although costs are higher at the peak periods, net social benefits are derived.

The reference prices for the two situations are P_1 and P_2, respectively. P_2 reflects a market situation created by an implicit incentive to raise the flow level from Q_2 to Q_3. This leads to inefficient use of the major routes, because the passenger flow exceeds the socially desirable level. Finally, Q_4 stands for maximum utilization of existing road space. Beyond this point, capacity overload will occur, and the passenger flow will rise at unbearably high social marginal costs. At this point, it will become necessary to expand the capacity of the main road network.

This is what happens in situation *III*, where capacity is expanded through the widening of streets and avenues, together with the construction of elevated roads and expressways. Although the marginal benefit curve shifts in response to the eased traffic conditions, the private and social marginal costs rise at a much faster pace. In fact, the curve $(SMC)\, A_{III}+N_{III}$ slopes toward a variation θ relative to $(SMC)\, A_{II}+N_{II}$, while this relative variation is represented by ρ for the curve $(PMC)\, A_{III}+N_{III}$, and $\theta > \rho$

In situation *III*, the marginal costs to the community rise sharply because wider streets, elevated roads and expressways spoil the urban landscape, limit green spaces, and increase noise and air pollution. Besides exerting these negative effects on the quality of urban living, situation *III* originates from a point where the principal road network is already used inefficiently. That is, the curve $(SMB)\, A_{III}+N_{III}$ crosses only $(PMC)\, A_{III}+N_{III}$ (point *C*), and not the social marginal cost curve. This signifies that at Q_4, when road space is being fully utilized, the market price P_3 covers only the private cost. It should be recalled that in situation *II* their price level represented social efficiency.

From what was said about situation *II* (pendular movements at peak hours), it is clear that in situation *III*, increases in peak-hour

flows ($\overline{Q}_i \overline{Q}_{i+k}$) will always be subject to congestion. Thus, situation *III* merely transfers capacity overload to a relatively near future, with the precise date depending on the volume of passengers Q_{i+k}.

However, if funds were channeled toward situation *IV* instead of toward situation *III* — that is, in the direction of underground rapid transit, for example, rather than road solutions — a great part of the demand for pendular movements would be shifted away from surface modes to the new system, and thereby enhance the mobility of city dwellers. The demand curve — and consequently the benefit curve — would shift to position (*SMB*) $A_{IV}+N_{IV}$ and cross curve (*SMC*) $A_{IV}+N_{IV}$ at point *D*, reflecting a total flow of car and bus passengers equal to Q_{s-j} at reference price P_{s-j}. This flow is not only substantially lower than the equilibrium flow of pendular movements on surface routes, but even leaves surplus road space.[28] Situation *IV* therefore represents an integrated transportation system in which the various modes (cars, buses and metros [20]) are put to best use in order to meet the demand for different volumes of daily passenger flows and different types of trips.

3 — Conclusion: transportation systems and the spatial organization of metropolitan areas

As emphasized throughout this study, transportation systems exert a significant impact on metropolitan growth and urban land-use patterns. The comparative social advantages of the alternative modes listed in table 1 are graphically depicted in figure 9. Since the horizontal axis (Q) now stands for the daily passenger flow per kilometer, increases in Q can be attributed to increases in the volume of passengers per day, in the average distances traveled, or in both. Thus, the characteristics of the various transport systems can be related to different land-use patterns.

In situations *I* and *II*, cars and buses use an uncongested main road network to link an infinite number of origins and destinations, thereby offering maximum accessibility. The dense network, utilized at relatively low social marginal costs, is highly efficient for non-pendular trips. However, if the main road network serves the central areas better than the metropolitan periphery, there will be only limited dispersion of retail commerce, personal services and recreational activities beyond the central business district. New business districts may even appear within the metropolitan center. In contrast,

[28] Even considering the lower private marginal costs in situation *IV*, it may be advisable to continue imposing charges on private users (at the price level P_i of situation *II*) in order to reduce the attractiveness of private modes relative to underground mass transportation.

[29] Commuter trains might also be included.

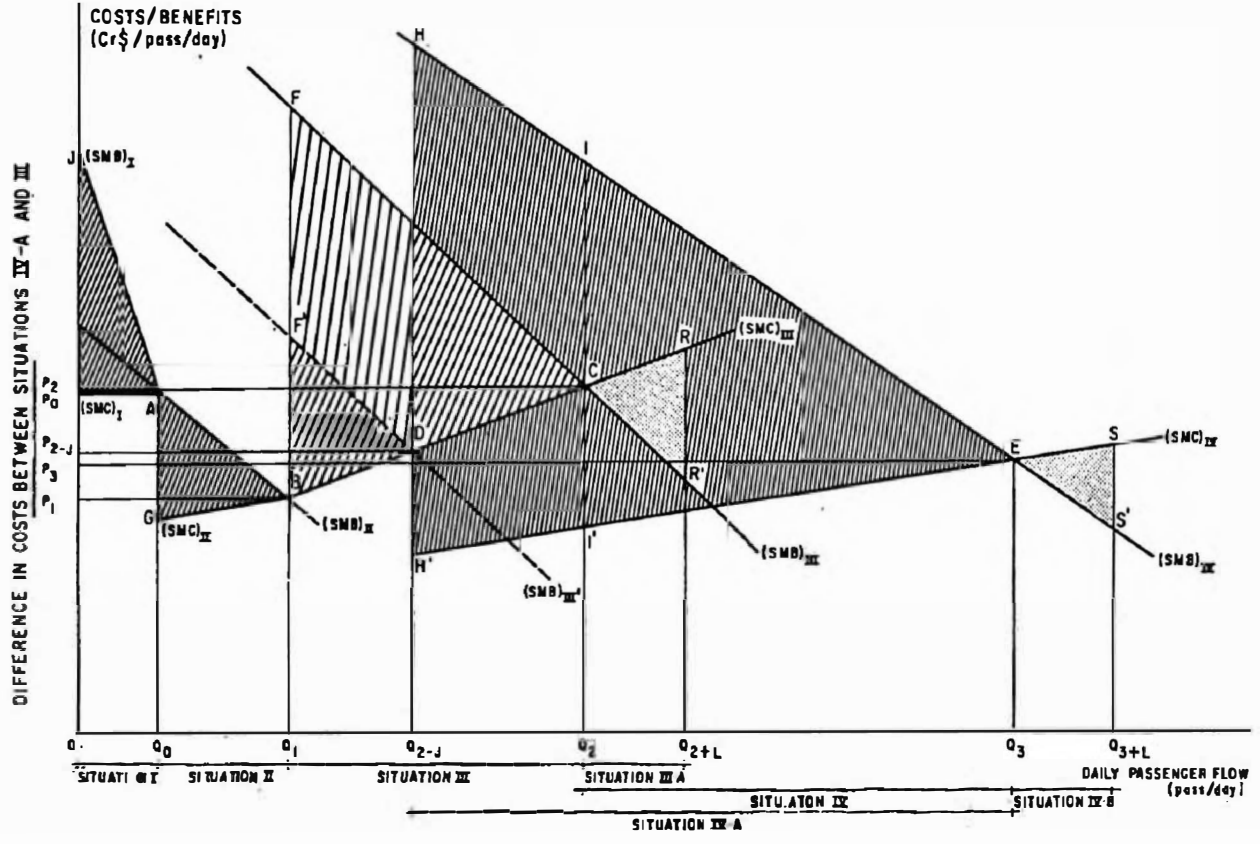
Table 1

Comparative Social Advantages of Four Transportation Systems and of Potential Intermodal Schemes

Situation	Transportation Mode	Description of Infrastructure	Type of Movement	Social Marginal Costs and Benefits	Net Marginal Benefit (Area)
I	Cars	Dense network	Non-pendular	$(SMB)I$ and $(SMC)I$	JP_oA
II	Cars and buses	Dense network	Non-pendular	$(SMB)II$ and $(SMC)II$	ABG
III	Cars and buses	Dense network	Pendular	$(SMB)III$ and $(SMC)III$ up to Q_2	FCB
III-A	Cars and buses	Congested dense network	Pendular	$(SMB)III$ and $(SMC)III$ above Q_2	Loss
IV	Mass rail transport	Rarefied or linear network	Pendular	$(SMB)IV$ and $(SMC)IV$ between Q_2 and Q_3	IEI'
IV-A	Cars, buses and mass rail transport	Complementarity between long-distance linear traffic and local traffic	Pendular	$(SMB)IV$ and $(SMC)IV$ between Q_{2-1} and Q_3 $(SMB)III'$ and $(SMC)III'$ between Q_1 and Q_{2-1}	$HEH' + F'DB$
IV-B	Mass rail transport	Congested rarefied or linear network	Pendular	$(SMB)IV$ and $(SMC)IV$ above Q_3	Loss

FIGURE 9

COMPARATIVE SOCIAL ADVANTAGES OF INTERMODAL TRANSPORTATION SYSTEMS



if the dense network covers the whole metropolitan area, it will favor the type of widespread dispersion observed in the United States. [30]

In situation *III,* pendular trips are made along the same dense network, but at higher marginal costs to the community due to the larger number of vehicles in circulation. Even so, the system is still efficient. As in situations *I* and *II,* however, if there is readier access to the central areas than to the metropolitan periphery, the tendency toward limited dispersion will be reinforced because the trend will be extended to housing, as well as to secondary and tertiary activities. [31] In situation *III,* efficient utilization of existing road space (and consequently the extensive use of intra-urban areas) is limited by Q_s. As of this point, if there is no support from mass linear modes (commuter trains) or rarefied networks (metros), additional passenger movements will be made at a net loss to the community.

Between points Q_s and Q_{s+t} (situation *III-A*), congestion cancels the possibilities that a dense road network offers for urban dispersion, if it is assumed that access to the peripheral areas is limited. The outcome will be further concentration in the central areas, where real estate prices will rise with traffic congestion. This partly explains the recent patterns of metropolitan expansion in Brazil. On the one hand, those in the middle and upper income brackets are willing to face high housing costs (paying more per square meter and living in smaller apartments in densely populated areas) in order to avoid the friction due to inefficient transportation and to take advantage of the public and private services that are progressively favoring the central areas. [32] On the other hand, those in the lower

[30] The role of transportation systems in determining the spatial aspects of metropolitan growth is emphasized by American writers. See, for instance, A. Z. Guttemberg, "Urban Structure and Urban Growth," *Journal of the American Institute of Planners,* 1960, pp. 104-110.

Nonetheless, given the fact that growth is a continuous process, no strict cause-and-effect relationship can be established between the availability and efficiency of transportation services and land-use patterns. It should also be borne in mind that the effect of a dense road network varies with urban size. In a city of one hundred thousand inhabitants, for example, such a network might lead to an additional concentration of activities in the central zones. While the same might happen in a metropolis of two million, the net outcome is more likely to be a dispersion of activities.

[31] An interesting historical point is that, until they were substituted by buses, streetcars contributed to dispersing activities and to facilitating access to the center city in Rio de Janeiro.

[32] It is precisely the environmental attraction and growing concentration of services in areas inhabited by the well-to-do that has led to real estate speculation.

income groups are obliged to live in the peripheral areas, where both housing and transportation (commuter trains [33]) are cheaper. [34] The slums located in central areas are something of an exception to this pattern, frequently acting as the first locational choices of the poorest migrants in order to eliminate transportation costs.

For pendular flows above Q_2, the mobility of those who reside in the peripheral areas can be guaranteed, and the rational use of land in the central areas assured, only in the presence of efficient mass rail transportation. Such a system, basically linear in design, can meet a demand up to flow level Q_3 at a relatively gradual rise in social marginal costs. More distant points can be reached at lower costs, since traffic is channeled along corridors that link high loading/unloading sites. In turn, new residential and employment centers can be established, and the supply of urban space expanded as the metropolitan limits push outward.

Considering that part of the pendular flow shifts from cars and buses to the mass transportation system, for services $\overline{Q_2 - Q_3}$, the net social benefits resulting from adoption of the latter alternative will rise from triangle *IEI'* to triangle *HEH'*. Despite the reduction in the net benefits derived from situation *II* (the curve drops to *FDB*), the fact that the triangular area *HEH'* is larger than the quadrilateral area *FCDF'* indicates that this loss is largely offset. Furthermore, in an integrated system in which metro (and/or pre-metro) and commuter train services complement car and bus services, the net social benefits related to pendular movements are actually represented by the sum of the areas of *HEH'* and *F'DB*. This type of system (designated *IV-A* in table 1 and figure 9) allows for dispersion not merely at the intra-urban level as in situations *I*, *II* and *III*, but on the metropolitan scale.

It may therefore be concluded that, when the outer rings of a metropolitan area are deprived of infrastruture in general and transportation services in particular, the emphasis on conventional road transportation solutions will further aggravate the metropolitan center/periphery duality. Even the differences between zones within the metropolitan center may be magnified, since varying degrees of traffic congestion make some areas more amenable than others.

[33] In Brazil, deteriorating rail services often oblige the poor to travel by bus at a significant disadvantage, given the fact that costs rise with increases in the volume of passengers transported per kilometer by this mode.

[34] Due to socio-economic dualism, in the developing countries, the poor have little chance to take advantage of the trade-off between living space (housing costs) and accessibility (transportation costs).

This vicious circle can be broken only through the introduction of rail or non-conventional bus solutions that allow for gains in scale along high-density corridors. Such solutions should be designed to promote rational metropolitan growth via decentralization, which can be achieved by integrating mass transportation with new employment poles and thereby inducing far-reaching changes in metropolitan locational patterns.

A mixed-integer programming model for the Brazilian cement industry *

Christine Ann Assis **

1 — Introduction

Most planning for the installation of new industrial plants is done by examining each project separately. Future demand for the product is projected, and the cost structure of the plant is analyzed to see if the project is viable. From the point of view of the national economy, this method has several faults. First, it isolates the project for one plant without studying demand and supply for the entire sector. Second, no account is taken of demand and supply conditions for intermediate products. Third, traditional methods of evaluation permit considering only one scale plant at a time. Fourth, such methods exclude dynamic planning which covers changes in the relevant variables over time. In all, traditional methods do not look at the trade-offs among timing, scale, and locations for the whole sector, and because of this may not minimize costs for the sector over the long run.

In order to correct these problems, this study develops a mathematical programming model in which total costs of investment, production, transportation, importing and exporting are minimized

* This article was taken from the author's Ph.D. thesis by the same title (The Johns Hopkins University, May 1977). The author would like to thank her advisers, Drs. Bela Balassa and Charles Mallar, for their comments and criticism. Thanks are also due to Mr. Ardy Stoutjesdijk and Mr. Alexander Meeraus of the Development Research Center (International Bank for Reconstruction and Development) and to the Center for the use of their computer for the estimation of the model. The author also thanks Dr. Clovis de Faro, IPEA/INPES, for his comments on an earlier version of this paper.

** Assistant Professor, Department of Industrial Engineering, Pontifícia Universidade Católica, Rio de Janeiro.

over time. The period to be analyzed is from 1976 to 1990. Economies of scale are assumed to exist in investment costs. Therefore, linear programming, in which all costs must remain constant over changes in scale (i.e. proportional to output), is not an appropriate form for the model. Thus, mixed-integer programming is used. In this type of programming, some variables must take on the integer values of zero and one, whereas other variables are allowed to be continuous.

The model is in a general form which can be applied to any industry. In order to estimate the model, the Brazilian cement industry is chosen as an example. Various aspects of the model make it particularly relevant for the analysis of this industry. The model allows one to examine the possibility of clinker shipment from limestone-abundant regions to limestone-poor regions. Clinker is the intermediate product from which cement is made. It can be shipped more easily than cement since no particular care must be taken to keep it from getting wet and because it can be shipped in bulk. No clinker is currently being shipped in Brazil, but the government has plans to encourage such shipment.[1] The second problem which is particular to the cement industry is the influence of fuel oil prices on the total cost of cement. Fuel oil is the most important input, by cost, in the manufacture of cement. Since Brazil must import 80% of her oil, the decision to produce or import cement must be made by valuing fuel oil at its economic cost rather than at the current price being paid by the cement producers.

The remainder of this paper is divided into five sections. In section 2, the specification of the model is presented. The model follows in section 3. A description of the data needed to estimate the model is found in section 4. In section 5, the versions of the model and the solution procedure are discussed. Finally, in section 6 the conclusions are presented.

2 — Specification of the model

Due to the large size of the country and the lack of a good transportation system between the North and the South, the study includes only the latter. Eighty-five per cent of cement demand originates in the South, which includes nine states (Minas Gerais, Espírito Santo, Rio de Janeiro, São Paulo, Paraná, Santa Catarina, Rio Grande do Sul, Mato Grosso, Goiás) and the capital, Brasília (Distrito Federal). Thirty-one of a total of 42 existing locations of cement plants are in this region, and plants in six more locations are scheduled for completion by 1978.

1 Banco Nacional do Desenvolvimento Econômico (BNDE), Portaria 4/74, "Critérios de Prioridade para o Subprograma Cimento, Programa de Insumos Básicos."

The set of demand locations includes one location for each state and one for Brasília. Since the Brazilian population is concentrated mainly in the state capitals, we will use the capitals as demand centers for calculating transportation costs.

Of the 37 plant locations, 23 have been eliminated as sites for future expansion. Some of the sites eliminated are located in populated areas where expansion is restricted due to public feelings about pollution (e.g. the plants in Contagem, MG, and São Paulo, SP). Others were eliminated due to lack of limestone for expansion at the site and transport cost considerations. Finally, computational considerations made it necessary to limit the number of expansion sites in order to maintain the number of integer variables at a reasonable level (see section 5).

One final product is considered: a metric ton of cement. Although three types of cement are produced in Brazil (common portland, pozolanic, and slag), the user views these as perfect substitutes. Some white cement is made by one plant, but this is used mainly for decorative purposes. Because production and demand are so small, white cement is not included in this model. Although 18% of all cement shipments are made in bulk, data are not available to estimate separate demands for cement in sacks and cement in bulk.

Only one intermediate product is present in the production of cement. Clinker is the product which leaves the kiln. After it is ground with gypsum, it is called cement. Although some countries ship large amounts of clinker, Brazil has never shipped clinker among factories. We are including clinker as an intermediate product that can be traded among factories mainly because we feel that shipments of clinker to the South of Brazil, where they can be ground with pozolanic rock, may be a feasible alternative for increasing production where limestone is not readily available.

Inputs are divided into two categories: those that are solely domestic (limestone, gypsum, electricity, labor, and miscellaneous inputs) and those that are imported (fuel oil). Some inputs do not include a transport cost, for example, labor and electricity. Other inputs are a small portion of total costs. For these, the input price includes the average transport cost to the plant. This is not a serious assumption and greatly reduces the size of the model since supply sites for the inputs do not have to be included in the model. But for limestone and fuel oil this assumption would not be acceptable. First, we would like to differentiate among plants due to their limestone costs. The transport cost is 50% or more of the limestone cost to a plant. Second, fuel oil is the most important (by cost)

input into making cement. Since fuel oil must be shipped from the coast, this greatly increases the costs of the plants located in the interior of the country.

Each plant is divided into two productive units. The first unit produces clinker by using capacity in the quarry, raw-grinding and kiln sectors. The second makes cement by utilizing capacity in the clinker-grinding and shipping sectors.

The model covers the period from 1976 to 1990. This time span is divided into five periods of three years each. Planned expansions for period one (1976-1978) are assumed to begin production as scheduled during that period. In the following four periods, the decision to construct new plants must be made. The length of the period is three years since this is the approximate time needed to plan and construct a new cement plant.

Only one export market is included in the model since exports have never been of great importance. This market can be reached by shipping to the nearest port. The bound on exports will increase as a fixed proportion of Brazilian demand.

3 — The model

The objective function to be minimized is

$$Min\ \xi = \sum_{t=1}^{5} \delta_t (\phi_{kt} + \phi_{rt} + \phi_{st} + \phi_{mt} - \phi_{et})$$

where the index t measures time periods, such that time period one is 1976 to 1978. There are five time periods each three years long.

The total costs to be minimized are: capital costs (ϕ_{kt}), domestic resources costs (ϕ_{rt}), transportation costs (ϕ_{st}), import costs (ϕ_{mt}). From this sum the export revenues (ϕ_{et}) are subtracted. Each cost item (ϕ_{kt}, ϕ_{rt}, ϕ_{st}, ϕ_{mt}, ϕ_{et}), is calculated for the midpoint of each period (e.g. 1977 for period one), which is assumed to be an average value for the period. The discount factor δ_t discounts costs to 1974 by using a formula calculated on an annual basis rather than on periods. The discount rate is 8% per year.[2] Therefore ξ represents total costs incurred from 1976 to 1990 discounted on an annual basis to 1974.

[2] This rate was chosen because it reflects the rate paid by Brazil on external bonds during this period. For example, in December 1972, Brazil offered $35 million in external bonds at 8.25% interest with maturity in December 1987. Other bonds were offered during the early 1970s with interest rates varying from 7 to 8% per year.

The effect of economies of scale on investment costs is to cause these costs to increase less than proportionately as capacity (h) increases. In order to incorporate these economies of scale into the model, investment costs must be linearized (see section 4 for how this was done). The result of the linearization is a fixed-charge function in which the fixed charge (ω) represents the intercept of the function and the variable cost (υ) represents the slope.

$$\phi_{kt} = \sum_{i \in I} \sum_{m \in M} \sigma_m \left(\omega_{mit}\, y_{mit} + \nu_{mit}\, h_{mit}\right) \qquad t \in T \tag{1}$$

Set I is the set of plant locations, and set M is the set of plant sectors (or machines). The variable y is restricted to the values zero or one. When new capacity is to be constructed (i. e., $h > 0$), then the entire fixed-charge must be accrued ($y = 1$). When no new capacity is to be constructed ($h = 0$), then no fixed-charge is paid ($y = 0$). The capital-recovery factor (σ_m) is used in order to convert capital costs into an equal stream of annual payments (ϕ_{kt}) which is sufficient to repay the original cost (ϕ_k) and the interest payments over the period the equipment is in use. Therefore, the capital recovery factor is

$$\sigma_m = \frac{\phi_{kt}}{\phi_k} = \frac{\rho\,(1+\rho)^Z\, m}{(1+\rho)^Z\, m - 1}$$

where

ρ = rate of interest
Z_m = life of the investment

When Z_m equals 25 years and the rate of interest is 8%, the value of the capital recovery factor is .0937. [8]

Domestic resource costs are those of the domestic inputs used in production.

$$\phi_{rt} = \sum_{c \in C_{rd}} \sum_{i \in I} p_{cit}\, d_{cit} \qquad t \in T \tag{2}$$

where

p_{cit} = price of domestic inputs paid by plant i at time t
d_{cit} = purchases of domestic inputs by plant i at time t
C_{rd} = set of domestic inputs

8 A real rate of interest of 8% corresponds to the lending policy of BNDE: monetary correction plus 8%. During 1973 inflation was at a level that the 20% limit on monetary correction was not effective.

Transport costs (ϕ_{st}) must be calculated for final goods, intermediate goods, and inputs.

$$\phi_{st} = \sum_{c \in C_f} \left\{ \sum_{i \in I} \sum_{j \in J} \mu_{cijt}\, x_{cijt} + \sum_{j \in J} \mu_{cjt}\, m_{cjt} + \right.$$
$$+ \left. \sum_{i \in I} \mu_{cit}\, e_{cit} \right\} + \sum_{c \in C_{rp}} \left\{ \sum_{i \in I} \mu_{cit}\, m_{cit} \right\} +$$
$$+ \sum_{c \in C_m} \left\{ \sum_{\substack{i,\, i' \\ i \neq i'}} \mu_{cii't}\, x_{cii't} + \sum_{i \in I} \mu_{cit}\, m_{cit} + \right. \qquad (3)$$
$$+ \left. \sum_{i \in I} \mu_{cit}\, e_{cit} \right\} + \sum_{c \in C_{rd}} \left\{ \sum_{i \in I} \mu_{acit}\, d_{cit} \right\} \qquad t \in T$$

where

μ_{cijt} = transport cost for one unit of product c from plant i to demand center j during time t

x_{cijt} = number of units of good c shipped from plant i to demand center j during time t

μ_{ajt} = transport cost for one unit of imported product c from the port of entry nearest the demand center j during time period t

m_{ajt} = number of units of good c imported and shipped to demand center j during period t

μ_{cit} = transport cost for one unit of good c between plant i and the nearest port during time t

e_{cit} = number of units of good c exported from plant i during period t

m_{cit} = number of units of good c imported by plant i during period t

$\mu_{cii't}$ = transport cost for one unit of intermediate product from plant i to plant i' during period t

$x_{cii't}$ = number of units of intermediate product shipped from plant i to plant i' in time t

μ_{acit} = transport cost from nearest domestic supply center to plant i in period t

C_f = set of final goods

C_m = set of intermediate goods

C_{rp} = set of imported inputs

J = set of demand centers

Import costs (ϕ_{mt}) are the sum of the cost of importing final goods, intermediate products, and inputs.

$$\phi_{mt} = \sum_{c \in C_f} \sum_{j \in J} p_{mcjt}\, m_{cjt} + \sum_{c \in C_{rp}\ UC_m} \sum_{i \in I} p_{mcit}\, m_{cit} \qquad t \in T \qquad (4)$$

where

p_{mcjt} = import price of final good c during time period t

p_{mcit} = import price of imported inputs and intermediate products during period t

Revenue from exports (ϕ_{et}) is included as

$$\phi_{et} = \sum_{C \in C_f\ UC_m} \sum_{i \in I} p_{ect}\, e_{cit} \qquad t \in T \qquad (5)$$

Exports include final goods and intermediates where

P_{ect} = price for exported good c during time t

This objective function must be minimized subject to seven types of constraints: (1) material balance, (2) capacity, (3) market requirements, (4) export bounds, (5) maximum investment, (6) integer, and (7) non-negativity.

Material balance constraints. The material balance constraints for final products state that production of a final good must be greater than or equal to domestic shipments of the final good plus exports.

$$\sum_{p \in P_i} a_{cpit}\, Z_{pit} \geqq \sum_{j \in J} x_{cijt} + e_{cit} \qquad \begin{matrix} c \in C_f \\ i \in I \\ t \in T \end{matrix} \qquad (6)$$

where

a_{cpit} = the input-output coefficient for good c in process p at plant i during time t (a_{cpit} is negative for an input, positive for an output)

Z_{pit} = a level of production of one ton in process p at plant i during time t

P_i = set of processes for plant i

For the intermediate product clinker, net output at a clinker plant i plus clinker shipments from plant i' ($i' \neq i$) to plant i plus imports to plant i must be greater than or equal to clinker shipments from plant i to plant i' and clinker exports from plant i.

$$\sum_{p \in P_i} a_{cpit} Z_{pit} + \sum_{\substack{i' \\ i' \neq i}} x_{cii't} + m_{cit} \geq \sum_{\substack{i \\ i' \neq i}} x_{cii't} + e_{cit} \qquad (7)$$

$$c \in C_m$$
$$i \in I$$
$$t \in T$$

By net output of clinker, we mean that produced by the plant which is not used in cement production. If net output is negative, then interplant shipments and imports must be positive.

For domestic inputs, the material balance constraints state that purchases of these inputs must be greater than or equal to their use. (The a_{cpit} are negative for inputs.)

$$\sum_{p \in P_i} a_{cpit} Z_{pit} + d_{cit} \geq 0 \qquad (8)$$

$$c \in C_{rd}$$
$$i \in I$$
$$t \in T$$

For imported inputs, the material balance constraints state that imports of these inputs must be greater than or equal to their use.

$$\sum_{p \in P_i} a_{cpit} Z_{pit} + m_{cit} \geq 0 \qquad (9)$$

$$c \in C_{rp}$$
$$i \in I$$
$$t \in T$$

Capacity constraints. Total capacity required in each plant in each productive unit during each time period must be less than or equal to initial capacity plus the difference between capacity expansion (h_{mit}) and retired capacity (s_{mit}) during all previous periods.

$$\sum_{p \in P_i} b_{inpit} Z_{pit} \leq k_{mi} + \sum_{\substack{t' \in T \\ t' \leq t}} (h_{mit'} - s_{mit'}) + \sum_{\substack{t' \in T \\ t' \leq t}} \gamma_{mit'} \qquad \begin{matrix} m \in M \\ i \in I \\ t \in T \end{matrix} \qquad (10)$$

where

k_{mi} = initial capacity at plant i

γ_{mit} = required additions to capacity in machine m at plant i during time t

b_{inpit} = required capacity in machine m at plant i during time t (in tons)

Since various plants have plans for increasing capacity in 1976 to 1978 (our period 1), it is assumed that these plans are fixed and will be completed on time so that they will form part of effective capacity during period one. Therefore, in period one, capacity will be $k_{mi} + \gamma_{mi1} + h_{mi1} - s_{mi1}$. It is initially assumed that fixed additions to capacity occur only in period one (i.e., $\gamma_{mi2} = \gamma_{mi3} = \ldots = \gamma_{mis} = 0$). In another version of the model, certain fixed additions are required in period two, and the effect of this on the minimum cost of the investment program is determined. This is done by including positive values for the γ's in period two.

It is also assumed that no capacity is retired ($s_{mit} = 0$) during the period. Since the forecast period is only 15 years long, this is a reasonable assumption.

Market requirements constraints. Each demand center must receive at least the quantity it requires of each final good. Since demand can be satisfied by domestic shipments or imports, the market requirements constraint states that the sum of the shipments of the final good from each plant to one demand center j plus imports to that center must be greater than or equal to the demand for the final good at the demand center during time t.

$$\sum_{i \in I} x_{cijt} + m_{cjt} \geq r_{cjt} \qquad \begin{matrix} c \in C_j \\ j \in J \\ t \in T \end{matrix} \qquad (11)$$

where r_{cjt} = required demand of final good c at demand center j during time t.

Export bounds. In order to limit exports to reasonable levels, bounds must be placed on the amount that can be exported. If no bounds were placed, the model would imply that very large amounts could be exported without affecting the export price. Since the export price is a parameter in this model, bounds will be placed on the level of exports.

$$\sum e_{cit} \leq \bar{e}_{ct} \qquad c \in C_j\ UC_m \qquad t \in T \tag{12}$$

where $\bar{e}_{ct}$ = upper bound on exports of good c during time period t.

Maximum investment constraints. An upper bound must also be determined for the maximum investment at one location during one time period. This bound should be based on the largest capacity which is used to calculate the investment function. Since the investment function is only valid for the range of capacities that was used to estimate it, the largest capacity should be used as the upper bound on investment.

$$h_{mit} \leq \bar{h}_{mit}\, y_{mit} \qquad m \in M \qquad i \in I \qquad t \in T \tag{13}$$

where

h_{mit} = 2 640 thousand tons

Integer constraints. The fixed-charge variable must be constrained to be either zero (no fixed charge occurs when capacity expansion is zero) or one (the total fixed charge occurs when there is any increase in capacity).

$$y_{mit} = 0 \text{ or } 1 \qquad m \in M \qquad i \in I \qquad t \in T \tag{14}$$

Non-negativity constraints. Finally, the variables must be constrained to be non-negative.

$$Z_{pit'}\ x_{cijt'}\ e_{cit'}\ m_{cit'}\ m_{cjt'}\ h_{mit'}\ d_{cit'}\ y_{mit'} \qquad (15)$$
$$x_{cii't} \geq 0$$

4 — Data

In order to obtain the most accurate solution possible, attention must be paid to the collection of data and the estimation of the parameters. The data necessary for the estimation of the model are: plant capacities, demand projections, input-output tables, domestic and import prices, capital costs and transportation costs.

Plant capacities. Information on plant capacities in 1974 was furnished by the Sindicato Nacional da Indústria de Cimento (SNIC), the cement producers association. This organization also furnished a projection of capacity expansion through 1981.

Fourteen locations were chosen as expansion sites: Pedro Leopoldo, MG; Mesquita, MG; Uberaba, MG; Carandaí, MG; Lagoa Santa, MG; Cantagalo, RJ; Jacupiranga, SP; Apiaí, SP; Capão Bonito, SP; Salto de Pirapora, SP; Rio Branco do Sul, PR; Campo Largo, PR; Canoas, RS; Corumbá de Goiás, GO.

Demand estimates. Demand for cement is estimated by using a pooled time-series cross-section model with observations for nine states and seven time periods (1968 to 1974).[4] Due to the special nature of the construction of Brasília, data for that city are not included in the pooled model. The equation estimated was

$$\log CON = 1.762 + .915 \log RYC + (\textit{dummy variables } [-.83] \; [3.32] \textit{ for the slope and intercept of each of eight states})$$

$$R^2 = .947$$

$$DW = 1.880$$

where

CON = per capita purchases of cement in kilograms

RYC = per capita state income deflated by wholesale price index for products for domestic use (Column [12], *Conjuntura Econômica*).

Using the estimated income elasticity (.915), two demand forecasts are made for the period from 1975 to 1990. One forecast (hereafter called "demand I") was based on the assumption that real per capita income will grow at the rate of 5% per year. Since the average rate of population growth is 2.9%, this assumption

[4] Instituto Brasileiro de Geografia e Estatística (IBGE), *Anuário Estatístico do Brasil*, 1974.

amounts to an annual growth rate of approximately 8% for gross national product. The other forecast ("demand II") is based on a growth rate of 7% for per capita income.

For Brasília, a better explanation of cement consumption is obtained by regressing total consumption on real federal government expenditures. This relationship, though, is not expected to continue through our entire prediction period. Therefore, it is assumed that total cement consumption will increase either 10% (demand I) or 14% (demand II) per year.

Demand assumption I predicts that total cement consumption for this region will be 38.671 thousand metric tons by 1989. This is a 208% increase over the 1974 amount, or a growth rate of 7.8% per year. For demand assumption II, the increase over 1974 is 302%, or a growth rate of 9.7%.

Input-output tables and domestic prices. The southern part of Brazil has been divided into four regions based on the quantity and accessibility of limestone in each. Region one consists of the states of Minas Gerais, Espírito Santo, Rio de Janeiro, and São Paulo. The states of Goiás, Mato Grosso, and the federal capital, Brasília, make up region two. Region three is composed of Paraná and Santa Catarina. The state of Rio Grande do Sul is region four. Each of the plants in a region is assumed to use the same processes. For each region there is a set of input-output tables for the production of clinker and cement. SNIC developed these tables by taking the individual tables for each plant and weighting each coefficient by the plant's contribution to total output.

The prices of the inputs are also the weighted averages of the actual prices paid by the factories, and they include the transportation costs. These data were also compiled by SNIC.

The major inputs into the cost of clinker are: labor (5.9 to 17.7%), limestone (.5 to 15.2%), fuel oil (56.6 to 77.5%), electricity (4.6 to 11.7%), and grinding media and kiln lining (1.3 to 7.8%). Other expenses are for sand, clay, slag, iron ore, diesel oil, grease, gasoline, coal, and explosives. Transportation for all inputs except limestone and fuel oil are included in their costs for these two inputs. Transportation costs are calculated separately.

The major inputs into the production of cement are: clinker (71.6 to 78.9%), labor (1.1 to 2.2%), gypsum (1.4 to 3.8%), electricity (2.8 to 5.7%), grinding media (1.5 to 2.9%), and sacks (9.7 to 13.6%). Twenty sacks are used to bag a ton of cement. Other costs include slag, pozolanic rock, diesel oil, grease, and chemicals.

Import prices. The prices at which clinker and cement could be imported into Brazil are important because the level of these prices

will determine whether new plants will be constructed or not. Since Brazil has not imported large amounts of cement in recent years, the import prices were calculated by estimating European export prices and the freight rate to Brazil.

Due to the uncertainty in predicting future import prices, the model is estimated using two different sets of import prices. For set one, the cement import price is US$ 40 per ton, and the clinker import price is US$ 35 per ton. This corresponds to a European export price for cement of US$ 24 and for clinker of US$ 22. (Bagged cement is more costly to ship than clinker, which is shipped in bulk.) For the other set, the cement import price is US$ 45 per ton, and the clinker import price is US$ 39.38. This corresponds to an export price of US$ 29 per ton of cement and US$ 27 per ton of clinker.

Fuel oil is the only imported input used in cement production. Since the 1974 import price was not available, we took the 1973 price of imported fuel oil (CIF) into Brazil, US$ 44.68 per ton, and inflated it by the United Nations fuel oil price index. This results in a price of US$ 111 per ton. This price is used in the model, and domestic transport cost is added for those plants that are not on the coast.

Export prices and bounds. Since the southern region of Brazil has never been a large exporter of cement, it was decided to limit exports to 3% of projected demand in each period. In order to compete with exports from Europe, Brazilian cement could not cost more than US$ 25 per ton (FOB) if we assume a European export price of US$ 24 per ton. Clinker would have to be priced at US$ 22 per ton. If it is assumed that the European export price is US$ 29 per ton, then the Brazilian export price could be US$ 30 per ton of cement or US$ 26.25 per ton of clinker.

Transportation costs. In the model, transportation costs for cement, clinker, fuel oil, and limestone are minimized. Since a complete shedule of rates between all locations is not available, it is necessary to estimate the cost of shipping one metric ton one kilometer. To do this the total transport cost per ton (COST) is regressed on the distance in kilometers (KM) in order to obtain a fixed cost (a) and a variable cost per kilometer (b).

$$COST = a + b\ (KM)$$

The estimated coefficients are then combined with a table of distances to obtain complete estimated transport cost tables.

Investment costs. Investment costs are included in the model by means of a fixed-charge function which linearizes the effect of economies of scale on the investment costs (C):

$$C = \omega\, y + v\, h$$

where the definitions of the variables are the same as in section 3. The data used to estimate this function for the two sectors of the cement plant (clinker production and clinker grinding) were furnished by a cement equipment manufacturer who has produced two-thirds of the equipment used in the Brazilian cement industry. The costs included are for cement-making equipment (grinding mills, kilns, etc.), auxiliary equipment, spare parts, electrical equipment, process-control equipment, erection, civil engineering, roads and fences, water-supply plant, purchase of the quarry, opening of the quarry, quarry equipment, offices, laboratory, repair shop, freight, insurance, consulting engineer's fee, contingencies, and working capital.

Exchange rate. The exchange rate used is the official rate for December 1974: Cr$ 7.327 = US$ 1.000. All costs are in terms of 1974 *cruzeiros.*

5 — Estimation of the model

Since the solution of the model depends crucially on the projections for cement demand and on the import prices for cement and clinker, the model was solved several times using different values for these parameters. In all, the model was solved 12 times.

Two sets of demand projections are used. For each set of projections, the model is estimated five times. Two different import prices for cement are used, and for each import price, the model is estimated twice. For the first version, a US$ 45 per ton import price for cement is used and increases in demand are satisfied completely by imports. This version is called the "all import model", and it serves as a basis for comparison with the model described in section 3. The second version allows for the construction of new plants or imports. This model is described in section 2 and is called here the "investment model". If new investment in the Brazilian cement industry is to be undertaken, then the cost of this version must be less than the cost of the all import model. The third version of the model estimates the all import model using a US$ 40 per ton import price, and the fourth version estimates the investment model while specifying a US$ 40 per ton import price. The fifth version does not allow for any imports; all increases in demand must be met by construction of new plants. This version, the "no import model", also serves as a basis for comparison with the estimation of the investment model.

Besides the 10 solutions described above, two other versions (one for each demand projection) are estimated. In these, capacity expansion in period two is limited to those projects which have been

approved by the government and which are to begin production by 1980. For periods three to five, the same assumptions hold as for the investment model with the US$ 40 import price. The approved projects include expansions of 700 thousand tons in Jacupiranga, SP; 350 thousand tons in Capão Bonito, SP; 350 thousand tons in Rio Branco do Sul, PR; and 170 thousand tons in Canoas, RS. Also, a new plant of 700 thousand tons is planned for Salto de Pirapora, SP. Loans for expansions beyond 1980 have not been approved. These versions are called the "approved projects models". The solutions to these versions are compared to the investment model solutions to determine the efficiency of the project selection process.

The model was solved on a CDC 6000 computer by using the APEX linear programming system. The algorithm used for mixed-integer programming (MIP) was "branch and bound". Due to the large number of integer variables, it was not possible to examine enough nodes to prove that the global optimum had been found. For each model the algorithm found one or two integer solutions. A solution was chosen so that it was within 2% of the optimal solution to the corresponding linear program (the linear program which would result if the integer constraints were dropped). The optimal solution of the corresponding linear program (*LP*) is the minimum value for the optimal solution to the mixed-integer program. Therefore, the optimal value of the mixed integer program must lie between the optimal solution to the linear program and the solution we have found. Indeed, as it is possible that a solution which has been found by this method could be the global optimum, but without allowing the algorithm to terminate (which would involve a large amount of computer time), it is impossible to prove that a solution is the global optimum.

Since the APEX system calculates Driebeck/Tomlin penalities at each node, it is expected that the first integer solution obtained will be a good one. Each penalty is the lower bound on the total change in the value of the objective function if the variable being examined were forced to take on an integer value. The algorithm chooses as the next variable the one with the largest penalty in one direction or another. Branching is done in the opposite direction of the largest penalty. Since the algorithm tries to find an integer solution as quickly as possible, it works directly down a branch without backtracking.

Because the amount of computer time depends crucially on the number of integer variables, limiting the number of these variables would lower the cost of each solution. For the model there are 112 integer variables (14 locations, two sectors, four time periods). For 11 of the 14 locations there is no reason to build one sector without the other. In each of these cases the cost of imported clinker is

higher than the local production cost, so that clinker would not be imported and ground at the site. Also, none of these sites has a production cost which is high enough to make interplant shipments possible. For each of these 11 locations, a constraint was added to the model:

$$y_{cit} - y_{kit} = 0 \qquad i = 1, \ldots, 11 \qquad t = 2, \ldots, 6$$

This constraint states that if the integer variable for the cement sector (y_{cit}) is one (zero), then the integer variable for the clinker sector (y_{kit}) must also be one (zero). Either both sectors or neither must be built. Including these constraints eliminates 44 decisions from the model, leaving 68 decisions.

For the other three locations, this system cannot be used. Two of these locations (Apiaí and Capão Bonito, SP) are located within the distance it is possible to ship clinker to the South of Brazil. Therefore, the possibility of producing only clinker at these sites must be included. The last site is Canoas, RS, which can receive clinker from Apiaí and Capão. Here the possibility of constructing just a clinker-grinding (cement production) sector must be included.

In order to reduce the number of integer variables even further, the models were first run as linear programs (i.e. without the integer constraints). By examining the values of the integer variables, it was decided to eliminate those whose continuous values were zero or very small. When the model was rerun with the integer constraints, these variables were set equal to zero. In most cases the alternatives which were eliminated were in the earlier periods and/or locations in the state of Minas Gerais (where initial capacity exceeds demand in all subsequent periods).

6 — Conclusions

By comparing the total costs of the three models, it can be seen how much costs increase if either the no import model or the all import model is chosen as a development strategy rather than the investment model. (Since the investment model allows for imports, exports and production, the total cost of this model must be less than the total costs of either of the other two models.)

Of the three models, the most expensive is the all import model (see table 1). With a US$ 40 per ton import price, this model is 1.7% (3.2%) more costly than the investment model if demand I (demand II) is assumed. When the import price is US$ 45 per ton, the all import model is 4.4% (7.0%) more costly than the investment model with demand assumption I (II).

Table 1

Total Costs for Periods One through Five
(Discounted to 1974)

	Demand I	Demand II
	(Cr$ 1 000)	
US$ 45/Ton Import Price		
All Import Model	50 570 530	63 010 594
Investment Model	48 431 856	58 879 433
US$ 40/Ton Import Price		
All Import Model	48 780 700	59 996 364
Investment Model	47 946 382	58 114 897
Capacity in Period 2 Set to Approved Projects	52 253 448	62 517 346
No Import Model	48 771 211	59 637 944

Although the no import model is more costly than the investment model, the difference is not very large. When demand grows as is assumed in demand assumption I, the no import model is 1.7% (.7%) more costly than the investment model which assumes a US$ 40 (US$ 45) import price. For demand II, the no import model is 2.6% (1.3%) more costly than the investment model with a US$ 40 (US$ 45) import price. The higher the import price, the less the difference in cost between the no import and the investment models.

As can be seen in table 1, an increase in the growth rate of real per capital income from 5% (demand I) to 7% (demand II) causes total costs to increase 24.6% (23.0%) for the all import model with the US$ 45 import price (US$ 40); 21.6% (21.2%) for the investment model with the US$ 45 import price (US$ 40); and 22.3% for the no import model. This means that a 2% increase in the rate of growth of per capita income results in a 1.3% (for the investment models) to 1.5% (for the all import models) annual cost increase.

If capacity expansion in period two is limited to those projects which have been approved, then total costs increase 9.0% (7.6%) over the investment model for demand I (demand II) with a US$ 40 per ton import price. The increase is less for demand II since the fixed additions are a smaller part of total capacity expansion. This increase in costs is an indication of the importance of economies

of scale. Since the required expansions in capacity range from 170 thousand tons in Canoas to 700 thousand tons in Jacupiranga and Salto de Pirapora, all are much smaller than the maximum size plant (2 640 thousand metric tons). Although the five smaller expansions cause some savings in transportation costs, this savings is far outweighed by increases in investment costs due to construction of less than maximum size plants.

Clinker shipment. Besides allowing comparisons of the various versions of the model, the solutions allow one to analyze the possibility of interplant clinker shipments. Only two of the plants in the northern region lie within a distance that make it profitable to ship clinker to Rio Grande do Sul. Although the variable cost difference between plants in São Paulo and Minas Gerais and plants in Rio Grande do Sul is Cr$ 87.30 per ton of clinker (Cr$ 230.54 as compared with Cr$ 143.24), the cost of rail shipment limits the maximum distance clinker can be shipped to approximately 1 200 kilometers. Shipments can be made between the plants in Apiaí and Capão Bonito, São Paulo, and the plant in Canoas, Rio Grande do Sul. For the investment models, the minimum cost solution is to construct clinker-grinding facilities in Rio Grande do Sul and to import the clinker. A US$ 40 per ton (US$ 45) import price for cement amounts to a US$ 35 per ton (US$ 39.60) import price for clinker. If no imports are allowed, then shipments of clinker originate in Apiaí, SP. Although Apiaí produces clinker at a cost of Cr$ 220.51 or US$ 30.10 per ton, including the investment cost of a maximum size plant, the transport cost is Cr$ 85.31, for a total delivered cost of Cr$ 305.82 (US$ 41.74). Therefore, clinker shipment to the South is profitable only if no imports are allowed or if clinker cannot be imported for less than US$ 41.74.

Locations and timing. All versions of the model select Salto de Pirapora and Capão Bonito to serve the São Paulo market; Corumbá de Goiás to supply the western states of Goiás, Mato Grosso, and Brasília; Canoas to grind clinker which is imported or shipped from São Paulo; and Apiaí to supply clinker to Canoas (in the no import model) or cement to São Paulo and Paraná. The no import models also choose Cantagalo to serve the market in Rio de Janeiro. When demand assumption II is made, Jacupiranga is also selected to serve the São Paulo market.

Salto is chosen to serve São Paulo since it is located closest to the city of São Paulo (120 kilometers as compared with 210 for Capão, 270 for Jacupiranga, and 340 kilometers for Apiaí). The state of São Paulo has the largest demand for cement. When demand I is assumed, in 1977 demand is 6 827 thousand tons. This rises to 17 260 thousand tons by 1989. To keep up with this demand, four maximum size plants must be built. For the investment model with

the US$ 40 import price, two maximum size plants are built at Salto (in periods three and four), one at Capão (period five), and one at Apiaí (period five). When the import price is increased to US$ 45 per ton, another plant is built in Capão during period four.

If no imports are allowed, Apiaí is selected to ship clinker and a plant is constructed in Cantagalo to serve the market in Rio. The second plant at Capão replaces the shipments from Apiaí and the plant in Cantagalo replaces the shipments from Capão to Rio.

Also for demand I, one plant of 2 171 thousand ton capacity is built in Corumbá de Goiás. (The size of the plant increases by 280 thousand tons when the import price increases.) This plant is built in period four and serves the market in Goiás, Mato Grosso, and Brasília. Together, the demand in these three states rises from 1 382 thousand tons in 1977 to 3 881 thousand tons in 1989. This increase is satisfied by construction of the plant in Corumbá and by shipments from existing plants in Minas Gerais.

When imports are allowed, a clinker-grinding sector is constructed at Canoas which can produce 1 664 thousand tons of cement per year. The sector is constructed in period four. In periods two and three, cement is imported into Rio Grande do Sul. For the no import model, clinker-grinding sectors are constructed in period three (1 590 thousand tons) and in period four (1 893.9 thousand tons). This plant also satisfies part of the previously imported demand in Santa Catarina and Paraná.

When demand assumption II is used, demand in São Paulo goes from 7 072 thousand tons in 1977 to 20 817 thousand tons in 1989. Therefore, one more plant must be built to meet the increase in demand. This plant is constructed in Jacupiranga. Besides this plant, the model which has a US$ 45 import price constructs a plant in Campo Largo, PR to replace imports into that state. For assumption II, demand in Paraná increases from 1 027 thousand tons in 1977 to 3 316 thousand tons in 1989, or 748 thousand tons over the increase predicted by assumption I. The construction of this plant is now profitable because the average investment cost per ton of capacity for a plant of 2 648.7 thousand tons is Cr$ 23.58 per ton per year less than that for a plant with a capacity of 748 thousand tons less.

The sites not chosen by the model are those in Minas Gerais and in Rio Branco do Sul, PR. The sites in Minas Gerais are not chosen because demand in that state increases only 2 278 thousand tons (3 574 thousand tons) when demand I (demand II) is assumed. Since existing capacity in the state in 1977 is 6 560 thousand tons, all increases in state demand can be met by using this capacity. Less shipments are made to Rio and São Paulo, where new plants are constructed.

As long as adequate quarries are located closer to the main demand centers of São Paulo and Rio, expansion in the state of Minas Gerais should be limited to increases in that state's demand. The increase in transportation costs is large enough not to warrant more expansion in that state at this time. When nearby quarries are exhausted (probably not for another 25 years), then the optimal solution would be to haul clinker from the state of Minas Gerais to São Paulo and grind it with gypsum at these sites. Since gypsum comes from the North of Brazil to the ports of São Paulo and Rio, this would save the transport cost from the coast to Minas Gerais. More savings would result since clinker is less costly to ship than bagged cement and since there are no moisture problems in storage. Therefore, the least cost solution would be to operate grinding facilities in these cities.

The location of Rio Branco do Sul, PR is not chosen because demand in that state increases just enough to justify construction of one plant of optimal size.

Scale. The solutions of the model provide two indications of the importance of economies of scale. One is the choice of scale: plants are never chosen with a capacity of less than two million tons. The other indication is the solutions of the approved projects models. The costs of these models, which specify construction of smaller than maximum size plants, are much larger than the costs of the corresponding investment models.

Those plants that are constructed in Salto de Pirapora, Apiaí, Capão Bonito, Jacupiranga, and Cantagalo are always of maximum size. The exception to this is for Capão Bonito and Cantagalo, where two maximum size plants cannot be constructed due to the constraint on limestone extraction. The second plants, though, are larger than two million tons of capacity. Since demand is not very large in the western states of Goiás, Mato Grosso, and Brasília for demand assumption I, a less than maximum size plant is constructed at Corumbá de Goiás. This plant is not constructed, however, until demand is sufficient to warrant construction of a plant with over two million tons of capacity. The clinker-grinding facility at Canoas is not constructed until it can be built at approximately 1 500 thousand tons of capacity. Before this, it is less costly to import cement.

Due to the importance of economies of scale, incentives should be given to increase the capacity of future cement plants up to the maximum capacity offered by the producers of cement-making machinery. Demand is growing at a rate which is adequate for the installation of maximum size plants.

Imports. As demand increases from 1977 to 1989, cement imports become more important. For the all import model, by period five,

imports reach 20 million (32 million) tons per year if demand assumption I (II) is made. This amounts to 52.4% (63.5%) of total consumption. With demand assumption I and the US$ 40/ investment model, cement imports amount to 11.5% of total consumption by period five. This falls to 3.7% if the import price is raised to US$ 45 per ton. The corresponding numbers for demand assumption II are 18.9% and 2.7%. Imports are a smaller percentage for the US$ 45/investment model since the larger increase in demand causes construction of a plant in Paraná which replaces imports into that state. With either demand I or a US$ 40 import price, this plant is not profitable.

For the approved projects models, cement imports amount to 8.9% of total consumption if demand I is assumed or 15.1% if demand II is assumed. This is less than the corresponding investment models since the approved projects model replaces imports into São Paulo and Paraná with domestic production.

Clinker imports occur in all versions except for the all import model. Since no construction is allowed in the all import model and no separate clinker-grinding facilities currently exist in Brazil, no clinker can be imported. For all other versions, clinker imports amount to more than one million tons per year by period five. For one version (US$ 45/demand II/investment model), clinker imports are actually larger than cement imports. All clinker imports are ground by the facility in Canoas for consumption in the state of Rio Grande do Sul.

For the all import models, the cost of imports can reach US$ 1.4 billion per year by 1989 (for a US$ 45 import price and demand assumption II). Some doubt is placed on whether the world market price could remain constant if Brazil were to purchase the quantities required by the all import models. If so, the cost of the all import models could be even higher. Import costs for the investment models are much less, but range from a low of US$ 109 million per year for the US$ 45/demand I/investment model to a high of US$ 453 million for the US$ 40/demand II/investment model.

The states of Santa Catarina, Rio, and Rio Grande do Sul receive some imports in at least one period in all versions of the model. The capital cities of all three states are ports, so that no domestic transport cost is charged to imports. Since all expansion sites must pay some transport costs to these cities, imports have a large advantage over domestic production.

Cement is imported into Rio Grande do Sul in the periods before the clinker-grinding facility begins operation. Thereafter, only clinker is imported.

São Paulo receives some imports in all models except the US$ 40/ demand I/approved projects model. Due to the construction of the approved projects (most of which are located in São Paulo), no imports are needed in this version.

The state of Paraná receives imports in the all import models, the US$ 40/demand I/investment model, and the demand II/approved projects model. If the import price increases to US$ 45 per ton, then domestic production is less costly than imports. Domestic production is also less costly when demand increases more rapidly, as it does for demand assumption II.

The only other states that are supplied by imports are the states of Espírito Santo and Goiás for the all import model when demand II is assumed. Demand is growing so rapidly in demand II that without allowing for construction of new plants, imports must be made into seven states (including Goiás, which is 1 506 kilometers from the coast). Because of this the total cost of this model increases very rapidly.

Exports. For all the models with the US$ 40 per ton import price, the corresponding export prices are US$ 22 per ton of clinker and US$ 25 per ton of cement. At these prices no exports are made.

When the import price is increased to US$ 45 per ton, the corresponding export prices become US$ 26.25 per ton of clinker and US$ 30 per ton of cement. At these higher prices, clinker exports reach the bound (94.4 thousand tons) for period one when demand I is assumed and for periods one to three when demand II is assumed (94.4 thousand tons, 118 thousand tons, and 147.8 thousand tons, respectively).

Exports come mainly from the plant in Cachoeiro do Itapemirim, ES. This plant also exports 377.6 thousand tons of cement during period one (demand I) or 218.3 thousand tons in period one and 129.6 thousand tons in period three (for demand II).

Clinker and cement exports do not appear very promising. In order for a plant to compete internationally, its production facilities must be located on the coast. Since most coastal limestone deposits have been depleted, construction of a plant on the coast is not possible. Only two plants, Canoas and Rio, are located in coastal cities, but neither plant has a quarry at the site. (The plant at Cachoeiro, which supplies the bulk of the exports in the minimum cost solution, is located 162 kilometers from the port at Vitória, ES.)

Capacity expansion. Total capacity expansion is smallest for the US$ 40 investment models. Capacity expansion equals 76.2% (70.2%) of the increase in total consumption when demand I (II) is assumed. When the import price is increased to US$ 45 per ton, capacity expansion increases to 95.4% (95.7%) of the increase in

consumption for demand assumption I (II). For the no import models, capacity expansion equals the total increase in consumption. The approved projects models provide for more installation of new capacity than the corresponding investment models, since part of their capacity expansion is fixed to non-minimum cost locations. If demand I (II) is assumed, capacity expansion provides for 82.1% (76.3%) of the increase in consumption.

Since existing capacity in the cement industry in period one amounts to 18 420 thousand tons, capacity expansion must occur at a rate which doubles installed capacity in 12 years for demand assumption I, or more than doubles capacity for demand II. In 1970, existing capacity was 7 883 thousand tons. The rate of increase from 1970 to 1977 was 12.9% per year. The largest rates of increase are for the no import models which predict a rate of increase of installed capacity of 6.5% (8.7%) if demand I (II) is assumed. These rates are below the actual rate from 1970 to 1977 and seem likely to be attainable. Since the base capacity was low in 1970, it would not be expected that the rate achieved previously could be maintained indefinitely. The US$ 40/investment models forecast the smallest rates of capacity expansion: 5.3% for demand I and 6.9% for demand II.

The problem of excess capacity is related to that of capacity expansion. Since there is no capacity expansion in period one, excess capacity in this period is equal to existing capacity minus demand plus imports. Excess capacity is greatest for the US$ 40/investment models since imports are made. Excess capacity is at a minimum for the no import models.

Excess capacity in period five must be zero because there is no reason to construct capacity ahead of demand during the last period of the model. Average excess capacity is largest for the US$ 45/investment model with demand II. Average excess capacity amounts to 8.5% for this model. With the higher import price and the larger demand, it is cost-saving to construct capacity ahead of demand, rather than import small quantities.

In summary, the major conclusions of this study are:

(1) A plan similar to the investment model should be adopted to determine investment in the cement industry. Sufficient capacity should be constructed to supply future increases in demand in the interior, whereas coastal cities should be supplied by imports.

(2) If, for other reasons, such as restraints on the balance of payments, a no import plan is adopted, it can be seen from our solutions that the total increase in costs over the investment model

will be from .7 to 2.6%. Most of this increase is due to increases in transportation costs, since cement is hauled from plants in the interior to coastal cities.

(3) Increases in demand in the state of Rio Grande do Sul, where limestone deposits are being rapidly depleted, should be satisfied by importing clinker or by shipping clinker from São Paulo and grinding it at Canoas.

(4) Due to economies of scale in investment costs, future cement plants should be built at maximum capacity (2 640 thousand tons of clinker per year).

(5) When limestone quarries are depleted in the states of Rio de Janeiro and São Paulo, then clinker should be produced in Minas Gerais and shipped to São Paulo and Rio for grinding.

(6) Cement is not a promising export product since no limestone quarries are located on the coast.

References

Efroymson, M. A., and Ray, T. L. "A Branch-Bound Algorithm for Plant Location." *Operations Research* 14 (May-June 1966): 361-67.

Feldman, E., Lehrer, F. A., and Ray, T. L. "Warehouse Location under Continuous Economies of Scale." *Management Science* 12 (May 1966): 670-84.

Kendrick, D. *Programming Investment in the Process Industries: An Approach to Sectoral Planning.* Cambridge: The MIT Press, 1967.

Kendrick, D., and Stoutjesdijk, A. "The Planning of Industrial Investment Programs: A Methodology." Washington, D. C.: Development Research Center, International Bank for Reconstruction & Development (IMRD), 1975 (Mimeographed).

Manne, A. S. *Investments for Capacity Expansion: Size, Location, and Time Phasing.* Cambridge: The MIT Press, 1967.

Manne, A. S., and Markowitz, H. M., eds. *Studies in Process Analysis: Economy-Wide Production Capabilities.* New York: John Wiley & Son, 1963, pp. 136-57.

Meeraus, A., Stoutjesdijk, A., and Weigel, D. "An Investment Planning Model for the World Fertilizer Industry." Washington, D.C.: IBRD, 1976 (Mimeographed).

Pratten, C. F. *Economies of Scale in Manufacturing Industry*. Cambridge: Cambridge University Press, 1971, pp. 3-28, 90-95.

Revelle, C., Marks, D., and Liebman, J. C. "An Analysis of Private and Public Sector Location Models." *Management Science* 16 (July 1970): 672-87.

Silberston, A. "Economies of Scale in Theory and Practice." *Economic Journal* 82 (March 1972): 369-91.

Sindicato Nacional da Indústria do Cimento (SNIC). *47 Anos de Indústria: 20 Anos de Sindicato*. Rio de Janeiro, 1973.

Book Review

Foundations of Brazilian economic growth

Syvrud, Donald. *Foundations of Brazilian Economic Growth*. Stanford/Washington, D. C.: Hoover Institution Press and American Enterprise Institute for Public Policy Research, 1974. 295 pp.

Eustáquio J. Reis •

In this study of recent Brazilian economic growth up to 1972, Syvrud attempts to identify the factors that enabled the economy to overcome the crisis of the mid-sixties and to register the "outstanding growth performance" witnessed from then until the mid-seventies. He also strives to assess the capacity of the economy to sustain this expansion, for "given the history of boom and bust, there is a degree of urgency to the question of how long the present ... boom will continue" (p. 3), and, in a broader context, to draw lessons for other economies faced with the problem of maintaining accelerated growth.

Such an undertaking requires that the author be extremely familiar not only with the state of affairs in Brazil, but also with the instruments of economic policy. Syvrud unquestionably meets the first requirement, since he represented the U. S. Treasury at the American Embassy in Rio de Janeiro from 1965 to 1969. During these first years of stepped-up growth, the country was "benefiting extensively from United States aid [and] nearly all relevant data passed regularly over Dr. Syvrud's desk" (p. xv). Nonetheless, the work can hardly be considered a major contribution to the understanding of the recent successes and failures of the Brazilian economy. Its official tone leaves the reader with the impression that he has perused a consular report: replete coverage of the several areas of economic policy, of the principal measures and innovations, of their rationale and impact — but no in-depth analysis.

The book is divided in two parts. The first surveys the development guidelines followed from 1947 to 1972. After an introduction in which the general questions and premises are presented, chapter 2

• Of the research institute of IPEA.

briefly reviews the ideologies and policies of the pre-1964 period, while chapter 3 covers, in somewhat more detail, those of the years 1964-1972. The second part of the work enters into particulars concerning the aims, instruments and performance of economic policy in specific fields. Thus, each of the remaining eight chapters is dedicated to a given area: monetary and credit policy, interest-rate policy, fiscal policy, wage policy, foreign-exchange and trade policy, agricultural policy, coffee policy, and, by way of conclusion, a final chapter on the policies related to the mobilization and allocation of savings. In each section, the effectiveness of the pre- and post-1964 measures is compared.

To place the work in proper perspective, it is necessary to recognize that Syvrud's approach is essentially technocratic. For him, questions of economic development can be reduced to the wisdom or error of the measures employed. Economic growth and therefore the answers to the inquiries that motivated the study rest in "political stability, sound economic management, the ability of policy makers to diagnose accurately problems in the context of a market oriented development strategy, and, finally, their ability to react promptly to these problems with an effective mix of policy instruments" (p. 3). Hence, "the economic growth of the 1950's, the ensuing collapse of the early 1960's, and the recovery and boom of the late 1960's can all be explained in terms of the effectiveness or ineffectiveness of domestic Brazilian economic policies" (p. 3).

It should be noted that economic policy is narrowly defined in the textbook sense of choosing and manipulating instruments to achieve goals that are in a way exogenous to the problem. At no point does the author try to place policy decisions or their impact in a broader context. The international situation in which decisions are implemented is practically overlooked, as is the domestic power structure. Even the discussion of the economic structure of the country, as well as its evolution over the period analyzed, leaves much to desire. Idle capacity, import structure, income distribution and consumption patterns are mentioned only in passing. Significantly, in this study of an economy in which import substitution has served as a key industrial development strategy, not one of the 65 tables refers to industrial structure or the import list.

Unlike the writers of textbooks, however, Syvrud, despite his "purely economic" and technocratic approach, does not utilize a theoretical framework capable of unifying his discussion and showing the cause-and-effect relationships between policy instruments and goals. The resulting impressionism especially weakens the first part of the book. In the introduction, the author identifies the bases of Brazilian economic growth in the period 1964-1972 as political

stability, the continuity of economic policy, a strategy for reconciling market mechanisms and planning, and a development ideology founded on the "challenge" to settle the hinterland.

According to Syvrud, "political stability" was important because of the "degree of freedom and continuity not available to economists in democratic countries. With delegated responsibilities for economic policies and the capability of avoiding the time-consuming, cumbersome channels of democratic politics, the economic authorities were free to innovate with new approaches to old problems and, through a process of trial and error, were able eventually to stabilize and restructure the Brazilian economy" (p. 4). Although the statement is not wholly false, the way the interested parties were involved in the planning and implementation of these policies demands closer attention.

In chapters 2 and 3, the forementioned lack of a theoretical framework severely biases the comparisons of the pre- and post-1964 ideologies and strategies, especially since the author is constantly concerned with justifying the diagnoses and decisions of the latter period. On the one hand, he argues that pre-1964 economic policy aggravated the situation despite the high rates of industrial growth: "These policies led almost inexorably to economic stagnation, international bankruptcy, runaway inflation and an intensified concentration of income" (p. 29). On the other hand, in examining the post-1964 period, he focuses on the difficult options facing the policymakers, and stresses the brilliant performance of this "outstanding group of professional economists" (p. 33).

With respect to the second part of the book, the decision to devote each chapter to a single area of economic policy undermines the unity of the work, leads to repetition, and makes it difficult for the reader to grasp the overall situation (the policies related to the financial market are dealt with in at least three chapters, for instance). Another point that should be mentioned is the emphasis on growth and inflation control. While the importance of these goals cannot be denied, especially since 1964, satisfactory evaluation of the economic policy of the period would require that the priorities themselves be questioned.

In terms of content, the main thread is the interventionist syndrome that guided pre-1964 policy and which, according to the author, prejudiced the allocation of resources and thus worsened inflation, the balance-of-payments disequilibria, the concentration of income, and necessarily led to economic stagnation. It is contended that the economic liberalism adopted since 1964 has improved resource allocation, and has thereby led to more balanced growth and even to social justice. The remaining interventionism is ascribed

to pragmatism, or realistic handling of the idiosyncrases of the Brazilian economy. An example is provided by the discussion of monetary policy in chapter 4. The interest and credit controls of the years 1953-1963 are said to have diminished the credit available to the private sector (as a percentage of total money supply and/or GDP), resulting in government control of credit, denationalization of industry, high costs of financial intermediation, and negative effects on income and investment levels. The trend is seen to have reversed in 1964-1972. The new approach was supposedly fundamental in renewing growth, despite its implying a readier acceptance of inflation. After all, even inflation has its pros since it facilitates credit control and the allocation of financing to priority sectors, as well as the maintenance or increase of state participation in industrial assets. This line of reasoning, based on nationalistic motivations, is hard to accept, especially for a period marked by denationalization. Also to be noted is the different manner in which the author regards the consequences of credit restrictions in the two periods.

Interventionism is again attacked in the section dealing with interest-rate policies, which are considered to have been the cause of stagnation and inflation. In turn, it is not inflation itself, but the mechanisms introduced to contain inflation, that are held responsible for the disequilibria and distortions observed in the financial market prior to 1964. It is against this background that the post-1964 innovations, including indexation, are analyzed. Unfortunately, Syvrud offers no major insights, and bypasses relevant questions such as the inflationary effects of indexation and its impact on the financial capacity of the private sector.

Like the preceding section, chapter 6, which concerns fiscal policy, is essentially descriptive. The composition of government revenues and expenditures and the role of the public sector in the Brazilian economy are merely surveyed, with no inquiries into problematic issues.

The discussion of wage policy in chapter 7 most clearly reveals the shortcomings of the analysis and the bias of the author. His major conclusion (in fact, his premise) is that "wage determination in developing, as in developed nations, should preferably be left to market forces" (p. 147). However, nowhere in the chapter is there concrete backing for this assertion. What is offered instead is a series of unfounded opinions. Syvrud contends, for example, that the drop in real wages in 1964-1967 was necessary to restore the economy to equilibrium, since "the 1964 level of wages was, *as it appeared to be,* above sustainable market levels, given the supply of labor, labor productivity, and the demand for various labor skills" (p. 161, italics mine). Or, to cite another passage, "gains in real income by the urban industrial workers during the 1950's were in part disadvanta-

geous to rural workers; however, some of the losses in real income by urban industrial workers since 1964 have benefited the rural workers" (p. 164). The author immediately excuses himself for not having included more precise information and attributes this to lack of data. Bargaining power and union activities are completely neglected, perhaps because they are not considered determinants of market forces.

The chapters reviewed should give a sufficient idea of the study. Of those remaining, special mention should be made of chapter 8, in which the balance-of-payments problems and the foreign-exchange and trade policies are detailed.

By way of conclusion, we might contemplate Syvrud's final remarks. "The Brazilian experience shows that a country can overcome defeatist traditions and ideologies and can develop if it applies sound economic policies to that end" (p. 275). "The lesson is that policy changes should be initiated as soon as the need arises; the longer the delay, the greater the change necessary, and the greater the economic, social, and political consequences, and, therefore, the greater the opposition from vested interest to such changes" (p. 274). To reflect on such conclusions is to reinforce the impression that the work is an extended eulogy to technocratic pragmatism.

Brazilian Economic Studies

BES n.º 3 — Edited by Fernando Resende

Government Policy and the Economic Growth of Brazil, 1889-1945, by Wilson Suzigan and Annibal V. Villela. (1) Introduction. (2) Long-Term Trends, 1889-1945. (3) The Monetary Crisis of the Early Republic and Economic Recovery, 1889-1913. (4) The Foreign-Trade Crisis and the Impact of World War I, 1913-1918. (5) The "Heyday" of the Export Economy, 1919-1928. (6) The Great Depression and the Stagnation of Real Income, 1929-1939. (7) The Growth of Real Income During World War II, 1940-1945. (8) Conclusion.

BES n.º 4 — Edited by Fernando Rezende

Brazilian Economic Policy in the Mid-Seventies, by Dionísio Dias Carneiro Netto — *Real Estate Investments and Financial Markets*, by João Sayad — *Dualism in the Urban Labor Market*, by Hamilton C. Tolosa — *Urban Unemployment in Brazil*, by David E. Goodman and Daniel R. Oliveira — *Brazilian Public Foreign Debt Policy, 1931-1943*, by Marcelo de Paiva Abreu — *Production, Employment and Agrarian Structure in the Cacao Regions of Bahia*, by Gervásio C. Rezende — Book Review: *Tyler, William G. — Manufactured Export Expansion and Industrialization in Brazil*, by Carlos von Doellinger.

IBGE
Centro de Serviços Gráficos