

ECONOMIA APLICADA

1999, 3(2)

USP
BCE

P19

Sumário

ARTIGOS

**Mercado relevante na análise antitruste:
uma aplicação do modelo de cidade linear 181**

César Mattos

**Trade and interdependence in the economic growth process:
a multiplier analysis for Latin America 205**

Eduardo A. Haddad, Geoffrey J. D. Hewings, Michael Sonis

**Rendimentos perdidos por trabalhadores
em condições inadequadas de saúde 239**

Ana Lúcia Kassouf

Eficiência na produção agrícola paulista e seus determinantes 263

José R. Vicente

Sazonalidade em séries temporais quadrissemanais – o caso do IMEC 289

Eliezer Martins Diniz

COMO EU PESQUISEI

My life philosophy: policy credos and working ways 309

Paul A. Samuelson

Nicholas Georgescu-Roegen about himself 321

Nicholas Georgescu-Roegen

Revista Economia Aplicada/Departamento de Economia da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo e Fundação Instituto de Pesquisas Econômicas.

--v. 3, n. 2 (1999)-

--São Paulo: FEA/USP-FIPE, 1999-

Trimestral

ISSN 1413-8050

1. Economia. I. Universidade de São Paulo. Faculdade de Economia, Administração e Contabilidade. Departamento de Economia. II. Fundação Instituto de Pesquisas Econômicas.

CDD - 330

Mercado relevante na análise antitruste: uma aplicação do modelo de cidade linear*

César Mattos[§]

RESUMO

O objetivo deste artigo é desenvolver um modelo simples que formalize a idéia de “mercado relevante” utilizada como unidade fundamental de análise nas investigações das agências antitruste. O modelo utilizado é uma adaptação do modelo de cidade linear de Hotelling. O mesmo permite aprofundar a discussão da definição de mercado relevante e suas implicações para a aplicação aos casos concretos da análise antitruste, estabelecendo quais as variáveis importantes para esse exercício. Sendo um conceito ainda pouco explorado pela literatura teórica, verifica-se que sua aplicação concreta pelas agências de concorrência, incluindo o Brasil, compreende, indistintamente, abordagens estáticas e dinâmicas, sem atenção às diferentes implicações delas decorrentes, o que deriva das próprias falhas no rigor conceitual. O artigo procura demonstrar que uma e outra abordagem podem gerar conclusões distintas, especialmente quanto ao papel dos preços absolutos e relativos na delimitação do mercado relevante.

Palavras-chave: mercado relevante, análise antitruste, modelo de cidade linear de Hotelling.

ABSTRACT

The target of this article is to introduce a simple model that formalizes the concept of “relevant market” used as the most important unity of analysis in the antitrust investigations of the foreign competition agencies around the world. The model is adapted from the linear city model of Hotelling. The model allows a more rigorous discussion about relevant market definition and its implications to real world analysis, defining the most important variables to be addressed. As this concept is poorly explored in the theoretical literature, there is a wide range of interpretations based on static and dynamic approaches on real world applications, without enough attention to the different implications involved, which reflects lack of conceptual strength. This article shows that the use of one or other approach brings different conclusions mainly those related to the role of absolute and relative prices on relevant market definition.

Key words: relevant market, antitrust analysis, linear city model of Hotelling.

§ Doutorando em economia pela UNB/DF e mestre em economia pela PUC/RJ.

* Gostaria de agradecer aos comentários de dois pareceristas anônimos da revista *Economia Aplicada* que em muito aperfeiçoaram o conteúdo deste artigo. Os eventuais erros remanescentes são de exclusiva responsabilidade do autor.

1 Introdução

Os objetivos deste artigo são:

1. aprofundar a discussão da definição de mercado relevante e suas implicações para a aplicação aos casos concretos da análise antitruste, estabelecendo, a partir de uma adaptação do modelo de cidade linear de Hottelling, quais as variáveis importantes para este exercício;
2. destacar as diferenças existentes na definição de mercado relevante a partir de uma perspectiva estática *vis-à-vis* uma dinâmica. Essa análise parte da constatação de que algumas agências antitruste, incluindo o Brasil, utilizam, indistintamente, abordagens estáticas e dinâmicas, sem uma clareza de que as mesmas embutem implicações distintas, especialmente no que tange a preços absolutos e relativos dos produtos.

Esses dois objetivos serão abordados conjuntamente ao longo do estudo. A próxima seção apresenta um resumo introdutório acerca da importância do conceito de mercado relevante na análise antitruste. A seção 3 explicita a distinção entre abordagens estática e dinâmica. A seção 4 apresenta o modelo de cidade linear de Hottelling. A seção 5 aplica a abordagem estática neste modelo, enquanto a análise dinâmica fica reservada para a seção 6. A última seção destina-se às conclusões.

2 O conceito de mercado relevante na análise antitruste

O conceito de mercado relevante é uma das peças centrais da moderna teoria de defesa da concorrência. Constitui o ponto de partida da análise das autoridades antitruste no que tange à avaliação do escopo dos problemas concorrenciais no âmbito tanto de atos de concentração como de processos envolvendo abuso de posição dominante. Trata-se de um conceito nascido da experiência americana em sua legislação antitruste, sendo ainda pouco explorado pela academia no que diz respeito a suas hipóteses subjacentes e implicações teóricas. Conforme Werden (1992):

“Market delineation is a critical stage in the structural analysis employed in many antitrust cases to help assess actual or potential market power. Market delineation plays a central role in cases involving mergers and acquisitions challenged under section 7 of the

Clayton Act and in many cases arising under sections 1 and 2 of the Sherman Act”¹

Como primeira aproximação, descrever o mercado relevante implica responder à seguinte questão: **que produtos fabricam e onde** estão localizados os agentes que **determinada firma** considera quando define as variáveis fundamentais de sua atuação no mercado. Ou seja, quem são os agentes que determinado empresário observa como seus concorrentes, influenciando ou constringendo, portanto, sua política de preços, investimento, estratégia comercial, dentre outras.²

Note-se que a menção “aos produtos que fabricam e onde estão localizados os agentes que **determinada firma** considera” no parágrafo anterior não é causal, pois o conceito de mercado relevante deve ser aplicado de forma individualizada, construído a partir de um agente específico. Se o mercado relevante for construído a partir de outro agente, o primeiro poderá ser completamente diferente, o que explica por que empresas muito similares são tratadas de forma distinta na análise das agências. No caso de um ato de concentração, o mercado relevante é próprio apenas àquelas empresas específicas que estão se juntando. No caso de abuso de posição dominante, o mercado relevante é específico àquela empresa na qual recaem as acusações.

É a partir desse mercado relevante construído que irá se definir o *market-share* de cada um dos agentes e, por conseguinte, os presumíveis efeitos anticoncorrenciais de uma concentração ou as possibilidades de abuso por parte de determinada empresa. No caso de um ato de concentração, quanto maior o número de empresas incluídas no mercado relevante,

1 De acordo com Viscusi *et alii* (1995 p. 62) *The major federal antitrust statute in the United States, the Sherman Act of 1890, was the political reaction to widespread growth of large-scale business combinations, or trusts, formed in the 1880's..... There are two main sections of the Sherman Act. Section 1 prohibits contracts, combinations, and conspiracies in restraint of trade.Section 2 prohibits monopolization, attempts to monopolize, and combinations or conspiracies to monopolize”...As a result of dissatisfaction with the Sherman Act during the first few decades, two additional statutes were enacted in 1914. The Clayton Act was designed to define anticompetitive acts more clearly...The second statute passed in 1914 was the Federal Trade Commission (FTC) Act. These three laws - the Sherman Act of 1890 and the Clayton and FTC acts of 1914 - comprise the substantive framework for US antitrust policy.” A redação completa dessas leis encontra-se em Viscusi (1995, p. 69/71). Duas discussões sobre as leis americanas e lições para o caso brasileiro encontram-se em Santacruz (1998) e Salgado (1997).*

2 A análise de mercado relevante recai, com grande frequência, na concorrência via preços. No entanto, a análise pode ser estendida também para outros tipos de concorrência, como a procedida via qualidade. Santacruz (1998) argumenta que essa é uma diferença importante entre a prática das agências de concorrência européia e americana. Enquanto as primeiras tendem a considerar outras estratégias além de preço, na definição do mercado relevante as americanas concentram-se mais na questão de preços. Para simplificação do argumento, consideraremos apenas a dimensão “preço” neste artigo.

menor o *market-share* das empresas fundentes e, por conseguinte, menores, em média, os impactos negativos presumidos sobre a concorrência.³ Analogamente, no caso de condutas abusivas, quanto maior o número de empresas incluídas no mercado relevante da empresa acusada, menor o seu *market-share* e, por conseguinte, menor, em média, a capacidade presumida de a mesma empreender comportamentos anticompetitivos.⁴

Enfim, o exercício de delimitação do mercado relevante pode incluir tanto a análise da demanda do produto como o lado da oferta. Nesse último caso, suponhamos duas firmas, “1” e “2”, cada uma produzindo um bem que não seja substituto um do outro. A delimitação do mercado relevante pelo lado da oferta implica avaliar a capacidade da firma 2, por exemplo, de reformatar suas plantas com o mínimo de custos afundados (*sunk-costs*) possível, visando ofertar no mercado da firma 1, caso os preços dessa última aumentem. Se essa reestruturação produtiva puder ser realizada de forma rápida (prazo mínimo de 1 ano) e suficientemente alta, pode-se incluir a firma “2” no mercado relevante da firma “1”.⁵ Neste artigo, no entanto, nos restringiremos a analisar a delimitação de mercados relevantes estritamente pelo lado da demanda.⁶

3 Mercado relevante: análises estática x dinâmica

A aplicação do conceito de mercado relevante aos casos concretos tende a misturar elementos de natureza mais estática com outros de natureza mais dinâmica, sem levar em

3 Um mercado relevante mais estrito não obrigatoriamente torna a análise de atos de concentração mais rigorosa. Se a delimitação de um mercado relevante for tão estreita que mesmo as empresas fundentes não possam ser consideradas no mesmo mercado relevante uma da outra, então a sua união não afetará os *market-shares* de ambas em seus respectivos mercados relevantes, o que implica não haver efeitos anticoncorrenciais.

4 A análise isolada do mercado relevante de uma operação ou de uma empresa acusada de abuso está ligada a um típico elemento de análise estruturalista nos moldes do paradigma estrutura-conduta-desempenho, no qual o *market-share* das empresas e o grau de concentração do mercado constituem as variáveis determinantes. É óbvio que o exercício do mercado relevante é apenas uma parte da análise que incorpora outros desenvolvimentos da teoria de organização industrial na análise antitruste, inclusive os provenientes da escola de Chicago. Para uma taxonomia das correntes antitruste e suas principais características, ver Khemani (1995) e Mattos (1998).

5 Um caso interessante na jurisprudência do CADE acerca do uso deste conceito é o do ato de concentração CIBA/Sandoz nos segmentos de herbicidas, inseticidas e fungicidas. A despeito de um fungicida não matar insetos e nem um inseticida eliminar fungos, o que implica não haver substitutibilidade pelo lado da demanda, ambos podem ser produzidos com as mesmas máquinas, com baixos custos de transformação. Assim, os fungicidas e inseticidas foram considerados no mesmo mercado relevante, dada a substitutibilidade do lado da oferta. Já com relação aos herbicidas, essa flexibilidade não era possível, o que implicou tratá-los como um mercado relevante à parte.

6 Como acentua Callison (1993), os EUA tendem a definir o mercado relevante apenas do ponto de vista da substitutibilidade do lado da demanda, sendo o lado da oferta transferido para o estágio posterior da análise antitruste relativo às barreiras à entrada. Já na Europa, a tendência é incluir na própria delimitação do mercado relevante a ótica da oferta. Para maiores detalhes técnicos dos condicionantes desta discussão, ver Gheventer (1998).

consideração que os mesmos não obrigatoriamente serão coincidentes a partir de uma análise teórica mais rigorosa. A questão fundamental reside na identificação do conceito de substitutibilidade com o de elasticidade-preço cruzada da demanda entre os bens. Sendo assim, iniciamos essa seção procurando apresentar as diferenças entre as duas abordagens em um plano mais intuitivo tanto na delimitação do mercado relevante de produto quanto no geográfico.

Mercado relevante de produto

Análise Estática: Consideram-se dois bens no mesmo mercado relevante de produto se os mesmos forem considerados substitutos, dadas as suas características e seu uso, **independente da elasticidade-preço cruzada da demanda** entre os mesmos;

Análise Dinâmica: Consideram-se dois bens no mesmo mercado relevante de produto se os mesmos forem considerados substitutos, dadas as suas características e seu uso, o que, obrigatoriamente, se reflete em uma **elasticidade-preço cruzada da demanda suficientemente alta**. Ou seja, a análise dinâmica demanda que a relação de substitutibilidade entre os bens resulte em uma elasticidade-preço cruzada da demanda significativa. Note-se que, nesse caso, é fundamental que, para um bem “A” estar no mesmo mercado relevante de um bem “B” a variação provável da quantidade vendida de “B” em resposta a variações de preços de “A” seja suficientemente alta. Em outras palavras, é fundamental que a **elasticidade-preço cruzada da demanda de “B” em relação ao preço de “A” seja suficientemente alta.**⁷

Mercado relevante geográfico

Análise Estática: Uma região “A” estará no mesmo mercado relevante geográfico da região “B” quando o **fluxo atual** de produtos de uma para outra for suficientemente grande. Se o fluxo atual de importações de um determinado produto for nulo, por exemplo, o mercado geográfico, nesta análise estática, nunca será mais abrangente que o território nacional.

7 Não obrigatoriamente requer-se que o contrário se verifique, ou seja, que a elasticidade-preço da demanda de “A” em relação ao preço de “B” seja suficientemente alta.

Análise Dinâmica: Uma região “A” estará no mesmo mercado relevante geográfico da região “B” quando, **independentemente do fluxo atual** de intercâmbio de produtos entre as duas, se constata que **a variação provável dos fluxos de comércio de “B” para “A”, em resposta a incrementos no preço do produto relevante em “B”, seja suficientemente alta e rápida (menos de 1 ano)**. Em uma análise dinâmica, por exemplo, mesmo que o fluxo atual de produtos importados seja nulo, cabe indagar se uma determinada variação de preços relativos implicará ou não o aparecimento de um fluxo positivo e substantivo de importações para definição de um mercado geográfico relevante nacional ou internacional.

As *Merger Guidelines* (MG) americanas para fusões horizontais de 1992, elaboradas em conjunto pela *Federal Trade Commission* (FTC) e pela seção antitruste do Departamento de Justiça Americano (DOJ), são claras em conferir uma interpretação dinâmica para o conceito de mercado relevante, utilizando um exercício de “**um aumento pequeno, mas significativo e não transitório de preços**”, para fins de análise das respostas da concorrência.⁸ Tal exercício dinâmico está relacionado, em última análise, com a magnitude das elasticidades-preço cruzadas das demandas do empresário a partir do qual se constrói o mercado relevante.^{9, 10}

Um dos principais indicadores que têm sido utilizados pelas autoridades antitruste para construir o mercado relevante tanto nos EUA como no Brasil e que, aparentemente, tenderia a ser considerado próprio a uma análise estática, são os preços relativos dos produtos. Um elevado diferencial sinalizaria menor probabilidade de um bem pertencer ao mesmo mercado relevante de outro.

Intuitivamente, essa relação é bastante razoável, pois não há como manter bens substitutos próximos com preços muito diferentes em equilíbrio. Os modelos clássicos de equilíbrio de mercado, onde se pressupõe bens homogêneos de concorrência perfeita e oligopólio (Cournot e Bertrand), por exemplo, resultam em preços iguais.

8 As Diretrizes conjuntas da FTC/DOJ entendem que a definição do que seria um “aumento pequeno, mas significativo e não transitório” seria de mais ou menos 5%. Por isso, este teste do monopolista hipotético é chamado de “teste dos 5%”

9 Como mostra Werden (1992), a primeira vez na história da jurisprudência americana em que se utilizou o critério da elasticidade-preço cruzada da demanda foi em 1953 no caso Times-Picayune. O autor também relata a história da definição teórica de mercados baseada no conceito de elasticidade-preço cruzada da demanda realizada por Bain, Machlup e Nutter.

10 Na verdade, constata-se na história da jurisprudência americana uma miríade de abordagens para a delimitação dos mercados relevantes, umas de natureza mais estática e outras mais dinâmica. Werden (1992), citando estudos de Bock e Massel de 1960 e 1962, respectivamente, procura listar, exaustivamente, todas essas abordagens. Por exemplo, a elasticidade-preço-cruzada da demanda, nesses estudos, é considerada um critério distinto do critério “diferença de preços relativos”. Procuraremos mostrar adiante que esse último está incluído no primeiro dentro de uma análise mais formal baseada no modelo proposto.

Demonstraremos que a avaliação desse diferencial no modelo proposto está relacionada com uma abordagem dinâmica e não estática, como poderia parecer à primeira vista. Ou seja, avaliar os preços relativos visando dimensionar o mercado relevante também é uma forma de analisar a elasticidade-preço cruzada da demanda, o qual constitui, por excelência, um exercício de análise dinâmica. Mais do que isso, demonstraremos nesse modelo que, dentro de uma abordagem estática, curiosamente, não cabe uma avaliação de preços relativos, mas apenas de preços absolutos.

A interpretação mais estática é consistente com definições mais próximas à Comissão de Concorrência Européia, tal como pode ser constatado nas definições constantes em documento daquele órgão de 1993:

Mercado relevante de produto (estático)

“A relevant product market comprises all those products and/or services which are regarded as interchangeable or substitutable by the consumer, by reason of the products’ characteristics, their prices and their intended use.”

Ou seja, na definição de mercado relevante de produto estática do caso europeu não se requer que a substitutibilidade implique uma elasticidade-preço cruzada da demanda significativa.

Mercado relevante geográfico (estático)

“The relevant geographic market comprises the area in which the undertakings concerned are involved in the supply and demand of products or services, in which the conditions of competition are sufficiently homogeneous and which can be distinguished from neighbouring areas because the conditions of competition are appreciably different in those areas.”

O mesmo se pode concluir para a definição de mercado relevante geográfico. É nítido que não se pode depreender dessas definições, obrigatoriamente, qualquer exercício de efeitos de variações de preços pequenas, mas significantes e não transitórias presente nas *Merger Guidelines* do FTC e DOJ, o qual representa o entendimento americano mais recente acerca do conceito e próprio de uma abordagem dinâmica. Nesse último documento, a caracterização dinâmica do exercício envolvido na definição de mercado relevante é explícita tal como se segue:

Mercado relevante de produto (dinâmica)

*“the agency will begin with each product (narrowly defined) produced or sold by each merging firm and ask what would happen if a hypothetical monopolist of that product imposed at least a **small but significant and non transitory increase in price**, but the terms of sale of all other products remained constant. **If, in response to the price increase, the reduction in sales of the product would be large enough** that a hypothetical monopolist would not find it profitable to impose such an increase in price, then the agency will add to the product group the product that is the next-best substitute for the merging firm’s product.”*

Mercado relevante geográfico (dinâmica)

*“the agency will begin with the location of each merging firm (or each plant of a multiplant firm) and ask what would happen if a hypothetical monopolist of the relevant product at that point imposed at least a **small but significant and non-transitory increase in price** but the terms of sale at all other locations remained constant. **If, in response to the price increase, the reduction in sales of the product at that location would be large enough** that a hypothetical monopolist producing or selling the relevant product at the merging firm’s location would not find profitable to impose such an increase in price, then the agency will add the location from which production is next-best substitute for production at the merging firm’s location.”*

À primeira vista, uma interpretação mais dinâmica parece captar melhor a essência do significado de mercado relevante do que um exercício estático. No entanto, é razoável postular que ambas as abordagens sejam pertinentes, a depender do problema específico que se pretende examinar: os potenciais efeitos de atos de concentração, onde uma análise dinâmica parece ser mais pertinente, ou os impactos de condutas anticompetitivas, onde uma abordagem estática pode ser a mais apropriada.

Isso ocorre porque há motivações diferentes em cada tipo de análise. No caso de atos de concentração, examinam-se quais as implicações potenciais do **aumento** de poder de mercado da nova empresa. No caso de condutas anticompetitivas, examinam-se quais as implicações efetivas de um **dado** poder de mercado de determinada empresa. Ou seja, a questão relevante

no caso de condutas é quais as possibilidades de abuso em vista desse **dado** poder de mercado, o qual depende, em última análise, do constrangimento imposto pelos outros concorrentes sobre as ações da empresa no momento presente.

A próxima seção introduz o modelo de cidade linear ligeiramente adaptado para os propósitos deste artigo.

4 O modelo de cidade linear

Inicialmente, é importante considerar que a diferenciação entre os conceitos de mercado relevante de produto e geográfico desaparecem no modelo de cidade linear. Conforme Basu (1993):

“there is a certain analogy between the economics of location and the economics of product brands. This was evident to Hotelling (1929) who observed that the problem of two firms selling a homogeneous good at two different locations on a line could, alternatively, be thought of as two firms choosing to sell cider of two different degrees of sourness from within a continuum of possibilities.”¹¹

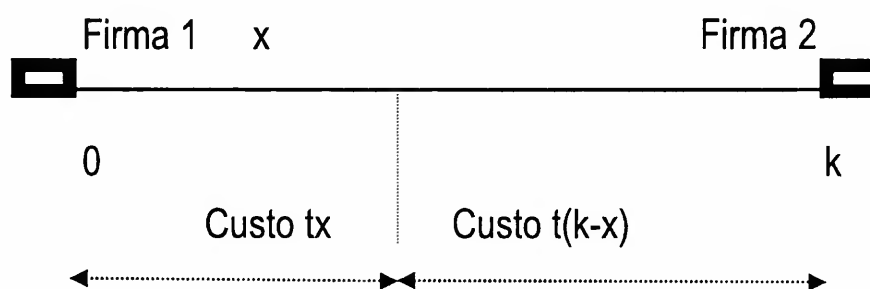
Para apresentar a estrutura do modelo de cidade linear seguiremos de perto a análise de Tirole (1988 p. 97/98). Primeiro, consideremos uma cidade linear de comprimento igual a “k”,¹² estando os consumidores distribuídos uniformemente ao longo da mesma, existindo apenas duas firmas localizadas nas duas extremidades, vendendo cada uma um bem, conforme a Figura 1 abaixo. Observe-se que “k” pode ser tomado não como uma medida da distância geográfica entre as duas firmas, mas como a diferenciação horizontal entre os dois produtos vendidos por ambas.

11 Werden (1992) enfatiza o fato de que a separação entre os dois conceitos ocorre mais por conveniência didática do que por rigor teórico. Segundo o autor: *“These two terms are convenient and very commonly used in antitrust; however, they can be misleading. There is but one relevant market with product and geographic dimensions. The product and market dimensions of relevant markets must be delineated in the context of each other, and the separation into product and geographic dimensions can be rather fuzzy.”*

12 Isso não implica dizer que um produto é “k” vezes melhor do que o outro. Para os consumidores próximos à firma 1, a utilidade líquida do produto da mesma é superior à da firma 2. O inverso é válido para os consumidores cujas preferências estão situadas próximas à firma 2. Pode-se supor que o valor de “k” constitui uma variável em que a relação geográfica e de diferenciação para cada consumidor já foi devidamente normalizada.

Além disso, os consumidores apresentam custo de transporte igual a “ t ” por unidade de comprimento ou diferenciação horizontal (o que pode incluir o valor atribuído pelos consumidores ao tempo gasto e ao esforço despendido) e demandas unitárias (ou seja, consomem uma unidade ou nenhuma). O coeficiente “ t ” para o caso de “ k ”, representando diferenciação horizontal e não distância geográfica, poderia ser interpretado como uma medida da taxa marginal de substituição entre os dois bens para cada consumidor. Assim, o tamanho da cidade linear pode estar representando “ k ” quilômetros ou “ k ” unidades de diferenciação horizontal entre os produtos. Daqui por diante, tomaremos o problema da cidade linear indistintamente como uma questão de diferenciação de produto horizontal e distância geográfica.

Figura 1
Cidade Linear com Duas Firmas



Assim, o consumidor localizado em “ x ” terá um custo de transporte tx para se locomover até a firma 1 e comprar o bem e $t(k-x)$ para ir até a firma 2.

Um ponto de fundamental importância é a distinção entre diferenciação horizontal e vertical. Enquanto a primeira capta as diferenças de gostos de cada consumidor dentro da população, a diferenciação vertical representa o elemento de concordância de todos os consumidores acerca do diferencial de qualidade entre os bens.¹³

Nesse sentido, suponhamos que “ U_1 ” e “ U_2 ” constituem as utilidades brutas dos consumidores das firmas 1 e 2, respectivamente.¹⁴ Observe-se que, para efeitos deste artigo, as utilidades brutas representam a utilidade total (antes de deduzido o preço pago) obtida pelo agente que mais obtém satisfação do consumo daquele bem ou daquele que

¹³ Para maiores detalhes da diferença entre as duas, ver Tirole (1988, p. 96-99).

¹⁴ Supõe-se neste exercício que as utilidades possuem uma representação cardinal.

esteja localizado no mesmo ponto que a firma, não tendo qualquer custo de transporte a ser deduzido. No caso do bem da firma 1, este consumidor será justamente aquele localizado no ponto “0” da figura 1, auferindo U_1 antes de descontar o preço. O mesmo vale para aquele consumidor localizado em “K” auferindo U_2 antes de descontar o preço.

Note-se que ao diferenciarmos as utilidades brutas derivadas por cada firma (U_1 e U_2), introduzimos um elemento que captura a chamada “diferenciação vertical”. Assim, o modelo aqui representado contempla ambas as fontes de diferenciação: a horizontal, medida ao longo da cidade linear, e a vertical, medida pela diferença $U_1 - U_2$ nos eixos verticais.

Sendo “ p_1 ” e “ p_2 ” os preços praticados pelas duas firmas, dados exogenamente, tem-se que o excedente do consumidor localizado em “ x ” será dado pela diferença entre esse valor, o preço e o custo de transporte:

$$\begin{aligned} &U_1 - p_1 - tx, \text{ se comprar na firma 1;} \\ &e \\ &U_2 - p_2 - t(k-x), \text{ se comprar na firma 2;} \\ &e \\ &0, \text{ se não comprar.} \end{aligned} \tag{1}$$

Se a diferença entre os preços das duas firmas não exceder o custo de transporte mais a medida de diferenciação vertical ($p_2 - p_1 < tk + U_1 - U_2$), existe um consumidor “ x_1 ” localizado entre “0” e “k” que será indiferente entre comprar nas duas firmas. “ x_1 ” será dado por:

$$U_1 - p_1 - tx_1 = U_2 - p_2 - t(k-x_1) \tag{2}$$

$$x_1(p_1, p_2) = (p_2 - p_1 + tk + U_1 - U_2) / 2t \tag{3}$$

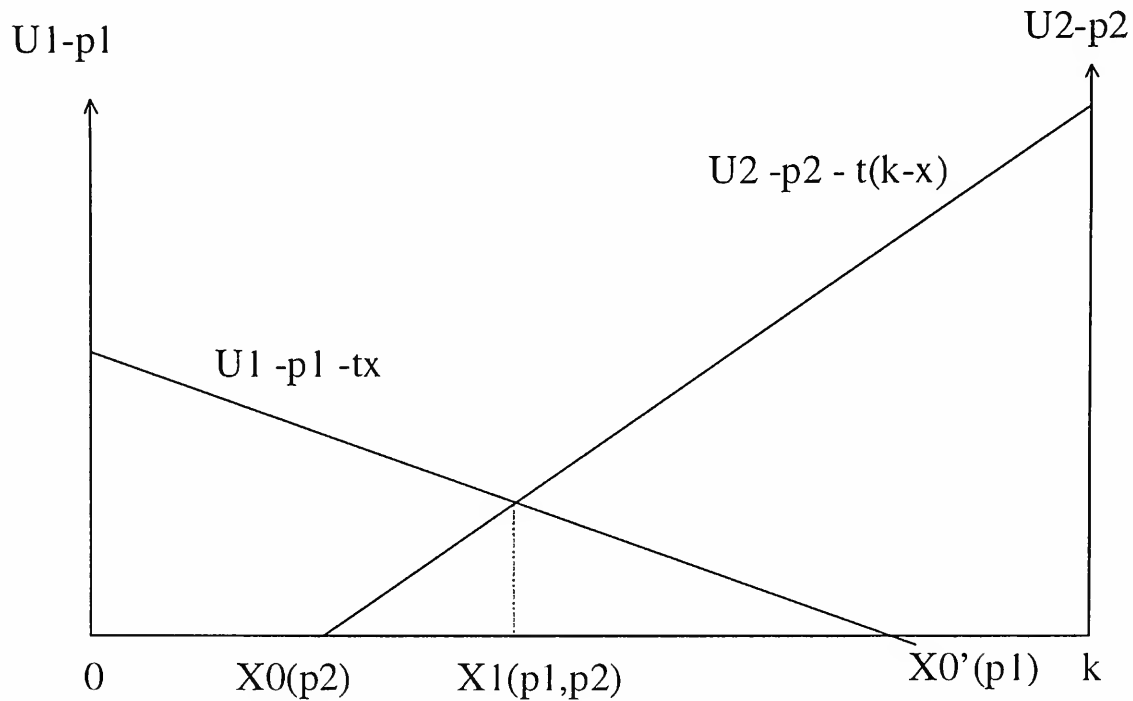
Havendo um total de “N” consumidores uniformemente distribuídos ao longo de “k”, as demandas das duas firmas no ponto de equilíbrio do modelo serão as seguintes:

$$D_1(p_1, p_2) = Nx_1(p_1, p_2) \tag{4}$$

$$D_2(p_1, p_2) = N(K - x_1(p_1, p_2))$$

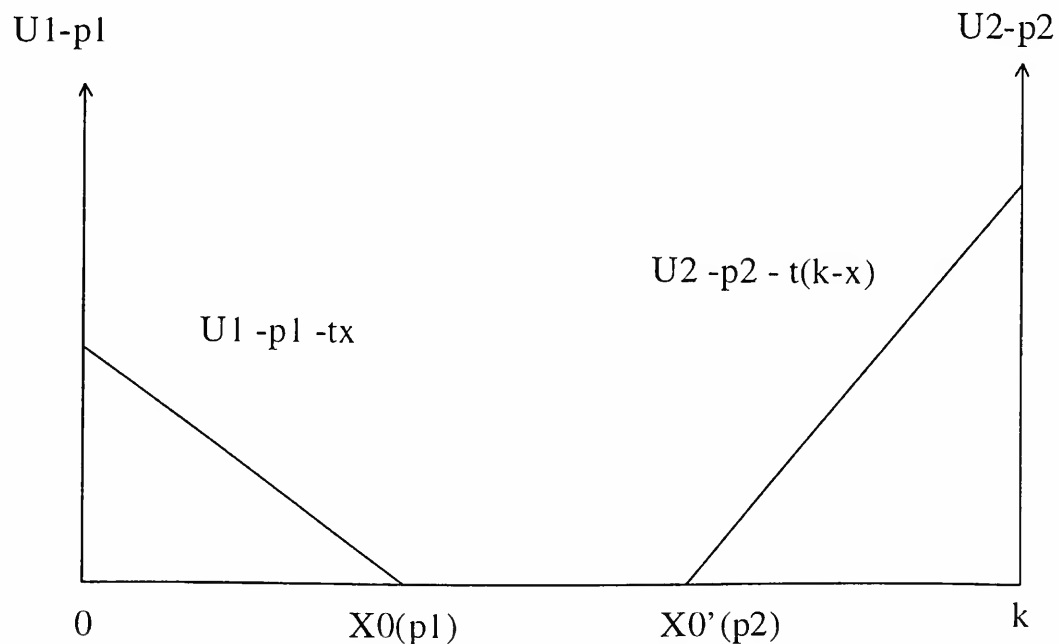
A Figura 2 mostra esse equilíbrio:

Figura 2
Modelo de Cidade Linear de Hotelling



Observe-se que a configuração mostrada na Figura 2, onde há um ponto de intersecção bem definido entre as curvas de demanda das duas firmas, pode não refletir o mundo real. Seguindo Tirole (1988), há mais duas situações importantes nesse tipo de modelo, sendo que, em ambas, as duas firmas não estarão no mesmo mercado relevante. Para nossos propósitos, exploraremos apenas uma delas, conforme Figura 3, a seguir:

Figura 3
Modelo de Cidade Linear de Hotelling O Caso de Monopólios Locais



Nesta situação, cada firma acaba detendo um monopólio local, pois dados “ p_1 ” “ p_2 ”, “ U_1 ”, “ U_2 ”, “ t ” e “ k ”, nenhuma das duas constrange o comportamento da outra. Ou seja, o único constrangimento das duas é a sua curva de demanda específica.

A situação da Figura 3 ocorre quando ambas as firmas possuem preços absolutos suficientemente altos, concentrando cada firma em seu próprio mercado cativo, tanto em termos geográficos quanto em termos de diferenciação horizontal.

5 Definição de mercado relevante: análise estática

5.1 O caso de duas firmas

Primeiramente, cabe lembrar que a definição de mercado relevante será sempre ligada a um agente econômico ou grupo de agentes em particular.¹⁵

Assim, escolhemos a firma 1 para empreender um exercício de definição matemática de mercado relevante. Inicialmente, vamos supor que apenas exista a firma 2 como concorrente em potencial da firma 1 dentro de uma perspectiva estática.

Observe que, se não fosse a firma 2, a firma 1 não teria uma demanda de apenas $D_1 = N(p_2 - p_1 + tk + U_1 - U_2)/2t$. A demanda da firma 1 seria dada por:

$$(U_1 - p_1 - tx_0) = 0$$

$$x_0 = (U_1 - p_1)/t \tag{5}$$

$$D_1(p_1,) = N(U_1 - p_1)/t \tag{5'}$$

que é maior ou igual a $D_1(p_1, p_2)$ das equações (3) e (4), dadas as hipóteses do modelo.

15 O modelo apresentado não segue à risca os mesmos procedimentos sugeridos nas MG para a delimitação de mercado relevante. Naquele caso, a construção desse mercado é feita a partir de uma firma definida, que vai agregando uma a uma as outras firmas na ordem do grau de substitutibilidade dos seus produtos com os da primeira. Esse procedimento pode subestimar o tamanho do mercado relevante, conforme observado por Shapiro (1994), o que implica tornar mais rigoroso o exame dos atos de concentração em questão ou as possibilidades de abuso de posição dominante por parte de uma empresa. Assim, no modelo proposto, pretende-se construir o mercado relevante de uma dada firma de forma simultânea com todas as candidatas potenciais e não uma a uma. Para tal, são efetuadas algumas hipóteses simplificadoras fortes visando isolar cada relação da firma a partir das quais se constrói o mercado relevante.

Defina-se x_0' como o ponto no qual a firma 2 não está constrangida pela firma 1:

$$(U_2 - p_2 - t(k - x_0')) = 0$$

$$x_0' = (p_2 - U_2 + tk)/t \quad (6)$$

De um ponto de vista estático, a zona de intersecção das duas curvas de demanda de ambas as firmas deve ser suficientemente grande, para que a firma 2 esteja no mesmo mercado relevante da firma 1 e vice-versa.¹⁶ Ou seja, a distância entre x_0 e x_0' deve ser suficientemente grande em relação ao total da distância entre as duas firmas “ k ”. Essa medida de suficiência, no entanto, não é dada endogenamente no modelo. Trata-se de um percentual que fica à discricção da autoridade antitruste definir *ex-ante*. Quanto maior a distância de x_0 e x_0' requerida por este último em relação a “ k ”, menos provável que a firma 2 esteja no mercado relevante da firma 1 e vice-versa. A distância entre x_0 e x_0' será dada por:

$$\begin{aligned} x_0 - x_0' &= (U_1 - p_1)/t - \{p_2 - U_2 + tk\}/t = \\ &= (U_1 + U_2 - p_1 - p_2 - tk)/t \geq 0 \end{aligned} \quad (7)$$

Suponhamos que a autoridade antitruste defina que para uma firma estar no mesmo mercado relevante de outra a distância mínima requerida entre x_0 e x_0' seja $Y\%$ de “ k ”. Assim, a desigualdade acima torna-se:

$$x_0 - x_0' = (U_1 + U_2 - p_1 - p_2 - tk)/t \geq Yk \quad (7'),$$

que é condição para que a firma 2 esteja no mesmo mercado relevante da firma 1 e vice-versa, dentro de uma análise estática. Assim, obtemos o primeiro resultado importante para a definição do mercado relevante de uma firma: o mercado relevante, do ponto de vista estático, será tão maior quanto mais altos os dois índices de utilidade bruta “ U_1 ” e “ U_2 ” menores os preços absolutos “ p_1 ” e “ p_2 ” menor o custo de transporte (ou menor a taxa marginal de substituição entre os dois bens) “ t ”, menor a distância geográfica (ou a diferenciação

16 Na análise estática com duas empresas, se uma firma 1 está no mesmo mercado relevante da firma 2, a recíproca também é válida, ou seja, a firma 2 sempre estará no mesmo mercado relevante de 1. Essa relação não se verifica na análise dinâmica.

horizontal), “k” ou menor o valor de “Y” arbitrado pela autoridade antitruste. Note-se que o mercado relevante do ponto de vista estático depende apenas da diferenciação horizontal e não da vertical. Ou seja, não depende das percepções comuns a todos os consumidores sobre o diferencial de qualidade entre os bens, mas apenas daquelas específicas a cada consumidor. Procederemos a uma análise mais detida dessas relações na subseção 5.3 adiante.

5.2 O caso de várias firmas e cidades lineares: premissas

Com a hipótese de que apenas existem duas firmas, vários fatores complicadores relacionados à aderência do modelo ao mundo real são eliminados. Particularmente, se a firma 2 está no mercado relevante da firma 1, a recíproca é verdadeira e esta última também estará no mesmo mercado relevante da firma 2. Introduzindo mais firmas ao longo da cidade, essa afirmação não é mais válida. Mais ainda, pode-se introduzir firmas fora da reta que definimos para a cidade, mas que estejam ligadas a pelo menos uma das firmas. Nesse contexto, a relação entre as firmas 1 e 2 não vai depender apenas das condições estabelecidas na cidade linear que as liga, mas nas várias relações entre essas duas e todas as outras e entre aquelas. Dentro desse instrumental de cidade linear, o problema matemático ficaria muito complicado, mas sem ganho adicional de poder explicativo.

Dessa forma, visando abrir o modelo para um número maior de firmas, adotaremos algumas hipóteses simplificadoras, a seguir:

1. A firma 1 está na extremidade de várias cidades lineares designadas 1-j (cidade linear entre a firma 1 e a firma j). No extremo de todas há uma outra firma da qual pretendemos verificar se pertence ou não ao mercado relevante de 1,
2. As firmas continuam produzindo apenas um bem;
3. Não há outras cidades lineares ligando essas outras firmas umas às outras, o que é uma hipótese forte. Ou seja, mesmo supondo que o produto da firma 1 é substituível pelo da firma 2 (a firma 2 está no mesmo mercado relevante da firma 1 no sentido da equação (7')) e pelo da firma 3, suporemos que **nunca** há transitividade (o produto da firma 2 não é substituível pelo da firma 3 e vice-versa). Isso decorre de uma hipótese anterior de que os consumidores nas várias cidades lineares apresentam preferências bem diferentes em relação aos das outras, sendo a única similaridade o desejo (mesmo assim diferenciado) de consumir o produto da firma 1,

4. Os custos de transporte (ou de substituição) entre a firma 1 e cada uma das firmas “j” não obrigatoriamente serão iguais, sendo que designaremos cada um de “tj”;
5. Não há comunicação entre os consumidores de duas cidades, impossibilitando o exercício de arbitragem, podendo haver diferenças nos preços da firma 1 entre as cidades. O preço da firma 1 será dado por “p1j” na cidade linear “j”, enquanto o da firma “j” será dado por “pj”.
6. O bem oferecido pela firma 1 nas várias cidades é o mesmo;
7. A autoridade antitruste fixou como distância mínima entre x_0 e x_0' um valor de “Ykjkj” para que as firmas estejam no mesmo mercado relevante. Ou seja, as duas firmas devem disputar, *ceteris paribus*, um valor fixo de, pelo menos, $y\%$ do mercado total de cada cidade “j”, “kjkj”. para que sejam consideradas no mesmo mercado relevante.
8. A utilidade daquele consumidor que mais aufere utilidade por comprar da firma 1 antes de deduzido o preço na cidade linear “j” o que definimos como “utilidade bruta” será representada por U_{1j} . Já no que se refere à utilidade do consumidor que mais aufere utilidade por comprar firma “j” na cidade linear “j” será representada por “Uj”

Sendo assim, podemos elaborar um modelo com várias cidades lineares construídas entre a firma 1 e cada uma das outras firmas “j” consideradas como candidatas a integrar o mercado relevante da firma 1.

5.3 O caso de várias firmas e cidades lineares: análise estática

Tendo em vista a análise procedida acima, para o caso de duas firmas, mais as hipóteses simplificadoras adotadas, segue-se que o modelo com várias cidades lineares apenas replica as conclusões do de duas. A condição necessária para que uma firma “j” esteja no mercado relevante da firma 1 na análise estática será a seguinte:

\exists pelo menos 1 consumidor em $x_1(p_{1j}, p_j) \leq kjkj$ na cidade linear situada entre a firma 1 e a firma “j”, tal que:

$$(U_{1j} - p_{1j} - t_j x_1(p_{1j}, p_j)) = (U_j - p_j - t(k - x_1(p_{1j}, p_j))) \quad (8)$$

Suponhamos que a autoridade antitruste definiu, “*a priori*” o tamanho mínimo da intersecção entre os mercados das duas firmas “1” e “j” como “yYkjkj”. Assim, a condição necessária e suficiente para que ambas as firmas estejam no mesmo mercado relevante será:

$$(U_j + U_{1j} - p_{1j} - p_j - t_{jk}k_j)/t_j \geq yYk_jk_j \quad \Rightarrow$$

$$(U_j + U_{1j} - p_{1j} - p_j)/(y+1)t_j \geq k_jk_j \quad (9)$$

O mercado relevante (tanto de produto como geográfico) da firma 1 será formado por todo o concorrente no qual a construção de uma linha entre este e a primeira respeita à desigualdade (9), sendo “y” definido pela autoridade antitruste.

Pela análise da desigualdade acima, obtemos as variáveis que condicionam o número de agentes no mercado relevante da firma 1. Quanto menores forem **ambos** os preços das firmas 1 e j nas cidades lineares 1-j, maior o mercado relevante da firma 1. Observe-se um ponto interessante que, nessa formulação, o mercado relevante de 1 não depende apenas negativamente de seu próprio preço, mas também do preço de seus concorrentes, ambos com o mesmo sinal. Assim, está demonstrado que, nessa definição estática, o que importa não são os preços relativos, mas sim os preços absolutos dos bens potencialmente substitutos ao da firma 1.

Graficamente, pode-se averiguar que quanto maior o preço do agente 1 (2) mais deslocada estará a sua curva de demanda para a esquerda (direita), reduzindo a área do triângulo abaixo do ponto de intersecção entre as duas. Assim, se o agente 1 (2) aumenta o seu preço, estará se concentrando nos consumidores que se localizam bem próximos ao seu estabelecimento e/ou aqueles que apresentam preferências muito fortes pelo seu produto em relação ao do agente 2 (1), ou seja, àqueles que diferenciam bastante os dois bens em favor de 1 (2). Ou seja, dentro dessa interpretação do mercado relevante, um incremento de preço dos rivais reduz o mercado relevante de um determinado agente, pois aqueles acabam se concentrando em seus consumidores mais cativos, seja por causa de localização geográfica, seja por causa de suas preferências.

O ponto importante é que isso independe completamente do efeito de uma variação de preços relativos sobre a quantidade do outro agente que é próprio ao que estamos denominando de uma análise dinâmica.

Quanto maiores as utilidades dos consumidores derivadas do consumo dos bens nas cidades lineares “j” onde atuam as firmas “j”, “U_j” e “U_{1j}”, maior deve ser o mercado relevante da firma 1, pois mais empresas satisfarão a desigualdade (9), *ceteris paribus*. Essa é uma característica particular da análise estática e se baseia em argumentos similares aos do preços absolutos: quanto maior a utilidade bruta dos consumidores “U_j” e “U_{1j}”, mais deslocadas estarão ambas as curvas de demanda para a direita (firma 1) e para a esquerda (firma 2).

A equação (9) também esboça relações mais intuitivas e observáveis na prática das autoridades antitruste. Primeiro, quanto maior o custo de transporte por unidade de distância e/ou a taxa marginal de substituição entre a firma 1 e cada uma das firmas j nas cidades lineares 1- j , menor o mercado relevante da firma 1. Segundo, quanto maior a distância geográfica ou menor a substitutibilidade entre os bens da firma 1 e das firmas j nas cidades lineares “ j ” “ kj ” menor o mercado relevante da firma 1. Enfim, quanto mais exigente a autoridade antitruste for em termos da “concorrência”(no sentido estático), ou seja, maior a intersecção requerida entre os mercados “ y ”,¹⁷ menor o mercado relevante a ser definido para a firma 1.

Quais as implicações da delimitação dos concorrentes integrantes do mercado relevante, conforme o “teste” da equação (9) para a análise antitruste? Quanto maior o número de firmas integrantes do mercado relevante de duas firmas que se concentram,¹⁸ menor será, em média, o dano presumível para a concorrência daquela fusão. Se nenhuma firma passa no “teste” da equação (9), constitui-se um monopólio, que deve gerar grande preocupação das autoridades de defesa da concorrência quanto à operação de fusão, podendo a mesma ser contestada.¹⁹ No caso da análise de condutas anticompetitivas, quanto maior o número e a participação das firmas concorrentes no mercado relevante, menor, em média, a capacidade da firma acusada de realizar as práticas alegadas.

6 A definição de mercado relevante com várias cidades lineares: análise dinâmica

Em uma análise dinâmica, o importante é analisar a elasticidade-preço cruzada da demanda. Assim, cabe avaliar a sensibilidade da demanda da firma “ j ” em relação aos preços da firma “1” Ou seja, é importante observar qual a resposta da demanda em termos de transferência dos consumidores da firma 1 para a firma j em resposta a aumentos “pequenos, mas significativos e não transitórios de preços” da primeira, conforme o teste proposto nas MG americanas. Primeiramente, temos que:

17 É importante destacar que essa forma de delimitar o mercado relevante está consoante à análise de Callisson (1995) do DOJ/USA. Segundo a autora, o importante é concentrar o foco da definição do mercado relevante nos chamados consumidores marginais “*who are most likely to switch in response to changes in relative prices.*” Esses são justamente aqueles que estão no intervalo definido por “ y ”

18 Em nosso exercício, ambas se tornam uma firma só, denominada de firma 1.

19 Supondo, obviamente, que as duas firmas fundentes estejam no mesmo mercado relevante. Caso contrário, não haveria razão para a autoridade antitruste contestar a fusão, já que o *market-share* de cada uma não seria afetada em seus respectivos mercados relevantes.

$$x_j(p_{1j}, p_j) = (k_j k_j - p_j + p_{1j} - t_j k_j k_j - U_{1j} + U_j) / (2t_j)$$

$$D_j = N * x_j$$

A elasticidade cruzada da demanda da firma “j” em relação ao preço da firma 1 na cidade linear j será a seguinte:

$$(\partial D_j / \partial p_{1j}) * p_{1j} / D_j = (N / (2t_j)) * p_{1j} / N(p_{1j} - p_j + t_j k_j k_j + U_j - U_{1j}) / (2t_j) = p_{1j} / (p_{1j} - p_j + t_j k_j k_j + U_j - U_{1j}) \geq 0$$

Mais uma vez, cabe introduzir uma variável que representa o poder discricionário da autoridade antitruste. A esta cumpre definir uma elasticidade mínima acima da qual a firma j será considerada no mesmo mercado relevante da firma 1. Suponhamos que seja definido $\varepsilon > 0$, como esse valor mínimo. Sendo assim, a condição:

$$p_{1j} / (p_{1j} - p_j + t_j k_j k_j + U_j - U_{1j}) \geq \varepsilon \quad (10)$$

deve ser satisfeita para que a firma j esteja no mesmo mercado relevante da firma 1.²⁰ Observe-se que, distintamente da análise estática, a equação (10) explicita o diferencial de preços $p_{1j} - p_j$ como condicionante da definição de mercado relevante. Quanto maior o diferencial, menor o mercado relevante²¹ da firma 1. Quanto maiores o preço absoluto da firma 1 e o preço relativo da firma j em comparação à firma 1 nas cidades lineares j, maior o

20 Simetricamente é preciso que $(\partial D_{1j} / \partial p_j) * p_j / D_{1j} = (N / (2t_j)) * p_j / N(p_j - p_{1j} + t_j k_j k_j + U_{1j} - U_j) / (2t_j) = p_j / (p_j - p_{1j} + t_j k_j k_j + U_{1j} - U_j) \geq \varepsilon$ seja satisfeita para que a firma 1 esteja no mercado relevante da firma 2.

21 Alternativa ou cumulativamente, também poderia se comparar a elasticidade da demanda da firma 1 ao seu próprio preço quando a presença da firma j restringe a sua demanda na cidade linear j e quando não restringe. Essa é uma consideração importante na qual as MG não são esclarecedoras. No caso em que a presença da firma “j” restringe o comportamento da firma 1, a condição $(U_j + U_{1j} - p_{1j} - p_j - t_j k_j k_j) / t_j \geq 0$ é satisfeita. Nesse caso, a demanda do consumidor indiferente será a dada pela equação 4. A elasticidade da demanda será dada por: $(\partial x_{1j} / \partial p_{1j}) * p_{1j} / x_{1j} = (-N / (2t_j)) * (p_{1j} / N(p_j - p_{1j} + t_j + U_{1j} - U_j) / (2t_j)) = -p_{1j} / (p_j - p_{1j} + t_j + U_{1j} - U_j)$. Já no caso do em que a demanda será $D_{1j}(p_{1j}, \quad) = N(U_{1j} - p_{1j}) / t$, a elasticidade da demanda será $(\partial x_{1j} / \partial p_{1j}) * p_{1j} / x_{1j} = (-N / t) * p_{1j} / N(U_{1j} - p_{1j}) / t = -p_{1j} / (U_{1j} - p_{1j})$. Observe-se que, dentro de uma perspectiva dinâmica na análise antitruste, para uma firma 2 estar no mercado relevante da firma 1, é condição necessária que a elasticidade de demanda em 10 (a firma j constringendo o comportamento da firma 1) seja maior, em valores absolutos, que em 10' (a firma j não constringendo o comportamento da firma 1). No entanto, isso não é sempre verdadeiro. Apenas quando: $U_j \geq p_j + t_j$, a elasticidade da demanda derivada da presença da firma j será maior que na ausência, também em valores absolutos. Além disso, a autoridade antitruste pode exigir que a presença de j constringa de forma suficientemente grande a variação de preços da firma 1 para que aquela possa ser considerada no mercado relevante dessa. Ou seja, pode exigir que a diferença entre as elasticidades-preço da firma 1 com e sem a presença j seja suficientemente grande. Essa condição pode ser dada por $U_j - p_j - t_j \geq h$ sendo “h” estabelecido pela autoridade antitruste.

mercado relevante da firma 1. Isso ocorre porque quanto maior o preço (relativo) da firma j em relação à firma 1, mais consumidores estarão preferindo consumir da firma 1. Mas, neste caso, um aumento de preço dessa última tende a transferir um número de consumidores, em média, maior para a firma j. Ademais, quando o preço relativo da firma 1 já é elevado, a mesma já se concentra bastante em seus consumidores cativos e, portanto, um aumento de preço da mesma não transferirá muitos consumidores para a firma “j”²²

Enquanto na análise estática, quanto maiores os preços absolutos de ambas as firmas, menor o mercado relevante da firma 1 (ver equação (7')), a análise dinâmica apresenta outras conclusões:

Contrariamente à análise estática, quanto maior o preço absoluto da firma j, maior o mercado relevante da firma 1. Isso porque mais consumidores estarão consumindo na firma 1 e, portanto, mais consumidores poderão, potencialmente, se transferir para a firma j em resposta a um “aumento pequeno, mas significativo e não transitório” de preços da firma 1,

O efeito do preço absoluto da firma 1 sobre o seu próprio mercado relevante é ambíguo e depende da seguinte derivada:

$$\frac{\partial (p_{1j}/(p_{1j}-p_j+t_jk_jk_j+U_j-U_{1j}))}{\partial p_{1j}} = [p_{1j}-p_j+t_jk_jk_j - p_{1j}]/(p_{1j}-p_j+t_jk_jk_j+U_j-U_{1j})^2 = [t_jk_jk_j - p_j]/(p_{1j}-p_j+t_jk_jk_j+U_j-U_{1j})^2$$

Se $t_jk_jk_j > p_j$, quanto maior “ p_{1j} ”, mais amplo o mercado relevante da firma 1. Se $p_j > t_jk_jk_j$, quanto maior “ p_{1j} ” menor o mercado relevante da firma 1. O conteúdo intuitivo dessas relações é claro. Se o custo de transporte e/ou a distância (ou diferenciação horizontal) forem altos em relação aos preços das firmas j, um aumento de preço da firma 1 tornará os consumidores mais dispostos a se transferirem para as firmas j, ampliando o mercado relevante de 1. Se ocorrer o contrário, preços das firmas j altos em relação ao custo de transporte ou à diferenciação, os consumidores estarão mais resistentes a essa mudança.

Além disso, similarmente à análise estática, quanto maiores as distâncias “ k_jk_j ”, os custos de transporte (ou de substituição) “ t_j ” entre as firmas 1 e j nas cidades lineares j e o parâmetro

22 Cabe notar que a utilização de preços relativos nunca definirá um mercado relevante de forma isolada pela equação 10. Se os preços relativos pudessem definir sozinhos o mercado relevante, poder-se-ia chegar à conclusão absurda que um chiclete estaria no mesmo mercado de uma passagem de ônibus se ambos custassem o mesmo preço. No entanto, neste caso extremo, o grau de substitutibilidade é zero, o que implica que o valor de “ k_jk_j ” é próximo de infinito, nunca satisfazendo (10).

fixado pela autoridade antitruste para a elasticidade da demanda “ ϵ ”, menor o mercado relevante da firma 1.

Na análise dinâmica, a diferenciação vertical passa a importar para a delimitação do mercado relevante, o que não ocorria para o caso da análise estática. Este termo ($U_j - U_{1j}$ no denominador de (10)) revela uma propriedade importante da análise dinâmica: o mercado relevante de uma firma varia em relação direta ao elemento de qualidade de seus bens percebido por todos os consumidores como superior aos bens das outras firmas. O sentido dessa relação é que quanto maior o elemento de qualidade superior da firma 1 percebido por todos os consumidores, maior o número desses que compram, *ceteris paribus*, da mesma. Assim, haverá um número potencial maior de consumidores para reagir a um aumento “pequeno, mas significativo e não transitório de preços” da firma 1, o que alarga o seu mercado relevante.

7 Conclusões

O modelo teórico apresentado permite definir o mercado relevante de uma firma como função inversa da distância geográfica e/ou da magnitude da diferenciação horizontal entre as firmas, o que é válido tanto para a análise estática como para a dinâmica. Essas variáveis estão sempre presentes nas análises das agências de concorrência aplicadas aos casos concretos.

Por outro lado, o papel de preços relativos e absolutos das empresas na delimitação dos mercados relevantes nunca foi suficientemente bem trabalhada na análise antitruste do ponto de vista teórico. A questão dos preços relativos, por exemplo, é tratada pelas agências de forma estritamente intuitiva: quanto maior a distância entre os preços, menor a probabilidade dos consumidores considerarem um bem como substituto do outro, pois, caso contrário, não haveria lógica em se consumir o mais caro. Apesar de bastante razoável, a pura intuição, mais do que o rigor formal, pode esconder relações importantes.

Primeiro, em uma análise estática, conforme o modelo desenvolvido, não cabe uma avaliação de preços relativos, mas apenas de preços absolutos para a delimitação dos mercados relevantes. Em uma análise dinâmica, a análise de preços relativos passa a ser fundamental, mas o sentido desse diferencial também é crucial. Se na equação (10) o preço relativo da firma, a partir da qual se constrói o mercado relevante, for grande em relação a uma determinada firma, menor a probabilidade dessa última estar incluída em seu mercado relevante, o que é consoante à aplicação tradicional do conceito pelas agências de concorrência. Mas se ocorrer o contrário, ou seja, o preço relativo da outra firma ser grande

em relação ao da firma a partir da qual se constrói o mercado relevante, a relação se inverte: maior a probabilidade daquela firma estar no mercado relevante da primeira firma.

O papel dos preços absolutos também são importantes em uma análise teórica mais rigorosa da definição de mercado relevante. Esse ponto, no entanto, nunca foi devidamente analisado para a melhor compreensão do conceito de mercado relevante.

O papel distinto das noções de diferenciação horizontal e vertical também tende a ser negligenciado na análise antitruste convencional. A ênfase tende a ser na diferenciação vertical e sem uma consideração acerca do sentido da mesma, ou seja, qual empresa possui atributos de qualidade superior percebidos unanimemente por todos os consumidores. Nessa análise convencional, quanto maiores as diferenças absolutas de qualidade, menor a probabilidade dos bens estarem no mesmo mercado relevante. Na análise estática, a equação (7') demonstra que a diferenciação vertical não possui nenhum papel na delimitação do mercado relevante, sendo que o importante é a magnitude absoluta da utilidade derivada de ambos os bens.

Na análise dinâmica, o mercado relevante de uma firma varia proporcionalmente aos seus atributos de qualidade superior (U_{1j}), mas inversamente aos atributos superiores das outras firmas (U_j). Assim, cabe não só avaliar as magnitudes relativas dos atributos de qualidade, mas também em qual dos dois bens esses são maiores.

Cabe ainda enfatizar o papel relevante para a descrição da autoridade antitruste por meio da escolha do parâmetro fixado, explícita ou implicitamente, por aquela. Esse aspecto não é normalmente considerado de forma clara quando se discute o conceito de mercado relevante.

Enfim, cabe assinalar a importância das autoridades antitruste brasileiras explicitarem a base teórica a partir da qual aplicam seus conceitos, especialmente o de mercado relevante, aos casos concretos. Nesse sentido, é louvável a iniciativa do Ministério da Fazenda, no momento em que a submissão deste artigo já se encontra em fase avançada, de publicar Diretrizes para a análise de atos de concentração no Brasil com apurado rigor técnico, especialmente no que tange às definições de mercado relevante, objeto deste artigo.²⁴ Essas diretrizes certamente contribuirão bastante para aprofundar o debate do antitruste no Brasil e possibilitar a maior aproximação dos economistas nacionais com essa importante questão.

24 Constitui-se no primeiro trabalho sério com significativo embasamento econômico das autoridades antitruste brasileiras nesta questão.

Referências bibliográficas

- Basu, K. *Lectures in industrial organization theory*. Blackwell, 1993.
- www.mj.gov.br/cade
- Callisson, J. E. Issues in relevant product market definition. US Department of Justice-Antitrust Division, 1994, *mimeo*.
- Department of Justice and Federal Trade Commission. *Horizontal Merger Guidelines*. April 1992.
- European Competition Commission. Market definition under the merger regulation. 1993, *mimeo*.
- Gheventer, A. A outra lâmina da tesoura - considerações sobre a oferta na definição de mercado relevante. *Revista do IBRAC*, v. 5, n. 6, 1998.
- Khemani, R. S., Dutz, M. A. The instruments of competition policy and their relevance for economic development. Chapter 2. In Frichstak, C.(ed.), *Regulatory policies and reform: a comparative perspective*. December 1995.
- Mattos, C. C. A. The recent evolution of competition policy in Brazil: an incomplete transition. *Journal of Latin American Competition Policies*, v. I, n. 2, August 1998. Também encontrado em CERES (Centro de Estudos em Reforma do Estado) FGV/EPGE/RJ Texto para Discussão 06 Outubro 1997
- Salgado, L. H. *A economia política da ação antitruste*. São Paulo: Ed. Singular, 1997.
- Santacruz, R. *Prevenção antitruste no Brasil - 1991/1996*. Tese de doutoramento apresentada ao Instituto de Economia da Universidade Federal do Rio de Janeiro - IE / UFRJ, março de 1998.
- Shapiro, C. Mergers with differentiated products: narrow markets under the Merger Guidelines. Berkeley University, 1994, *mimeo*.
- Tirole, J. *The theory of industrial organization*. The MIT Press, 1988.
- Viscusi, W. K., Vernon, J. M. & Harrington, J. E. *Economics of regulation and antitrust*. Second Edition. The MIT Press, 1995.
- Werden, G. J. *The history of antitrust market delineation*. Economic Analysis Group Discussion Paper. US Department of Justice. Antitrust Division EAG 92-2, July 1992.

Trade and interdependence in the economic growth process: a multiplier analysis for Latin America

Eduardo A. Haddad[§]
Geoffrey J. D. Hewings[†]
Michael Sonis^α

RESUMO

Este artigo propõe algumas abordagens alternativas para a análise do papel do comércio internacional e da interdependência econômica no processo de crescimento baseadas em contribuições anteriores de Machlup, Goodwin e Miyazawa. A análise empírica enfoca os países da América Latina: a economia mundial é dividida em dois grupos principais de países (América Latina e uma seleção de países desenvolvidos), com o resto do mundo formando um terceiro bloco agregado. Uma série temporal de matrizes de comércio, para o período 1978-1991, é construída para que se possa explorar a extensão dos impactos do crescimento de um país no resto do mundo e a simetria/assimetria destes impactos. Por meio da utilização do método proposto, revela-se que *insights* importantes acerca da estrutura de comércio internacional podem ser obtidos, propiciando valiosas informações para a análise da evolução dos regimes de comércio e da formação de blocos econômicos.

Palavras-chave: comércio internacional, análise de insumo-produto, decomposição de multiplicador, América Latina.

ABSTRACT

This paper illustrates alternative methodological approaches to the issue of trade and interdependence in the economic growth process with a focus on the countries of Latin America, drawing inspiration from earlier contributions by Machlup, Goodwin and Miyazawa. The world economy is divided into two main blocks of countries (Latin America and a selection of developed economies) with the rest of the world forming an aggregated third block. A time series of trade matrices for the period 1978-1991 has been constructed to explore the degree to which changes in one country spillover to the rest of the world and the degree to which the changes are symmetric or asymmetric. The approaches reveal that important insights into trade structure can be obtained, insights that will prove of value in the rapidly changing trade regimes of the current and next decades.

Key words: international trade, input-output analysis, multiplier decomposition, Latin America.

§ FIPE, Universidade de São Paulo, Brazil, and Regional Economics Applications Laboratory, University of Illinois, USA.

† Regional Economics Applications Laboratory, University of Illinois, USA.

α Bar Ilan University, Ramat Gan, Israel, and Regional Economics Applications Laboratory, University of Illinois, USA.

1 Introduction

The relationship between trade and growth has been a familiar topic of discussion in the development literature. More often, the question posed concerns the effects of international trade on economic growth, and thus focuses on trade as an active “agent” of growth. This active role played by international trade can be found in many different models. With the neoclassical free-trade model, in one extreme, trade is regarded as an important stimulator of economic growth. It enlarges a country’s consumption capacities, increases world output, and provides access to scarce resources and worldwide markets for products without which poor countries would be unable to grow. (Todaro, 1994) By imposing limits to this theory (fixed resources, full employment, and the international immobility of capital and skilled labor; fixed, freely available technology and consumer sovereignty; internal factor mobility and perfect competition; the absence of national governments in trading relations; balanced trade and international price adjustments; and trade gains accruing to nationals), Todaro (1994) concludes that trade can be an important stimulus to rapid economic growth, although it might not be a desirable strategy for economic and social development. The contribution to development depends on the nature of the export sector, the distribution of its benefits, and the sector’s linkages with the rest of the economy. It seems that, to the extent we are only interested in the effects of international trade on pure economic growth, there is a consensus that trade can provide an important stimulus to growth.

One of the central issues to be addressed in this paper explores the degree to which growth in one country affects growth in other countries. More specifically, the role international trade plays as a mechanism of transfer of income from growing countries is explored with a focus on the degree to which developing countries benefit from the growth of developed countries. Also explored is the asymmetry problem: does the economic growth of developed countries have a higher impact on the developing economies than these economies’ growth impact on the richer countries? This paper provides a multiplier analysis of the international economic dependency among Latin American countries and developed countries. In a sense, it contributes to the debate raised by Krugman (1991, 1993), under a new perspective, on regionalism versus multilateralism, by presenting an approach to evaluate the impacts, over time, of the creation of trading blocks on trade patterns within a general equilibrium framework.

Section 2 provides a brief review of the literature on international trade multipliers. Section 3 describes the matrix of international trade (MIT), which will be used as the data base for our estimates. In Section 4, the methodology used to analyze the impact of growth through international trade is described, and Section 5 presents the empirical results. Final remarks are provided in the last section.

2 Background

The **passive** role of trade related to growth issues was studied in depth, in the first half of this century, when Keynesians started to study the income multiplier effects in a national economy (see Machlup, 1939). Machlup (1943) presented a framework in which the multiplier technique was developed in certain respects (primarily to measure the effects upon national income and trade balances), applicable to the theory of foreign trade. The idea of the foreign-induced trade multiplier was presented, in a demand-driven two-country framework, in which an increase in autonomous income in country A generated, through import leakages, an increase in the income of country B, i.e., part of the increase in the income of A would be spent on imports from country B. Part of this income transfer from A to B, in a second round, through import leakages, would return to A through A's exports to B. This process would continue until the income transfers became negligible. The total effect of the initial increase in the income of A is the foreign-induced trade multiplier effect.

There have been many, diverse methodological contributions to understanding these issues. For example, Goodwin (1983) generalized this idea by implementing input-output techniques for a ten-region world payments matrix and presented the concept of the world matrix multiplier, which was, basically, Machlup's concept of foreign-induced trade multiplier expanded to a many-country context. Prior to Goodwin's contributions, Miyazawa (1960) revised the conventional foreign trade multiplier, considering a sub-multiplier process involving production, i.e., treating the imported intermediate goods required for the production of exported goods as an endogenous factor induced by the initial injection. By taking into account the intermediate products in the circular flow, Miyazawa's foreign trade multiplier provides a more accurate indicator of the direct and indirect effects on the economy. Other approaches, based on input-output systems, have attempted to analyze the structure of multi-regional trade flows. Feedback loop analysis has been used for both interregional national input-output tables (Sonis *et alii*, 1995a), and intercountry input-output tables (Sonis *et alii*, 1993 and 1995b). However, the data requirements of analysis based on interregional and intercountry input-output tables are more demanding, and it has often proved difficult to gather all the data necessary to carry out the type of analysis that is the subject of this paper, in the context of Miyazawa's foreign trade multipliers (see Ota, 1994, for a similar view).

Hence, this paper offers a less comprehensive perspective and one that necessitates some compromises and assumptions. As a result, in this paper, induced imports are restricted to finished goods only. The first step in the data assembly involves the construction of a time series (1978-1991) of matrices of international trade for 23 countries

and the rest of the world, in order to generate more precise estimates of foreign-induced trade multipliers. The next step is to analyze the impacts of growth on trade balances and growth in other parts of the world economy. To accomplish this task, the world economy was divided into two blocks, developed countries and Latin American countries, while the rest of the world was allocated to a third block. Trade flows for individual countries were shown in the first two blocks.

3 The Matrix of International Trade (MIT)

The data used by Goodwin (1983) were taken directly from the estimates of the marginal propensities to import developed by Thorbecke and Field (1974). These estimates, however, given by the shares of merchandise imports into region i from region j , expressed in f.o.b. terms, in the GNP of region i , overestimate the multiplier effects, as will become clear after we develop the matrix of international trade (MIT) framework. Furthermore, some world regions had to be considered exogenous in Goodwin's analysis, which does not necessarily happen in the MIT framework. The development of the MIT model follows.

Consider the following balance identity, which is valid for each country i in the world economy

$$C_i + I_i + G_i + X_i - M_i \equiv Y_i, \quad i = 1, 2, \dots, n \quad (1)$$

where:

C_i is total domestic consumption of country i

I_i is total investments of country i

G_i is total government spending of country i

X_i is total exports of country i

M_i is total imports of country i

Y_i is GDP of country i

Our concern here is the flow of goods and services among countries that participate in international trade. Therefore, the next step is to disaggregate total exports and total imports in the following way.¹ Rearranging equation (1), we have:

$$X_i + C_i + I_i + G_i \equiv M_i + Y_i \quad i = 1, 2, \dots, n \quad (2)$$

where:

$X_i + C_i + I_i + G_i = Z_i$ is the total demand for domestic output of country i

$C_i + I_i + G_i = F_i$ is the total **domestic** demand for domestic output of country i (exogenous outlays)²

$M_i + Y_i = E_i$ is the total expenditure of country i

Building up a trade table, we have:

$$X_i = \sum_{j=1}^n x_{ij} \quad (3)$$

where x_{ij} are the exports of country i to country j , and

$$M_i = \sum_{j=1}^n m_{ji} \quad (4)$$

where m_{ji} are the imports of country i from country j .

Notice that $[x_{ij}]$ and $[m_{ji}]$ are the same matrix.

Given these definitions, we can propose the design of matrices of international trade (MIT), that present structural similarities to the closed-economy input-output tables. We can also think of the MITs as a special case of a representation of a single-industry-economy

1 Modeling the aggregate demand components would increase the accuracy of the model, without any profound implication for the results.

2 Domestic here refers to the origin of income which generates the demand.

multirregional input-output tables. In this case, since the proposed tables incorporate the whole world economy, there is no reason to include net exports in the “final demand” portion of the table.

Thus, we will have an input-output-type table, in which the rows describe the distribution of a country’s domestic production throughout the world economy ($\sum_{j=1}^n x_{ij} + C_i + I_i + G_i$), and in the columns we can read the composition of a country’s domestic expenditures ($\sum_{j=1}^n m_{ji} + Y_i$).

Figure 1 shows the layout of the MIT. The mathematical structure of the system consists of a set of n linear equations with n unknowns. Like the input-output systems, the solutions here are straightforward mathematically, but there are differences in the economic interpretations of some of the results.

One very important definition in our framework refers to the extended import coefficients. These coefficients - analogous to the direct requirements coefficients in the input-output framework - show the proportion of the total expenditures which is used for imports.³ An extended import coefficients matrix can be derived, in which we can read, columnwise, the proportion of the total expenditures of a country that goes to external payments of output; as such, it reveals the direct linkages among countries (trade linkages). We are assuming here that for each dollar spent in a certain country, the structure of external expenditures in a given period is indicated by these coefficients, i.e., given the conditions for trade in a given period (e.g. tariffs, transport costs, production technique or consumers’ tastes), the result of the behavior of the agents is indicated by a certain fixed proportion of expenditures at that time. In an *ex-post* analysis, the set of actual expenditures is given by the MIT. Thus, we can define the matrix of extended import coefficients as:

$$[t_{ij}] = \frac{1}{E_j} [x_{ij}] \quad (5)$$

where t_{ij} is the share of total expenditures of country j used to pay the imports from country i . In the case of countries, a typical element t_{ij} is zero, for $i = j$.⁴

³ Note that the extended import coefficient is strictly smaller than the traditional import coefficient, which is equal to imports divided by the GDP.

⁴ A country does not export/import to/from itself. In the case of groups of countries as a “sector”, t_{ij} represents the total trade among those countries.

Given the assumption of fixed proportions of total domestic expenditures in a given period, one can derive the following “closed” system:

$$\sum_{j=1}^n t_{ij} Z_j + F_i = Z_i, \quad i = 1, 2, \dots, n \quad (6)$$

In matrix notation, (6) becomes:

$$TZ + F = Z \quad (7)$$

where:

T is the matrix of extended import coefficients ($n \times n$)

Z and F are column vectors ($n \times 1$)

Solving equation (7), it is possible to obtain the total demand for domestic output (=total domestic output) necessary to satisfy the total **domestic** demand for domestic output.

$$Z = (I - T)^{-1} F \quad (8)$$

where $(I - T)^{-1}$ is the Machlup-Goodwin (hereafter, MG) foreign-induced trade multiplier matrix⁵. The existence of such an inverse matrix is assured since $(I - T)$ is nonsingular and satisfies the Hawkins-Simon conditions, i.e. the sum of the elements of each column is strictly less than one.

To understand what the MG foreign-induced trade multiplier matrix reveals, first assume an autonomous change in the exogenous outlays in a country, i.e., an autonomous change in the total domestic demand for domestic output of a country, say a \$1M increase in investments in Brazil.⁶ Earlier, it was noted that the extended import coefficients show the

5 To be more consistent with Machlup's definitions, we should call this matrix, type I foreign-induced trade multiplier matrix, since all the internal components of aggregate demand are exogenous here.

6 Quoting Machlup (1939): "There are those who feel that primitive stories are unworthy of being embodied in a scientific article. I believe, however, that if more such stories were employed by writers when they develop their arguments, they might avoid a good many pitfalls, or their critics, at least, might discover them more quickly."

direct trade linkage among the countries. However, considering this static structure of the world economy that is implied, one that is not affected by different levels of output, **hypothetically**, there would be both direct and indirect trade linkage effects related to the other countries' economies. For instance, for the \$1M increase in Brazilian investment, it is assumed that, in a given period, Brazil has an optimal fixed menu of imports from other countries. By homogeneity of degree one, this increase will generate a general increase in imports from all the other countries, which will generate income for Brazil's trade partners. In turn, this increase in their income will generate more imports in order for those countries to balance their optimal structure of expenditures. These interactions involve many rounds of spending and respending and the total effect matrix of direct and indirect trade linkages is represented by the Leontief-type-inverse, namely, the MG matrix. The process described above does not really occur, since the constraints change constantly, but it explains the *ceteris paribus* round-by-round effects in which we are interested in this context.

4 The impact of growth through international trade

As shown in the last section, the structure of the MITs resembles to a great extent closed-economy input-output tables. In this section, we will introduce some techniques pioneered in input-output analysis that can be adopted in order to characterize the structure of international trade. Although there are many similarities between input-output tables and MITs, we still have to take care when applying some of the former analytical tools to the latter framework. Since they provide a rather precious set of temporal "photographs" of the international trade, the use of MITs to characterize the structure of international trade of the world economy can reveal a reliable "portrait" reflecting relevant aspects of trade relations. The first technique presented here relates to the multiplier product matrix "landscape" derived from the MG matrix to reveal visually some of the structural changes that have occurred over time; then we explore some suggestions of Goodwin (1983), followed by Miyazawa's (1966) distinction between internal and external multipliers.

The multiplier product matrix and trade linkages

The concept of multiplier product matrix (MPM) (see Sonis *et alii*, 1994) was developed under the notion of field of influence of structure change in an economy. The basic idea draws on the way in which changes in one or more elements in a matrix affect the whole system of interrelationships. The definition of the MPM is as follows: given $T = \|t_{ij}\|$ and

$B = (I - T)^{-1}$, the first-order of field of influence associated with a change, ε_{ij} , in the i^{th} row and j^{th} column is a matrix:

$$F[(i, j)] = \begin{pmatrix} b_{i1} \\ b_{i2} \\ \vdots \\ b_{in} \end{pmatrix} (b_{1j}, b_{2j}, \dots, b_{nj}) = \|b_{ir} b_{sj}\| \quad (9)$$

Let $b_{.j}$ and $b_{i.}$ be the column and row multipliers of the Leontief inverse. These are defined as:

$$b_{.j} = \sum_{i=1}^n b_{ij}, b_{i.} = \sum_{j=1}^n b_{ij} \quad (10)$$

Let V be the global intensity of the Leontief inverse:

$$V = \sum_{i=1}^n \sum_{j=1}^n b_{ij} \quad (11)$$

Then, the input-output multiplier product matrix (MPM) is defined as:

$$M = \frac{1}{V} \|b_{i.} b_{.j}\| = \frac{1}{V} \begin{pmatrix} b_{1.} \\ b_{2.} \\ \vdots \\ b_{n.} \end{pmatrix} (b_{.1} \ b_{.2} \ \dots \ b_{.n}) \quad (12)$$

It can be shown that the MPM has a cross structure; this cross structure can be explored to reveal a hierarchy of transactions such that there exists a cross (one row and column) in which the elements of this row (column) are larger than the corresponding elements of every other row (column). If this cross is now excluded, another cross with the same properties can be identified and the procedure repeated until all the rows and columns have been arranged. This new arrangement will be conducted in such a way that the centers of subsequent crosses will appear on the main diagonal, thereby providing a descending economic landscape. The procedure can be repeated for subsequent years; by maintaining

the ordering of rows and columns from a base year, it is possible to reveal immediately structural changes in the economy (or, in the case of the trade matrix, the strength of flows between countries). If there are no changes in the hierarchical structure of exchange, the ordering from one time period will be preserved in later time periods.

Goodwin's Net Foreign Balance (NFB)

The analysis that follows, developed by Goodwin (1983), enables us to measure the effects of a change in the vector of exogenous outlays (F) on the balance of trade of each country. This reveals the extent to which growth affects the balance of trade of all the countries in the world economy. Equation (13) provides the final effect on the net foreign balance for each country, given a change in the vector of exogenous spending. Note that the net foreign balances (NFBs) sum to zero.

$$NFB = \left[[I - \lambda][I - T]^{-1} - I \right] \Delta F, \quad (13)$$

where λ is a diagonal matrix of the column sums of T

Miyazawa's internal and external multipliers⁷

Miyazawa's framework of internal and external multipliers (Miyazawa, 1966, 1971), which was extended by Sonis and Hewings (1993, 1995), will prove useful in analyzing the impact of growth of one group of countries on the rest of the world through the trade network. This framework was used by Fritz (1995), in an input-output context, to analyze the transactions between polluting and clean sectors.

This dual approach can be considered by representing the extended import coefficients matrix in the following way:

$$T = \left(\begin{array}{c|c} T_{11} & T_{12} \\ \hline T_{21} & T_{22} \end{array} \right) \quad (14)$$

7 The presentation of this section draws on Sonis and Hewings (1993, 1995), with the necessary changes in notation and interpretation.

where T_{11} and T_{22} are the square matrices of extended import coefficients within the first and second groups of countries, and T_{12} and T_{21} are the rectangular matrices showing the trade relations between the two groups. One of the possible decompositions of T is given by:

$$T = \left(\begin{array}{c|c} T_{11} & 0 \\ \hline T_{21} & 0 \end{array} \right) + \left(\begin{array}{c|c} 0 & T_{12} \\ \hline 0 & T_{22} \end{array} \right) = T_1 + T_2 \quad (15)$$

It can be shown that the following Miyazawa formula, that will be used to analyze the trade between developed countries and Latin American countries, can be obtained (Sonis and Hewings, 1993) as follows:

$$(I - T)^{-1} = \left(\begin{array}{c|c} B_1 + B_1 T_{12} \Delta_2 T_{21} B_1 & B_1 T_{12} \Delta_2 \\ \hline \Delta_2 T_{21} B_1 & \Delta_2 \end{array} \right), \quad (12)$$

where $B_1 = (I - T_{11})^{-1}$ is the internal matrix multiplier of group 1, and $\Delta_2 = (I - T_{22} - T_{21} B_1 T_{12})^{-1}$ is the external matrix multiplier of group 2.

The matrix multiplier of interest, $\Delta_2 T_{21} B_1$, reveals the influence of group 1's internal propagation on group 2's income/output levels.⁸ This matrix is a rectangular matrix, whose elements, $\delta_{i_2 j_1}$ represent the increase in income/output in country i_2 due to a unit increase in country j_1 's exogenous outlays. In order to evaluate the total amount of income/output generated in group 2 by a unit increase in a country of group 1's output, their scalar column multipliers are computed.

$$M_{j_1} = \sum_{i_2} \delta_{i_2 j_1} \quad (16)$$

where M_{j_1} is country j_1 's column multiplier with respect to all countries in group 2.

8 In our context, internal propagation refers to internal direct and indirect import demand.

The multipliers of the matrix $\Delta_2 T_{21} B_1$ result from the interaction of three multiplier matrices, Δ_{22} , B_2 , B_1 , and the matrix T_{21} , where: $B_2 = (I - T_{22})^{-1}$ is the internal matrix multiplier of group 2, and $\Delta_{22} = (I - B_2 T_{21} B_1 T_{12})^{-1}$ is the Miyazawa external matrix multiplier, such that $\Delta_2 = \Delta_{22} B_2$

Δ_{22} includes the direct, indirect and induced effects of the group 2's import demand from group 1 on group 2's income/output. The sources of income/output induced by the group 1's external demand can be unveiled by looking at the column sums of these matrices with respect to group 2.

T_{21} income/output generated by direct import requirements of group 1

$T_{21} B_1$ income/output generated by direct and indirect import requirements of group 1

$B_2 T_{21} B_1$ income/output generated by internal propagation of group 1 and the induced direct and indirect production (income increase) of sector 2

$\Delta_{22} B_2 T_{21} B_1$ total foreign-induced trade multiplier of group 1. income/output generated by internal propagation of group 1 countries and the induced internal and external propagation of group 2 countries

Country j 's column sum with respect to these matrices are denoted m_{j1}^1 , m_{j1}^2 , m_{j1}^3 , and M_{j1} respectively. The following definitions will be employed in the empirical part:

$\frac{m_{j1}^1}{M_{j1}}$ share of direct import requirements in total multiplier

$\frac{m_{j1}^2 - m_{j1}^1}{M_{j1}}$ share of indirect import requirements in total multiplier

$\frac{m_{j1}^3 - m_{j1}^2}{M_{j1}}$ share of internal propagation of group 2 countries (stimulated by group of countries 1's import requirements) in total multiplier

$$\frac{M_j - m_{ji}^3}{M_{ji}} \quad \text{share of external propagation of group 2 countries in total multiplier}$$

In the next section, these methods will be employed to unravel some of the important structural changes in the MIT.

5 Summary of results

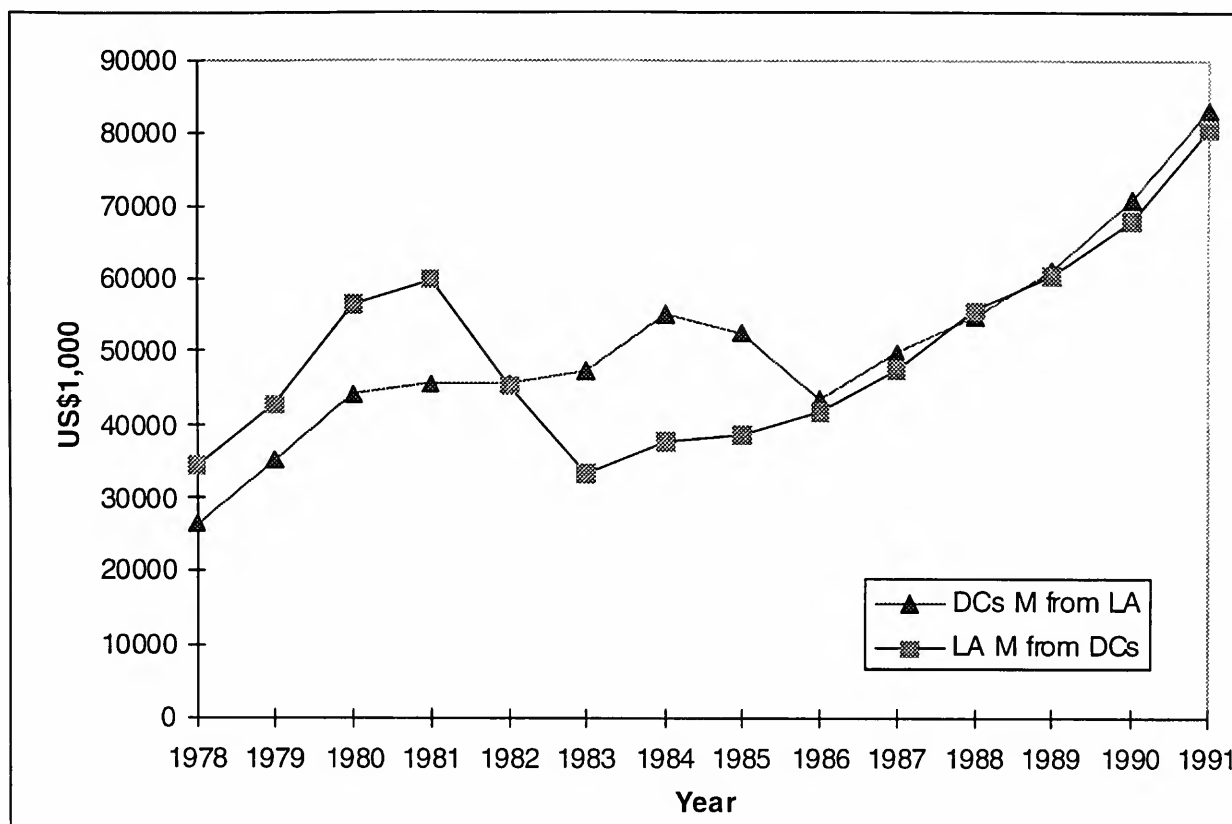
The data base

We used the MITs compiled especially for this study, using data for trade flows from the International Monetary Fund *Direction of Trade Statistics Yearbook* (several years), and for GDP from the World Bank *World Tables*. The data are presented in current US dollars for the basic tables. The definition of the level of country aggregation to be used was determined by data availability and countries' characteristics. We ended up with 23 individual countries and one closing group denoted Rest of the World. The countries are: USA, Canada, Japan, France, Germany, Italy, and UK (developed countries – DCs); Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Paraguay, Peru, Uruguay, and Venezuela (Latin America - LA); and Rest of the World (ROW).

The tables are presented on an annual basis, for the period 1978-1991. This follows the presentation of the basic IMF statistics and, although there are monthly data available, the use of yearly tables reduces the incidence of seasonal effects of trade, especially those related to primary production. Some limitations apply to these tables as they do to input-output systems; the estimates change very quickly in response to trade opportunities. However, the availability of easily compilable data allows for a constant updating of the tables, which is often not possible for input-output tables.

The data on the level of direct trade flows between LA countries and DCs reveal an equilibrium in the balance of trade between the two regions for the period. More precisely, the LA region presented a trade deficit in 1978-1981, and a trade surplus in 1983-1985. In the other years, the flows of goods and services from one region to the other, and vice-versa, nearly balanced (Figure 2).

Figure 2
Balance of Trade Between DCs and LA: 1978-1991



The extended import coefficients reveal some of the important differences between the two groups of countries. The coefficients for LA imports from DCs (6.54% of total expenditures, in the period average) are much higher than those for DCs' imports from LA (0.61% of total expenditures). However, the average total extended import coefficient for LA is smaller (11.66% against 13.20%), which reflects the importance of DCs' goods and services in the LA expenditures structure.

Import and export linkages

An attempt was made to compare the trade structure among LA and developed countries over time. We consider the hierarchy of import and export linkages - related to the column and row multipliers, respectively - and their economic landscape associated with the cross structure of the MPM. A sample of the results, depicted in Figures 3, 4, 5, reveal the cross structure for the years 1978, 1985, and 1991, the rows represent the hierarchy of export linkages while the columns provide similar detail for the import linkages. For the purpose of comparison, we decided to choose the year of 1978 as the *numéraire*. Hence, we can associate structural changes with deviations from the 1978 hierarchy. Figure 3 shows that the United States, followed by Germany, Japan and Brazil, presents far the highest export linkages. The trade connections for the other countries are more smooth or balanced.

Figure 3
1978: "Landscape" for First Order MPM

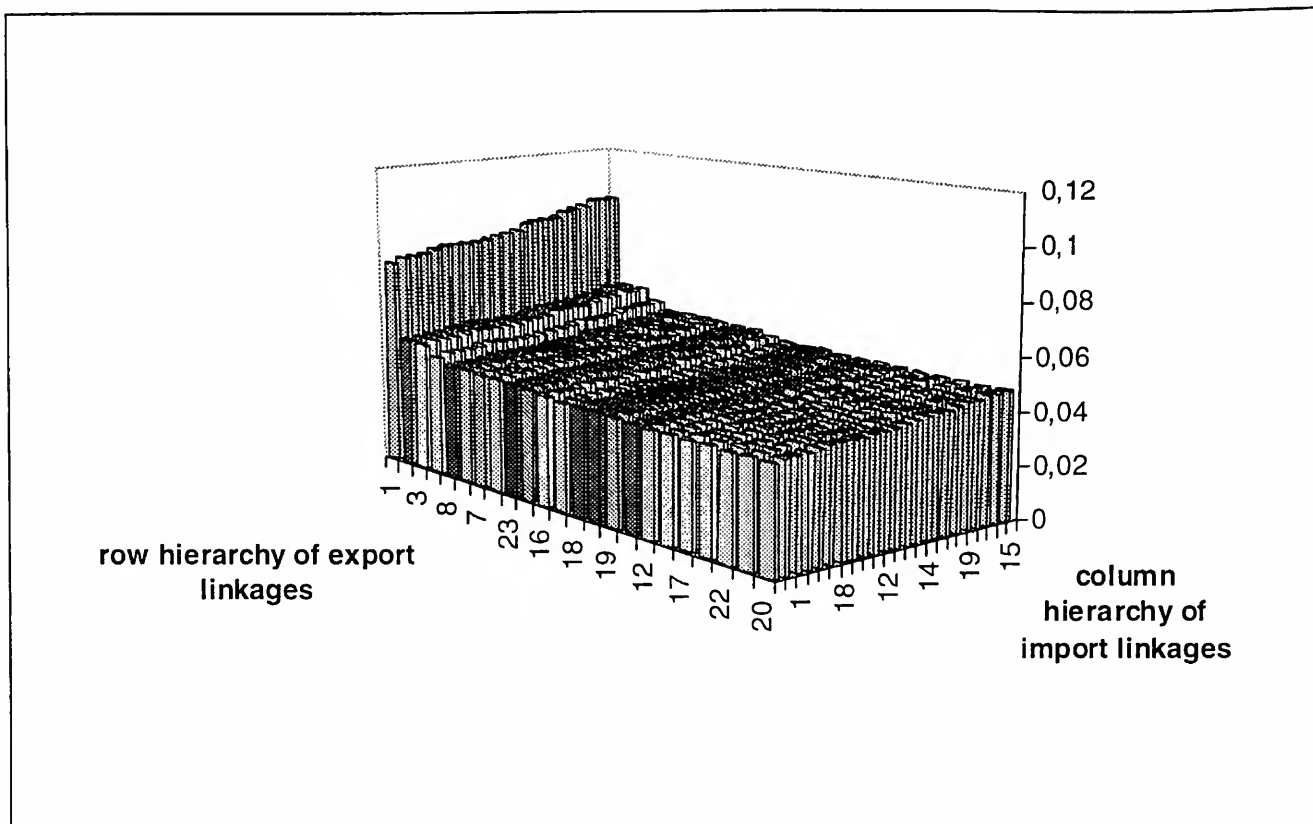


Figure 4
1985: "Landscape" Using 1978 Imposed Hierarchy

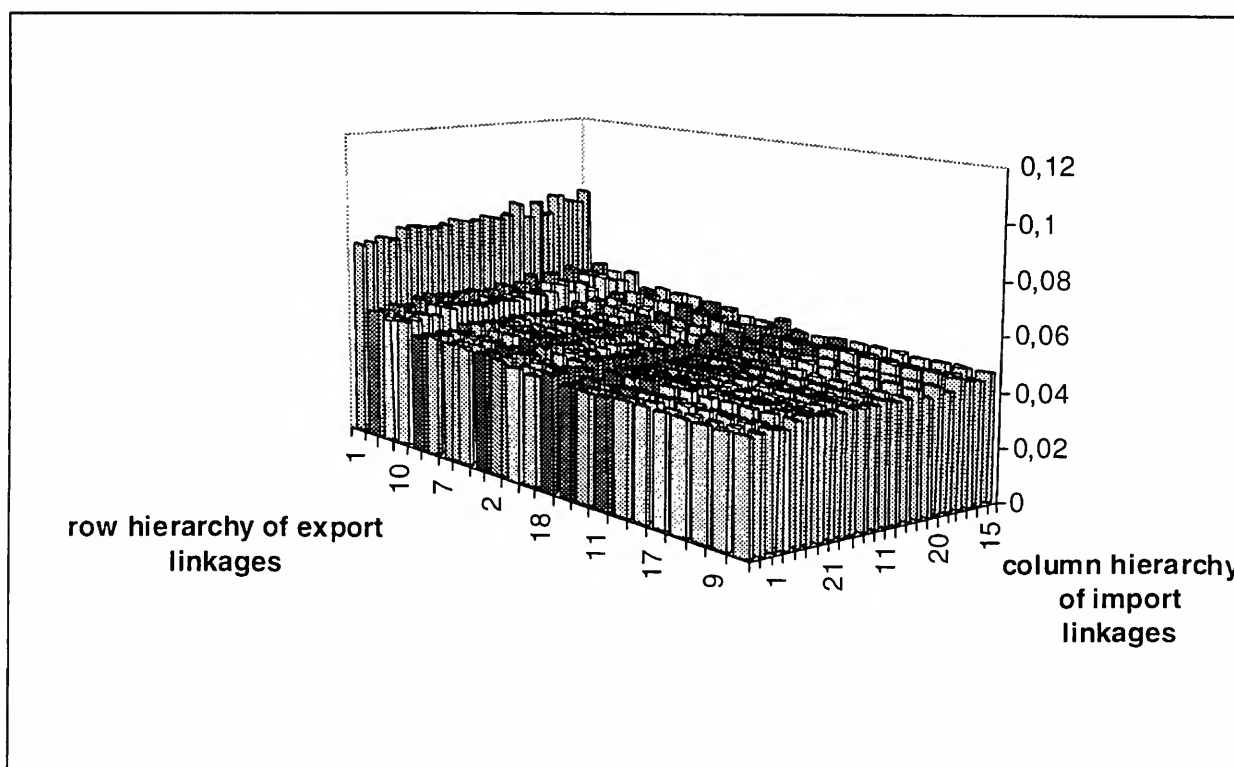
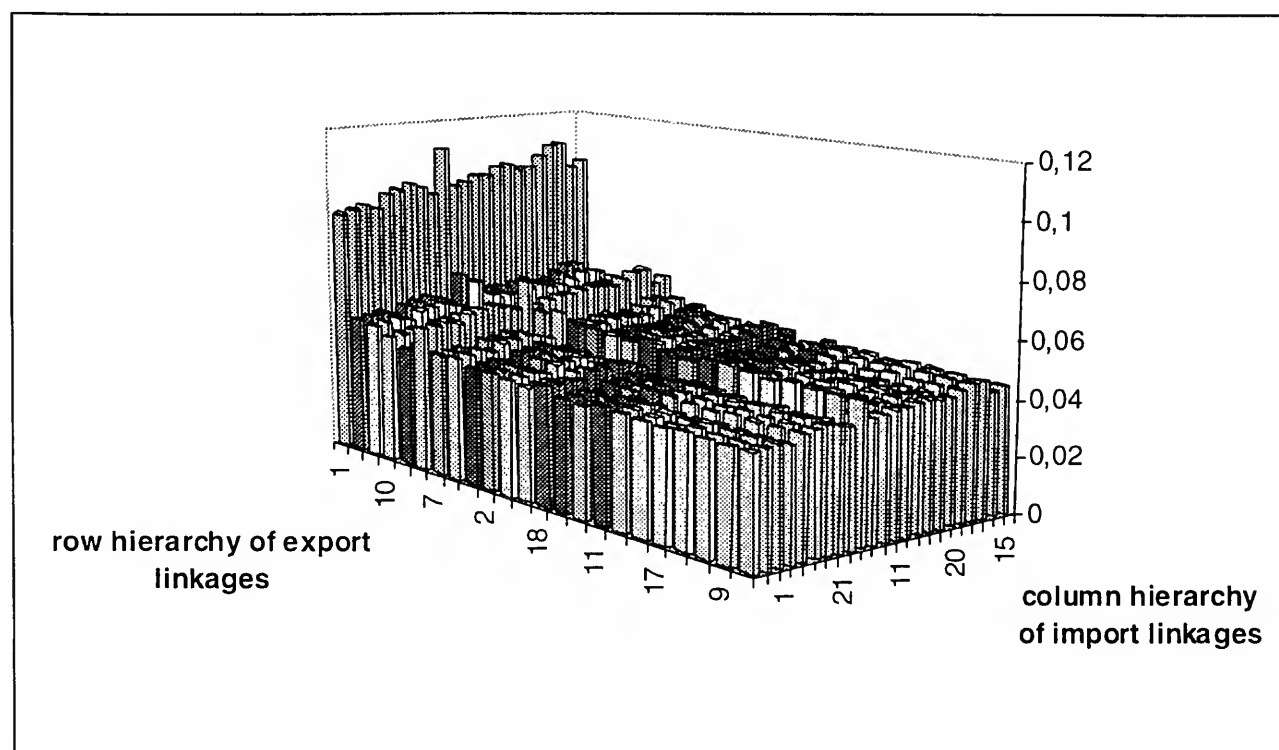


Figure 5
1991: "Landscape" Using 1978 Imposed Hierarchy



Inspection of the figures reveal some important differences over time. There is an increase in the relative linkage intensity of high-export-linkage countries, showing that these countries maintained their trade advantage over the period. Obviously, extending the structure into the 1990s, a period of greater trade liberalization, it would be likely that the changes would be more readily apparent.

Balance of payments consequences

For the analysis of the impact of growth on the balance of trade, we carried out a simulation, in which, for every year, there was an expansion of exogenous outlays of \$1000 by each country, individually. Table 1 shows the results for the year of 1985. Each column shows the net effects of an increase of \$1000 in the exogenous outlays of one country on the external payments of all the countries, including that country itself. It is clear that, for this particular year, benefits from individual countries growth, **in the form of positive net foreign balance**, were concentrated in some of the other countries. From injections in LA countries, it can be seen that DCs, especially the US, first, and Japan and Germany, were the countries that benefited the most; in general, benefits in LA were concentrated in three countries, namely, Brazil, Venezuela and Mexico.

Table 1 - Net Foreign Balance (NFB): Injections Country by Country of \$1000 (1985)

	1	2	3	4	5	6	7	8	9	10	11	12
1 USA	-67,67	98,91	14,64	10,83	12,58	9,82	20,06	7,54	29,03	12,56	31,16	33,41
2 Canada	12,55	-150,06	2,60	1,11	1,38	0,97	2,99	0,55	0,97	1,90	2,77	2,86
3 Japan	13,91	11,31	-69,47	4,78	9,71	3,55	9,40	2,56	11,71	2,97	8,46	10,23
4 France	1,93	2,57	1,15	-157,40	16,08	17,39	12,76	2,32	2,97	1,71	3,63	3,07
5 Germany	4,08	4,50	2,34	29,29	-166,63	23,40	24,15	4,80	6,75	3,36	9,05	5,44
6 Italy	2,11	2,39	0,88	15,35	14,13	-140,55	9,11	2,21	1,27	1,03	2,50	2,26
7 UK	3,09	4,75	1,20	13,69	12,74	7,96	-166,98	0,28	3,10	1,27	4,27	2,52
8 Argentina	0,25	0,18	0,27	0,28	0,45	0,61	0,11	-39,46	16,57	1,95	5,02	3,05
9 Bolivia	0,02	0,00	0,00	0,01	0,04	0,01	0,10	3,38	-133,31	0,03	0,25	0,10
10 Brazil	1,52	1,13	0,97	1,43	1,78	2,21	1,31	5,61	40,08	-55,73	10,62	2,47
11 Chile	0,19	0,18	0,23	0,23	0,43	0,35	0,42	0,80	3,17	0,75	-121,36	1,16
12 Colombia	0,24	0,12	0,11	0,15	0,66	0,20	0,21	0,36	0,19	0,03	0,92	-91,19
13 Costa Rica	0,07	0,03	0,00	0,03	0,12	0,04	0,05	0,02	0,00	0,00	0,01	0,04
14 Ecuador	0,33	0,05	0,05	0,03	0,08	0,04	0,03	0,09	0,04	0,02	1,88	1,38
15 El Salvador	0,06	0,03	0,02	0,01	0,15	0,01	0,01	0,01	0,00	0,00	0,00	0,07
16 Guatemala	0,08	0,03	0,02	0,02	0,10	0,08	0,02	0,00	0,00	0,00	0,03	0,15
17 Honduras	0,08	0,02	0,03	0,01	0,05	0,03	0,02	0,00	0,00	0,00	0,00	0,03
18 Mexico	2,69	1,13	1,08	1,25	0,49	0,62	1,21	0,40	0,17	1,16	0,84	2,74
19 Nicaragua	0,01	0,02	0,02	0,02	0,05	0,01	0,00	0,00	0,00	0,00	0,00	0,00
20 Paraguay	0,00	0,00	0,00	0,04	0,05	0,01	0,00	0,14	0,07	0,21	0,52	0,02
21 Peru	0,21	0,07	0,19	0,08	0,18	0,15	0,23	0,36	2,99	0,20	2,14	1,65
22 Uruguay	0,03	0,02	0,01	0,03	0,08	0,04	0,06	0,58	0,25	0,50	0,17	0,15
23 Venezuela	1,33	1,61	0,31	0,35	0,79	1,31	0,51	0,05	0,14	0,96	10,13	4,88
24 ROW	22,89	21,01	43,35	78,38	94,51	71,74	84,22	7,40	13,84	25,12	26,99	13,51
SUM	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
DC's share	55,7%	82,9%	32,8%	47,7%	40,0%	44,9%	47,0%	51,3%	41,9%	44,5%	51,0%	65,6%
LA's share	10,5%	3,1%	4,8%	2,5%	3,3%	4,1%	2,6%	29,9%	47,8%	10,4%	26,8%	19,6%
ROW's share	33,8%	14,0%	62,4%	49,8%	56,7%	51,0%	50,4%	18,8%	10,4%	45,1%	22,2%	14,8%
Brazil-Mexico- Venezuela's share in LA	77,9%	83,8%	71,3%	76,3%	55,6%	72,4%	70,6%	51,4%	63,4%	36,5%	66,4%	56,4%

continue

Table 1 - Net Foreign Balance (NFB): Injections Country by Country of \$1000 (1985) - continuation

	13	14	15	16	17	18	19	20	21	22	23	24
1 USA	68,10	36,76	71,54	33,80	58,55	57,85	13,58	22,22	24,99	11,32	42,80	15,54
2 Canada	3,37	2,64	2,65	1,42	2,78	1,92	3,54	0,88	1,97	0,88	3,32	2,34
3 Japan	15,02	11,61	8,43	4,98	10,03	5,25	7,52	7,14	7,63	3,65	5,38	15,14
4 France	2,73	2,03	1,54	1,53	3,37	1,41	8,89	4,85	3,21	3,59	3,96	9,02
5 Germany	8,09	8,01	5,61	4,78	4,91	3,53	5,37	5,81	5,94	8,56	4,71	16,42
6 Italy	4,77	3,08	2,49	1,08	2,73	1,20	4,40	1,99	2,36	3,06	5,29	6,18
7 UK	3,35	4,56	2,09	1,66	2,54	1,35	2,73	4,36	2,64	3,33	2,83	8,75
8 Argentina	0,69	1,01	0,66	0,27	1,37	1,13	6,75	15,66	8,11	15,37	0,97	0,84
9 Bolivia	0,01	0,02	0,00	0,00	0,01	0,01	0,03	0,06	0,59	0,06	0,01	0,01
10 Brazil	3,14	7,38	1,07	1,11	2,48	1,08	2,02	63,84	4,71	21,51	3,82	1,93
11 Chile	0,05	1,93	0,04	0,02	0,04	0,22	0,04	1,21	2,10	1,78	0,41	0,19
12 Colombia	1,32	3,30	1,62	0,44	1,00	0,05	0,90	0,04	1,52	0,07	1,57	0,19
13 Costa Rica	-175,26	0,15	6,33	2,77	4,98	0,03	5,92	0,00	0,01	0,00	0,07	0,03
14 Ecuador	0,12	-112,67	0,12	0,03	0,13	0,03	0,11	0,03	0,29	0,04	0,05	0,16
15 El Salvador	3,45	0,01	-167,52	4,33	1,08	0,00	0,75	0,00	0,00	0,00	0,00	0,01
16 Guatemala	6,91	0,17	17,99	-89,57	4,76	0,05	3,86	0,00	0,01	0,01	0,01	0,06
17 Honduras	1,63	0,01	2,31	0,23	-146,02	0,01	2,57	0,00	0,00	0,00	0,00	0,03
18 Mexico	3,59	2,98	13,80	8,02	5,30	-83,08	7,41	0,18	0,73	3,64	0,63	0,71
19 Nicaragua	2,19	0,00	0,57	0,76	0,94	0,03	-148,29	0,00	0,00	0,00	0,00	0,02
20 Paraguay	0,00	0,00	0,00	0,00	0,00	0,00	0,00	-147,15	0,03	0,89	0,00	0,02
21 Peru	2,40	4,41	0,52	0,03	0,18	0,06	0,43	0,03	-83,97	0,38	0,52	0,16
22 Uruguay	0,01	0,09	0,01	0,00	0,01	0,03	0,01	1,24	0,09	-121,80	0,05	0,05
23 Venezuela	18,64	0,26	12,19	5,88	17,42	0,15	1,50	0,13	1,37	0,25	-90,95	0,66
24 ROW	25,68	22,26	15,94	16,43	21,41	7,69	69,96	17,48	15,67	43,41	14,55	-78,46
SUM	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
DC's share	60,2%	61,0%	56,3%	55,0%	58,1%	87,3%	31,0%	32,1%	58,0%	28,2%	75,1%	93,5%
LA's share	25,2%	19,3%	34,2%	26,7%	27,2%	3,5%	21,8%	56,0%	23,3%	36,1%	8,9%	6,5%
ROW's share	14,7%	19,8%	9,5%	18,3%	14,7%	9,3%	47,2%	11,9%	18,7%	35,6%	16,0%	0,0%
Brazil-Mexico- Venezuela's share in LA	57,5%	48,9%	47,3%	62,8%	63,5%	42,7%	33,8%	77,8%	34,8%	57,7%	54,9%	65,1%

Figure 6
Average NFB Given Injections Country by Country of 1000
in Developed Countries: 1978-1991

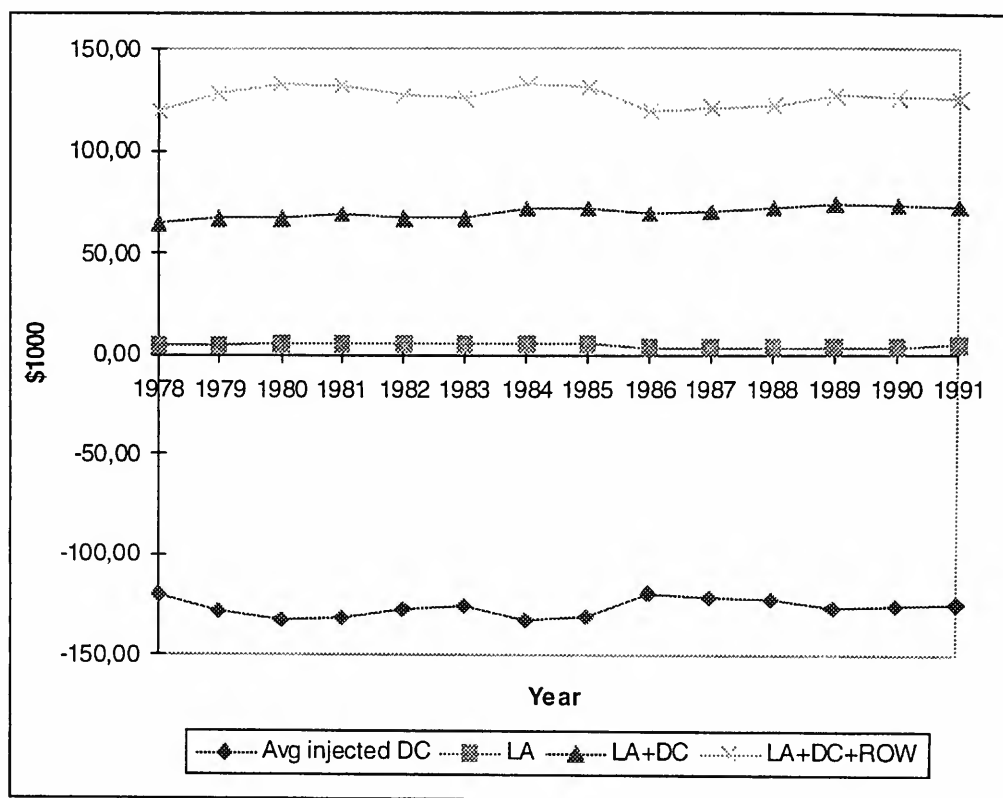
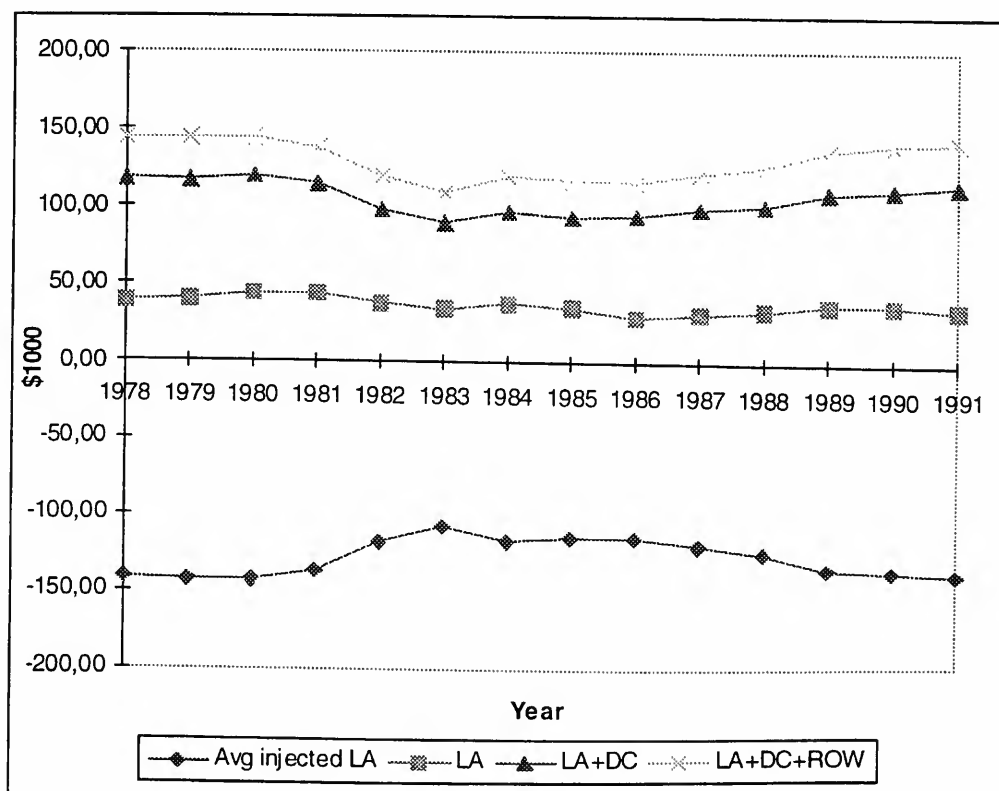
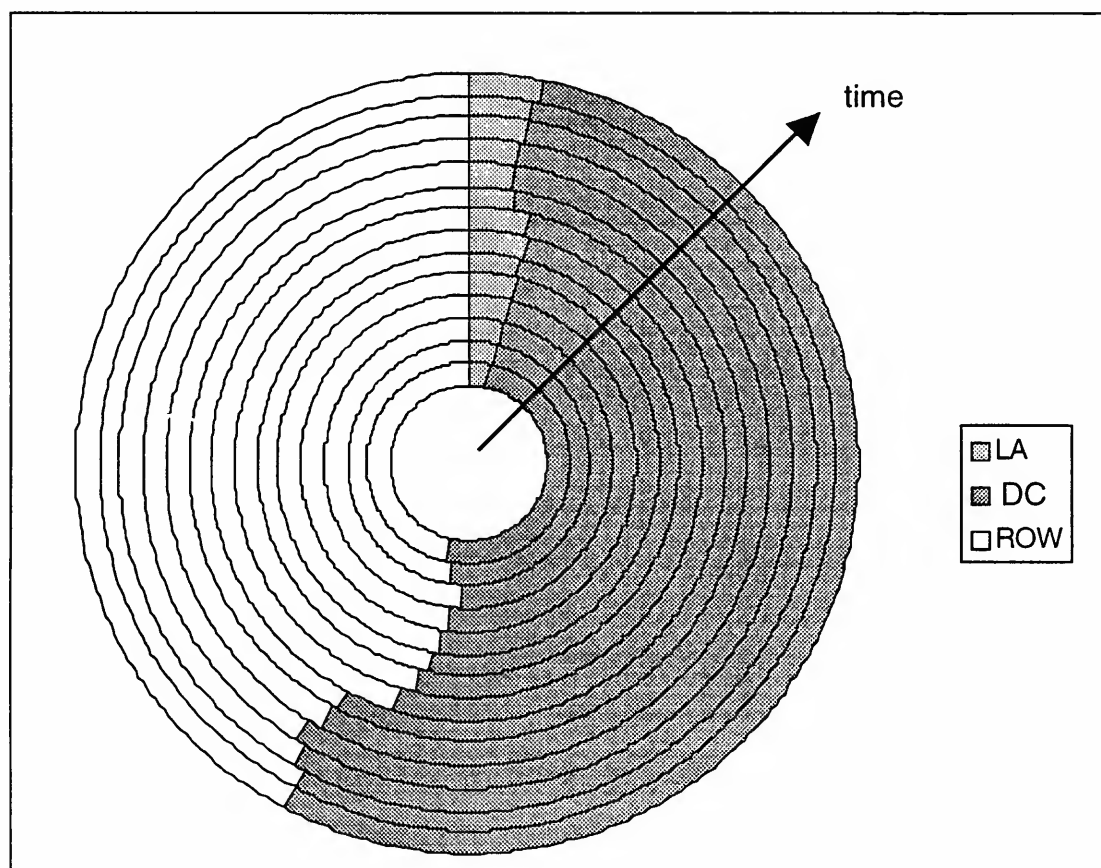


Figure 7
Average NFB Given Injections Country by Country of 1000
in Latin American Countries: 1978-1991



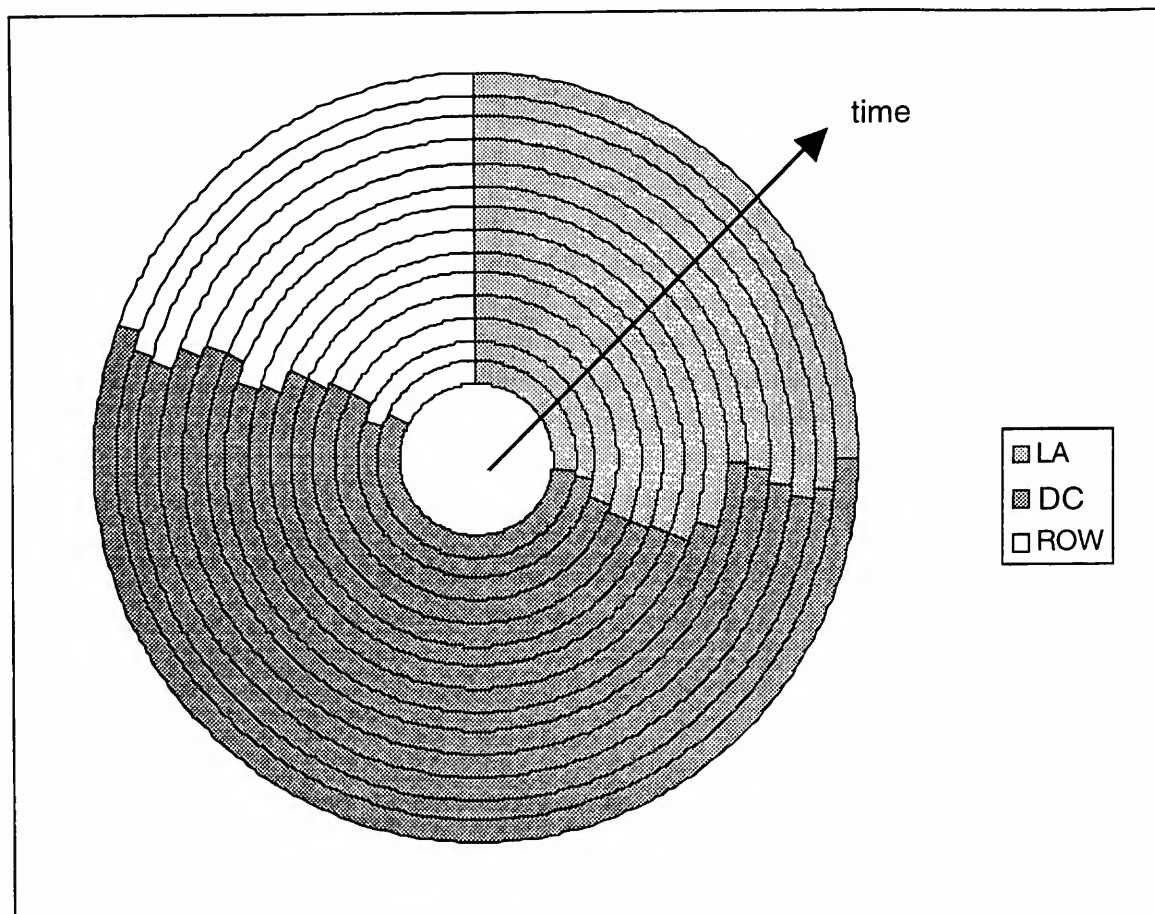
The trends established in the period 1978-1991 reveal the following highlights. First, the average benefit for LA countries from the growth in DCs is very small when compared to the average benefits that DCs achieve from the growth in LA (see Figures 6 and 7). This is the counterpart of the high dependency of LA on DCs imports and exports. Secondly, we should point out the trends in the distribution of benefits from growth in both DCs and LA countries (Figures 8 and 9). In both cases, the share of benefits to DCs increased over time, while the share of benefits to LA countries decreased. In **relative terms**, therefore, given the structure of international trade in the period 1978-1991, LA countries became potentially worse off than developed countries⁹ However, since trade flows are expressed in current US dollars, this might also reflect changes that occur as a result of movements in exchange rates, rather than true changes in trade relationships.

Figure 8
Average Percentual Distribution of Benefits from Growth
in Developed Countries: 1978-1991



9 *Relativity* implies that we are not considering the level of output in each country, but only the trade linkage structure. In a sense, we are dealing with marginal changes in output. Economy-size effects are not considered here.

Figure 9
Average Percentual Distribution of Benefits from Growth
in Latin American Countries: 1978-1991



Growth consequences

We estimated, for each year, both the effects of DCs growth on LA countries, and the effects of LA growth on DCs, with international trade as the mechanism generating such spillover effects. Table 2 shows the estimated ROW and LA foreign-induced trade multipliers for DCs, for the years 1978, 1985 and 1991, which are computed as the column sums of DCs across all the rows, in the ROW case, and excluding the ROW row in the LA case. The table also shows the ROW and DCs foreign-induced trade multipliers for LA countries, computed in a similar way. In our notation, $M1$, $M1(LA)$, $M2$, and $M2(LA)$ stand for the four multipliers. $M1$ indicates the effect of a \$1 increase in the exogenous outlays of a DC on the rest of the world ($M1(LA)$, the effects on LA only). $M2$ indicates the effect of a \$1 increase in the exogenous outlays of a LA country on the rest of the world ($M2(DC)$, the effects on DCs only).

Table 2 - Estimated M1, M1 (LA), M2, and M2 (DC) for Selected Years: 1978, 1985, 1991

	M1						M1(LA)					
	1978		1985		1991		1978		1985		1991	
	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value
USA	6	0,0362	6	0,0343	6	0,0397	6	USA	1	0,0078	1	0,0085
Canada	7	0,0307	7	0,0294	7	0,0327	7	Canada	2	0,0062	4	0,0061
Japan	5	0,0491	5	0,0539	5	0,0402	5	Japan	7	0,0023	7	0,0024
France	3	0,0832	3	0,0953	3	0,0843	3	France	6	0,0043	6	0,0031
Germany	2	0,0951	2	0,1158	1	0,1127	1	Germany	3	0,006	3	0,0044
Italy	4	0,0827	4	0,0895	4	0,0684	4	Italy	4	0,0062	2	0,0032
UK	1	0,1054	1	0,1025	2	0,0872	2	UK	5	0,0037	5	0,0028

	M2						M2(DC)					
	1978		1985		1991		1978		1985		1991	
	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value
Argentina	16	0,061	16	0,0316	16	0,0323	15	Argentina	15	0,023	16	0,0228
Bolivia	2	0,1591	11	0,0779	14	0,0754	6	Bolivia	6	0,0618	9	0,0622
Brazil	15	0,0617	15	0,0569	15	0,048	16	Brazil	16	0,0278	15	0,0301
Chile	9	0,1155	6	0,1005	6	0,1334	10	Chile	10	0,0691	7	0,0851
Colombia	10	0,0919	10	0,0819	11	0,0924	12	Colombia	12	0,0662	8	0,0733
Costa Rica	5	0,1546	1	0,1458	2	0,1869	3	Costa Rica	3	0,1159	1	0,1468
Ecuador	4	0,1549	4	0,1021	4	0,1362	4	Ecuador	4	0,0763	5	0,1104
El Salvador	3	0,1568	3	0,1216	3	0,1176	2	El Salvador	2	0,1031	2	0,0979
Guatemala	8	0,1183	8	0,0733	9	0,1301	9	Guatemala	9	0,0542	11	0,1022
Honduras	1	0,1716	4	0,1181	1	0,1926	1	Honduras	1	0,0932	3	0,157
Mexico	13	0,0827	9	0,088	8	0,132	11	Mexico	11	0,0791	4	0,1149
Nicaragua	7	0,1336	7	0,1336	3	0,182	7	Nicaragua	7	0,0523	13	0,0974
Paraguay	14	0,0812	13	0,0732	5	0,1339	14	Paraguay	14	0,0529	12	0,0837
Peru	11	0,0885	14	0,0725	12	0,0822	13	Peru	13	0,0543	10	0,0607
Uruguay	12	0,0879	8	0,0897	13	0,0803	8	Uruguay	8	0,0393	14	0,0514
Venezuela	6	0,144	7	0,0926	7	0,1333	5	Venezuela	5	0,0757	6	0,1081

The M1 multiplier maintains roughly the same rank for the three years, with UK and Germany presenting the highest multipliers, followed by France, Italy, Japan, USA and Canada, in this order. However, the rank changes for M1(LA). USA has the highest multiplier in the three years considered. We should also notice the position of Canada, one of the countries whose growth generated a larger impact on LA. By examining the MIFs more closely, we see that Canada's trade linkages with the USA are important in explaining this relatively high impact, since a considerable part of it is achieved through the linkages Canada→USA→LA. The consistent behavior of Germany, occupying the third place, should also be pointed out.

In the M2 case, the highest multipliers were found in small countries in the region, namely, Honduras, Costa Rica, Ecuador, El Salvador, and Nicaragua. Larger, more developed countries, such as Argentina and Brazil presented smaller multipliers. For the M2(DC), it is remarkable the positions gained in the rank by Mexico, whose growth impact on DCs increased over time, especially by the trade linkages with USA and Canada (an effect that pre-dated NAFTA). Also, the high multipliers shown by Honduras, El Salvador, Costa Rica, Ecuador, and Venezuela should be noted. Finally, the MERCOSUL countries (Argentina, Brazil, Paraguay, and Uruguay) contained the lowest multipliers.

Figure 10
Average ROW and LA Foreign-Induced Multiplier
of DCs, M1 and M1(LA): 1978-1991

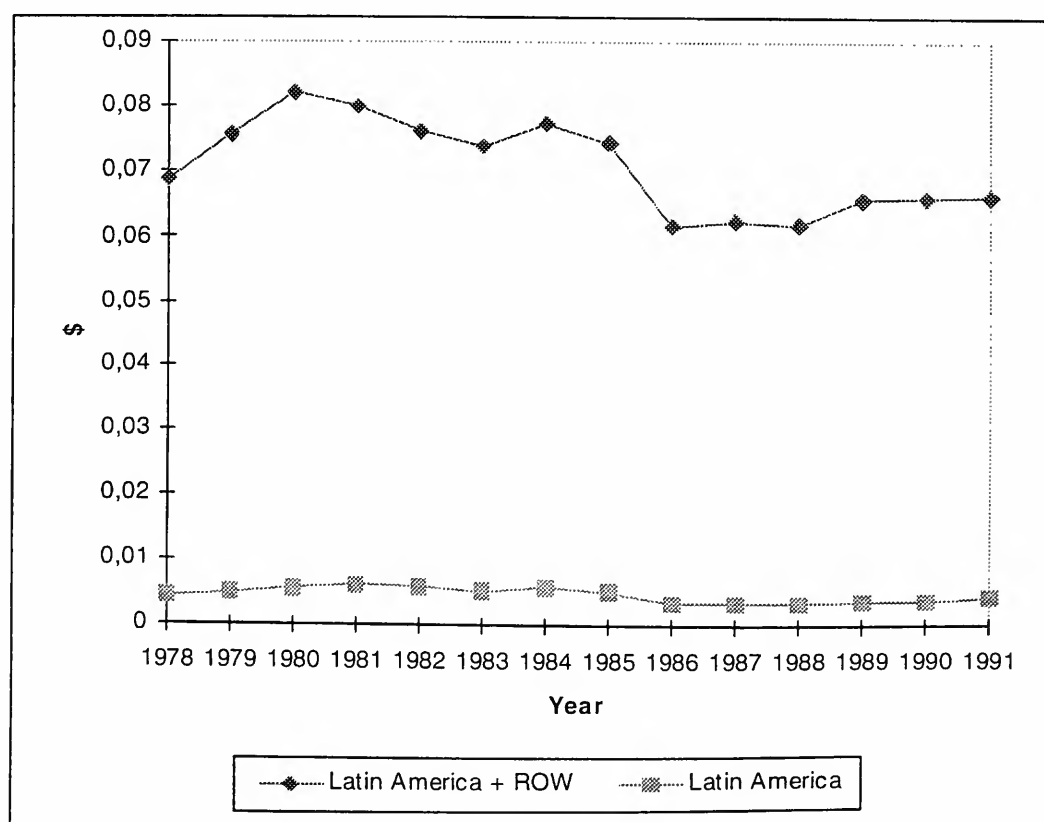
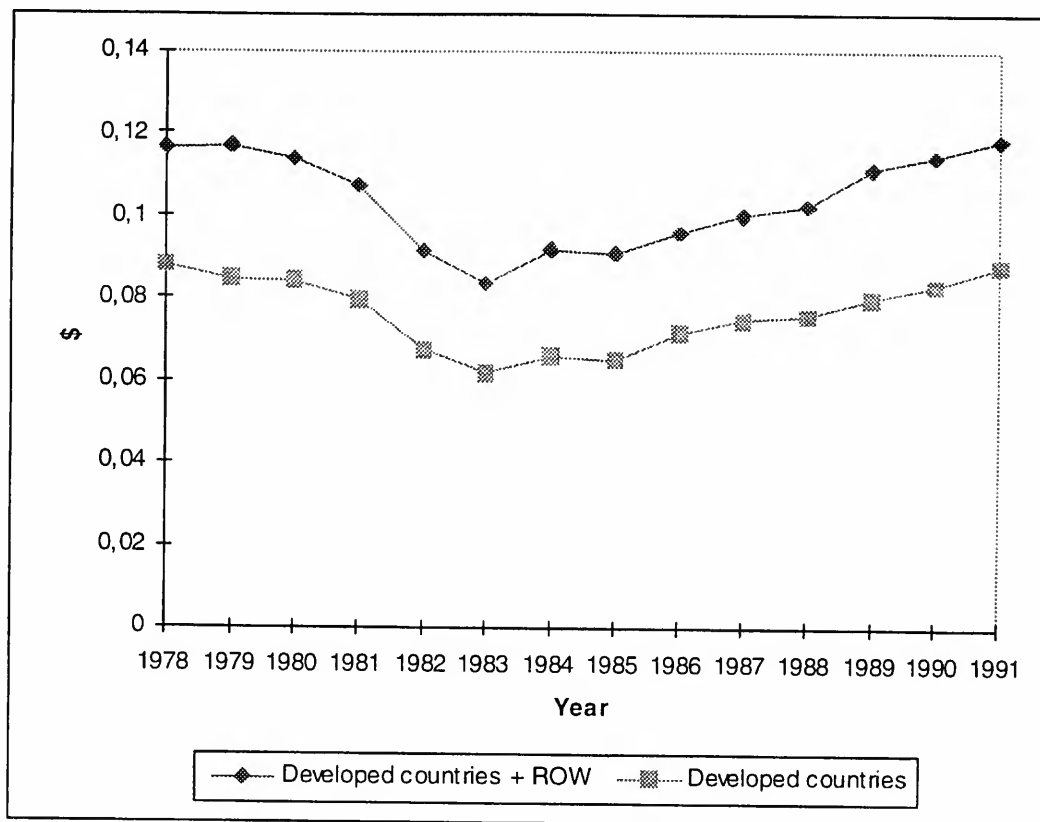


Figure 11
Average ROW and DCs Foreign-Induced Multiplier
of LA Countries, M2 and M2(DC): 1978-1991



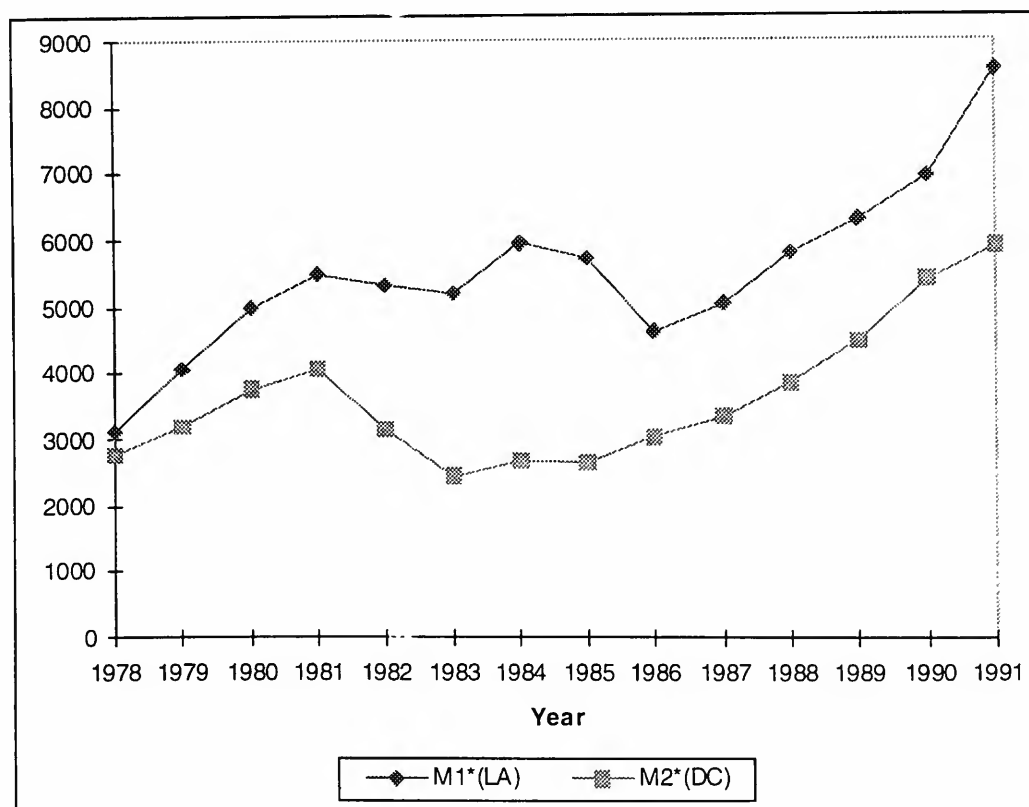
To compare the multipliers $M1(LA)$ and $M2(DC)$, we took the country averages for each year. Figures 10 and 11 show the behavior of $M1$ vs. $M1(LA)$, and $M2$ vs. $M2(DC)$, respectively. In the first graph, we see that the impact of DCs growth on LA is very small. On average, during 1978-1991, for each \$1 increase in the exogenous outlays in a DC, there is a \$0.0047 potential increase in income in LA. This indicator provides a useful measure of the degree to which DCs growth influences LA growth through trade linkages. By comparing $M1$ and $M1(LA)$, we can verify that the share of the impact of DCs growth for LA is very small (only 6.60% over the period). Both multipliers are also declining over time, indicating that changes in the structure of international trade, in the period 1978-1991, implied reduced indirect effects of DCs growth on LA countries.

If attention is now directed at the impact of growth in a LA country on the DCs, \$0.0764 was generated for each \$1 of increase in the former's exogenous outlays. First, the LA growth impact on DCs is much higher than the DCs impact on LA countries (approximately 16 times, in the average, for each \$1).¹⁰ Secondly, the share of the impact of LA growth that goes to DCs is incredibly high (73.44%).

¹⁰ Even for the total multipliers, the values of $M2$ are higher (the period average was 0.0709 for $M1$, and 0.1040 for $M2$).

The above facts combined would lead us to the conclusion that an increase in the exogenous outlays, of the same *level*, of all the countries would benefit disproportionately the DCs, increasing the income gap between LA countries and DCs in favor of the latter. However, one fact that we do not take into account when carrying out this type of analysis is that we do not consider the scarcity of money¹¹ in each country or region. To contemplate this issue, we should weigh the multipliers with the respective exogenous outlays, giving more importance to the multipliers of countries where money is less scarce (for weighted-multiplier schemes, see Hazari, 1970). Figure 12 shows the weighted partial multipliers, which refers to proportional changes in exogenous outlays (instead of absolute changes). These results reverse, in a sense, our prior conclusion, in that, now, LA countries benefit more than DCs from overall proportional changes in exogenous outlays in the world economy. The time series of the weighted multipliers reveal that the difference between them increased in the 1980's. This decade, known in Latin America as the "lost decade", was characterized by economic stagnation in LA, when its GDP grew at an annual rate of only 1.7%, therefore resulting in increases in the scarcity of money in the region, and, *ceteris paribus*, lower weighted multipliers $M1^*(LA)$.¹²

Figure 12
Weighted Multipliers: 1978-1991



11 By scarcity of money we mean the inverse of the total level of expenditures ($1/E$). Thus, in a country with low levels of expenditures, money will be relatively more scarce than in a country with higher levels of expenditures.

12 See Baer *et alii* (1991).

Figure 13
Share of Direct Import Requirements in Total Multiplier M1: 1978-1991

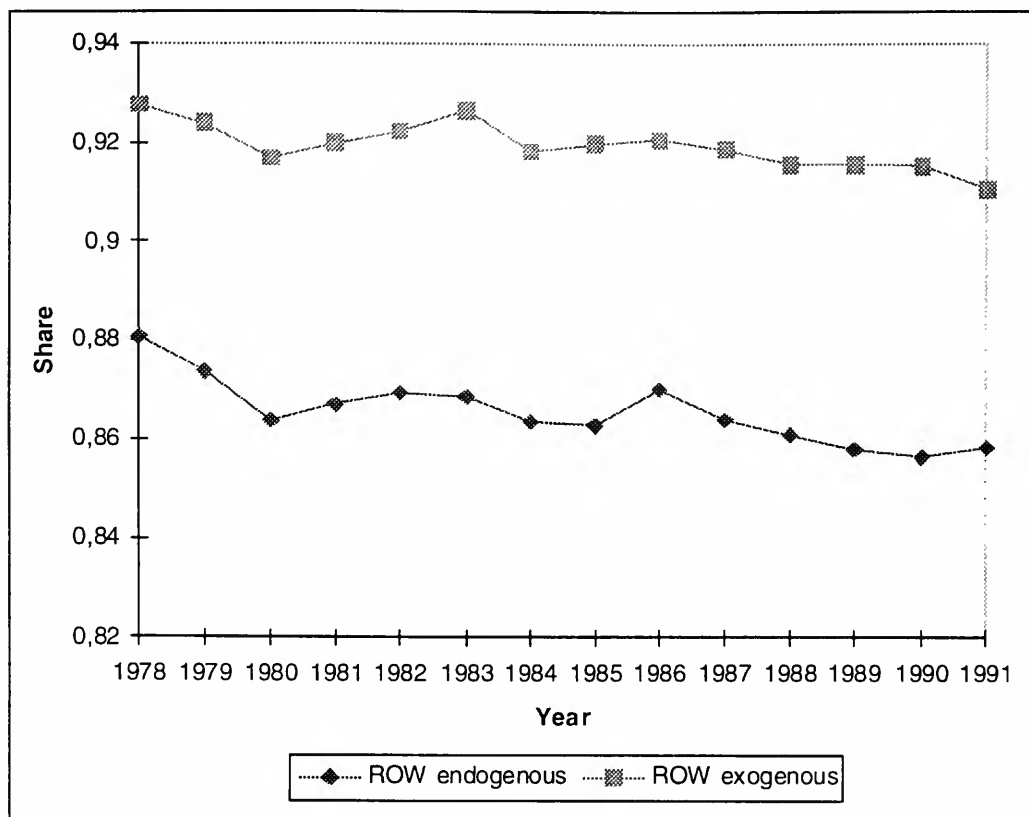


Figure 14
Share of Indirect Import Requirements in Total Multiplier M1: 1978-1991

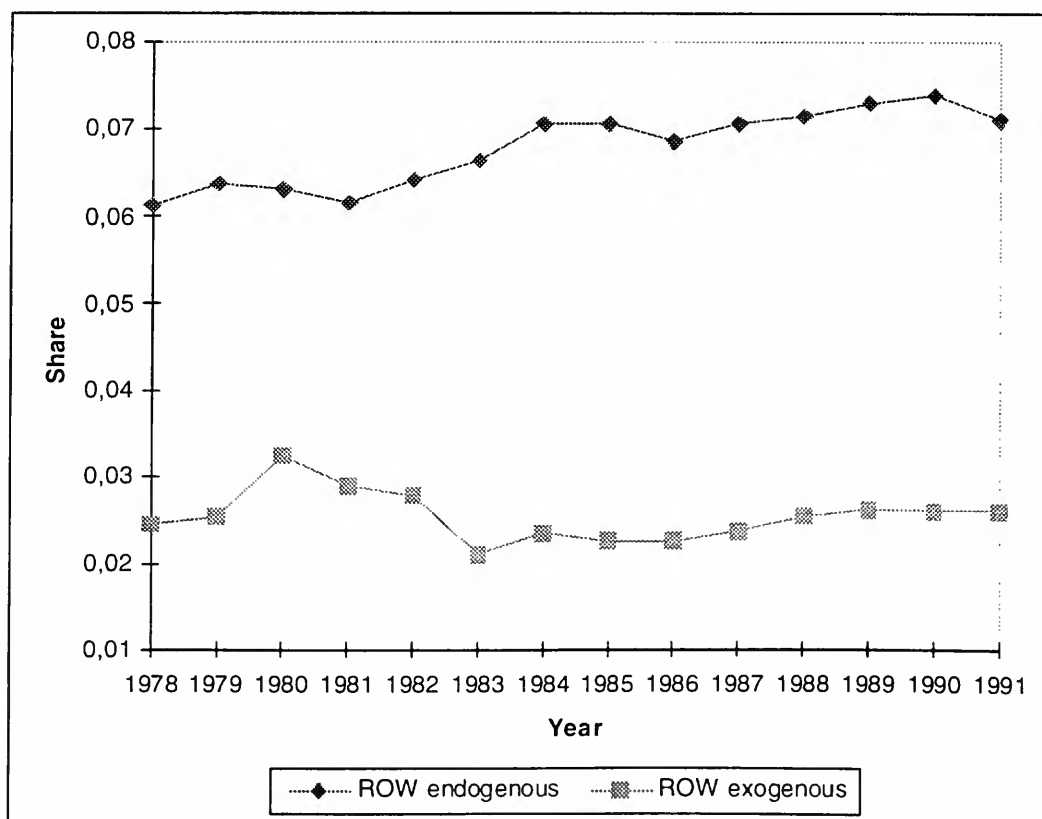


Figure 15
Share of Internal Propagation of LA Countries (and ROW)
in Total Multiplier M1: 1978-1991

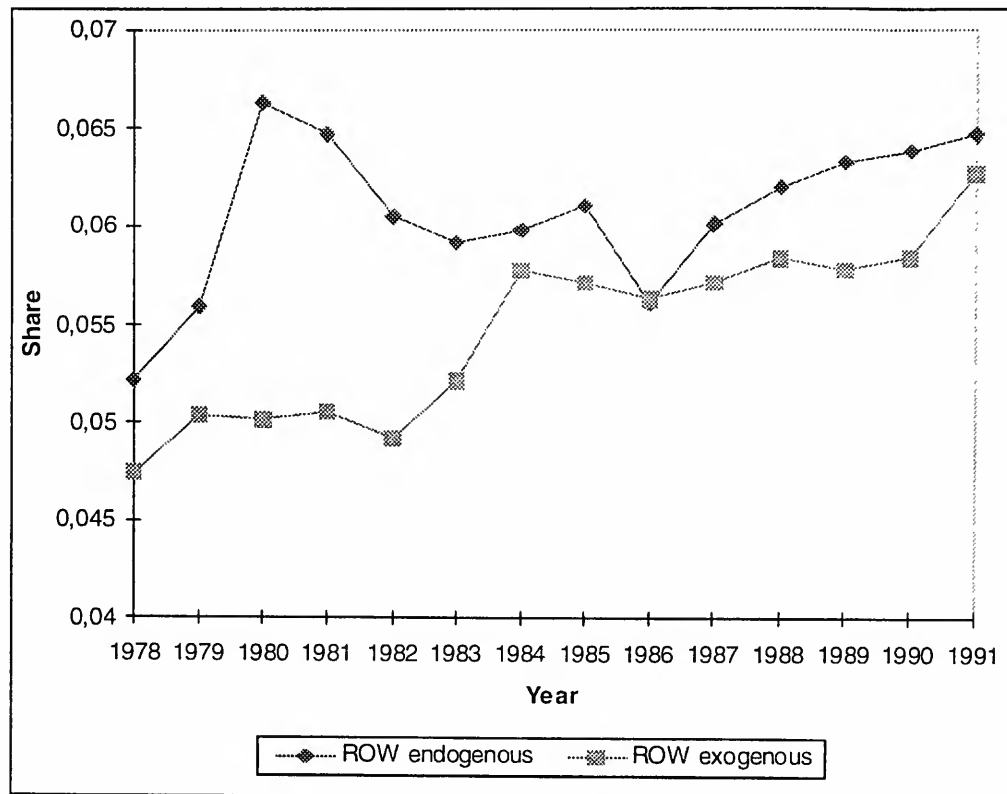


Figure 16
Share of External Propagation of LA Countries (and ROW)
in Total Multiplier M1: 1978-1991

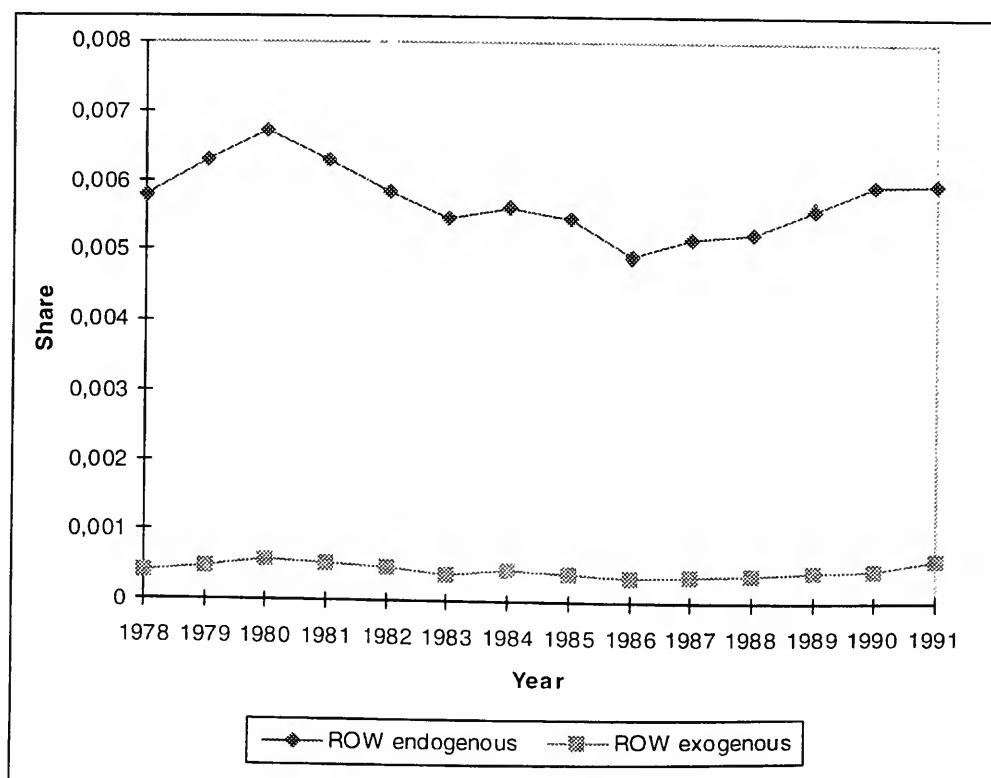


Figure 17
Share of Direct Import Requirements in Total Multiplier M2: 1978-1991

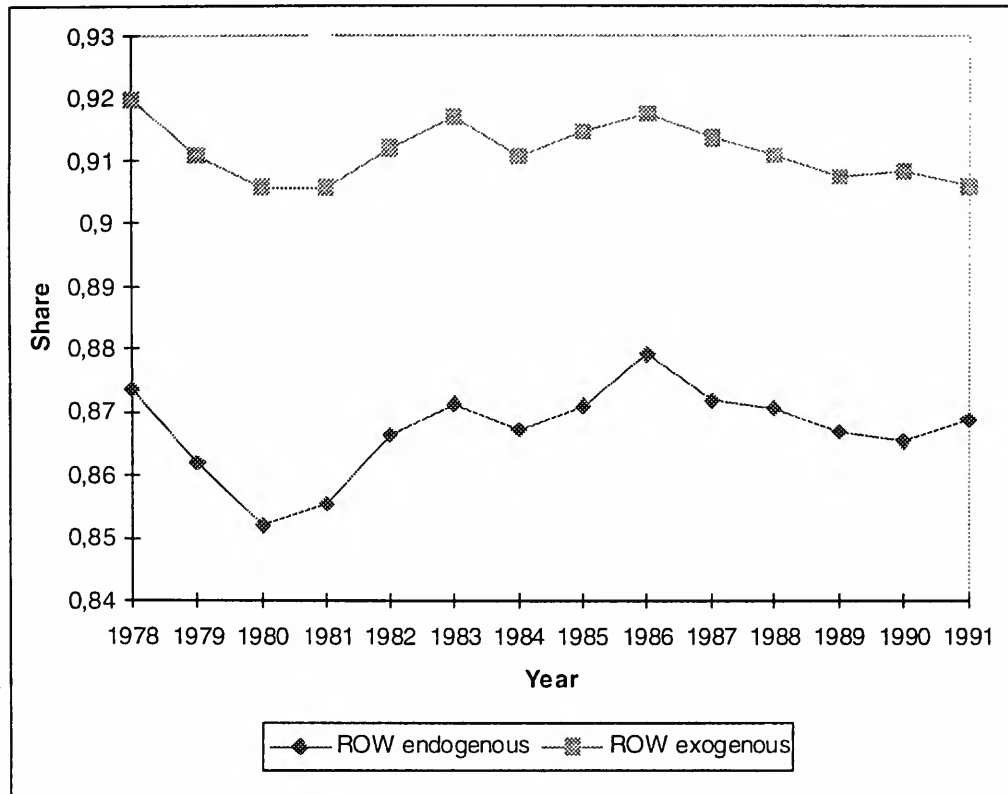


Figure 18
Share of Indirect Import Requirements in Total Multiplier M2: 1978-1991

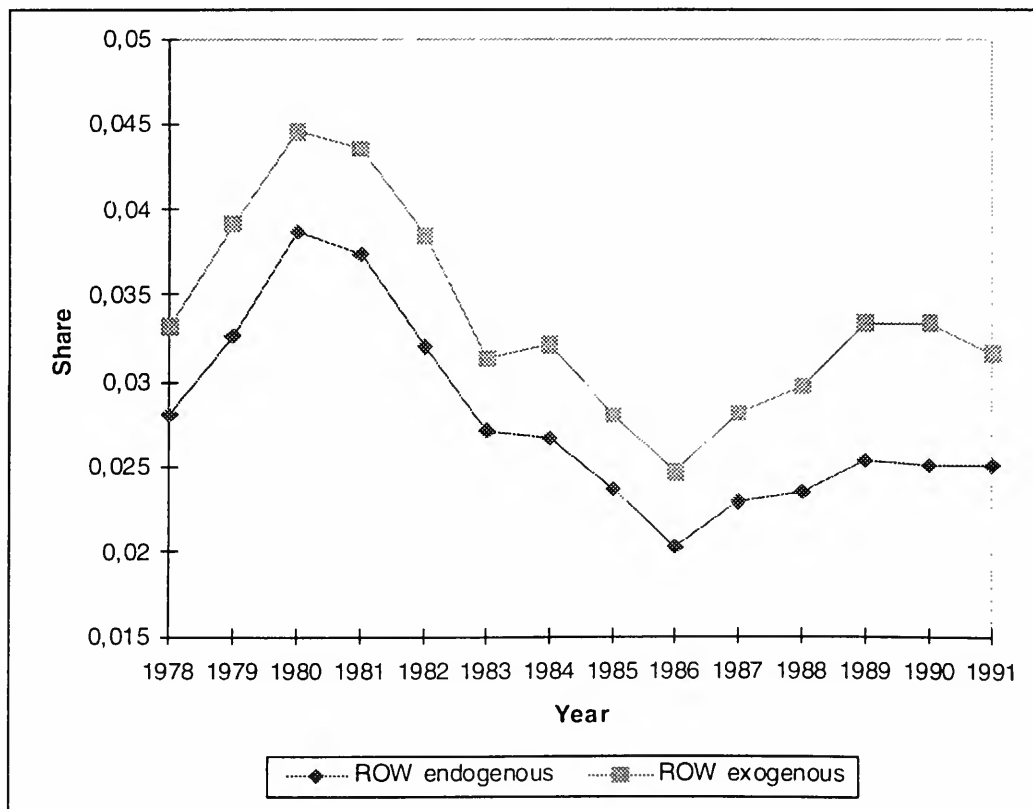


Figure 19
Share of Internal Propagation of Developed Countries (and ROW)
in Total Multiplier M2: 1978-1991

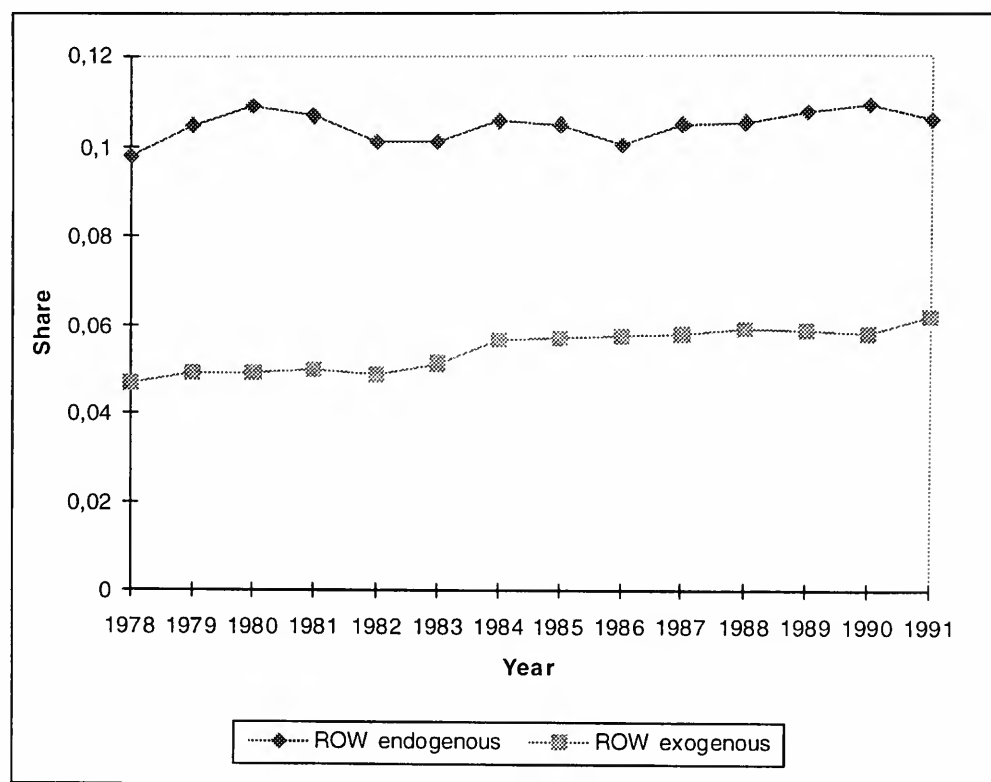
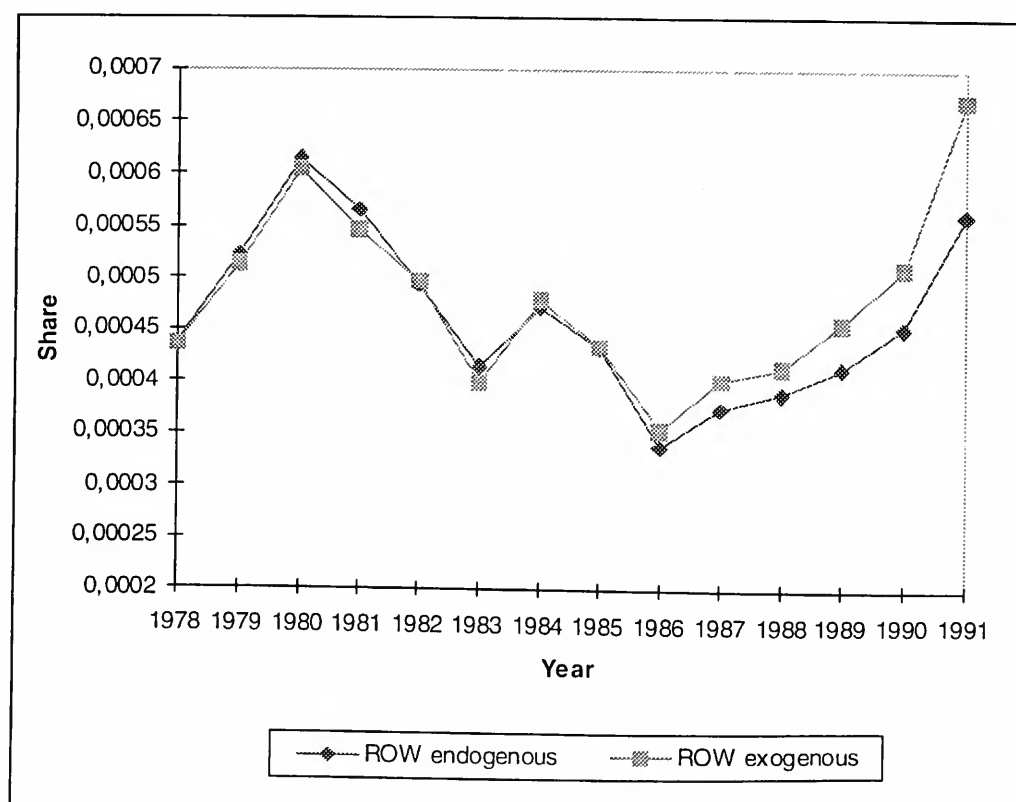


Figure 20
Share of External Propagation of Developed Countries (and ROW)
in Total Multiplier M2: 1978-1991



The decomposition (into internal and external effects) of the multipliers M1 and M2 indicates important trends in the composition of the external demands of DCs and LA countries (Figures 13-20).¹³ Regarding M1, we have: a) on average, DCs show decreasing shares of direct import requirements from LA countries (88.09% in 1978, and 85.85% in 1991); b) changes in the shares of indirect import requirements for DCs are positive in the period (6.11% in 1978, and 7.09% in 1991); c) and the shares of internal propagation of the LA countries stimulated by DCs, on average, are increasing over time (5.21% in 1978, and 6.46% in 1991). In the case of M2, a trend is more clearly perceived from 1980 on: a) LA countries show increasing shares of direct import requirements from DCs (85.20%, in 1980, and 86.91% in 1991); b) decreasing shares of indirect import requirements for LA are present (3.87% in 1980, and 2.50% in 1991); and c) increasing shares, but not monotonic, of internal propagation of DCs (9.77% in 1978, and 10.54% in 1991).

Putting together the trends of the decomposition of both multipliers, M1 and M2, there is evidence that both DCs and LA countries are reducing their external demand for LA goods and services. This reflects, to a great extent, the commodity recession of the 1980's.¹⁴ However, recent evidence suggests that the impact of the formation of MERCOSUL has generated significantly larger trade increases in percentage terms among the member LA countries than their trade with DC and ROW countries.

6 Final remarks

The analysis of the impacts of growth in DCs on LA countries, and the impacts of growth in LA countries on DCs, carried out in this paper, reveal the potential for analysis of trade flows using techniques pioneered in input-output analysis. In a broader sense, the use of the MIT framework provides an analytical tool to study different aspects of the structure of international trade. Different issues might be addressed from the use of MITs, which can be constructed for different purposes by varying, for instance, the country aggregation.

13 The shares were computed for ROW both endogenous and exogenous, in order to make the point clearer.

14 "The Latin American countries depend on a relatively small number of primary commodity exports for a large amount of their foreign exchange earnings. Ten nonfuel commodities have contributed at least 1% to total regional export earnings in the last fifteen years - namely coffee (9.2%), soybeans (4.7%), copper (3.6%), iron ore (3.6%), sugar (2.9%), beef (1.7%), cotton (1.6%), cocoa (1.3%), bananas (1.3%), and maize (1.2%)." (Lord and Boye, 1991)

In our specific study of the passive role played by trade in the growth process, involving DCs and LA countries, some features of their relations were revealed (growth impact on trade balance and indirect income/output growth), which might be used as indicators to help in the formulation and evaluation of development policies by international agencies, such as the IMF, the World Bank and the Inter-American Development Bank.

References

- Baer, W., Petry, J. and Simpson, M. *Latin America: the crisis of the eighties and the opportunities of the nineties*. Bureau of Economic and Business Research, University of Illinois, Urbana, 1991.
- Fritz, O. *An input-output analysis of indirect industrial pollution generation*. Unpublished Doctoral Dissertation, University of Illinois at Urbana-Champaign, 1995.
- International Monetary Fund. *Direction of Trade Statistics Yearbook* (1969-1975, 1981, 1987, 1992).
- Goodwin, R. M. *Essays in linear economic structures*. London: McMillan Press, 1983.
- Hazari, B. R. Empirical identification of key sectors in the Indian economy. *Review of Economics and Statistics*, v. 52, n. 3, 1970.
- Krugman, P. Is bilateralism bad? In: Helpman, E. and Razin, A. (eds.), *International trade and trade policy*. Cambridge: MIT Press, 1991, p. 38-55.
- _____. Regionalism versus multilateralism: analytical notes. In: Melo, J. de and Panagariya, A. (eds.), *New dimensions in regional integration*. Cambridge: University Press, 1993, p. 58-79.
- Lord, M. J. and Boye, G. R. The determinants of international trade in Latin American commodity exports. In: Urrutia, Miguel (ed.), *Long-term trends in Latin American economic development*. Washington D.C.. IDB, 1991.
- Machlup, F. Period analysis and multiplier theory. *The Quarterly Journal of Economics*, p. 1-27 November 1939.
- _____. *International trade and national income multiplier*. New York: AMK Reprints of Economic Classics, 1943.
- Miyazawa, K. Foreign trade multiplier, input-output analysis and consumption function. *The Quarterly Journal of Economics*, v. 74, n. 1, 1960.

- _____. Internal and external matrix multipliers in the input-output model. *Hitotsubashi Journal of Economics*, v. 7, p. 38-55, 1966.
- _____. An analysis of the interdependence between service and goods-producing sectors. *Hitotsubashi Journal of Economics*, v. 12, p. 10-21, 1971.
- Ota, H. The 1985 Japan-US-EC-Asia input-output table: its compilation and some results of analysis. *Journal of Applied Input-Output Analysis*, v. 2, n. 1, p. 54-78, 1994.
- Sonis, M. and Hewings, G. J. D. Hierarchies of regional sub-structures and their multipliers within input-output systems: Miyazawa revisited. *Hitotsubashi Journal of Economics*, v. 34, p. 33-44, 1993.
- Sonis, M. and Hewings G. J. D. Matrix sensitivity, error analysis and internal/external multiregional multipliers. *Hitotsubashi Journal of Economics*, v. 36, p. 61-70, 1995.
- Sonis, M., Oosterhaven, J. and Hewings, G. J. D. Spatial economic structure and structural changes in the EC: feedback loop input-output Analysis. *Economic Systems Research*, v. 5, n. 2, p. 173-184, 1993.
- Sonis, M., Hewings, G. J. D. and Guo, J. Input-output multiplier product matrix. *Discussion Paper 94-T-12*, Regional Economics Applications Laboratory, University of Illinois, 1994.
- Sonis, M., Hewings, G. J. D. and Gazel, R.. The structure of multi-regional trade flows: hierarchy, feedbacks and spatial linkages. *The Annals of Regional Science*, 29, p. 409-430, 1995a.
- Sonis, M., Guilhoto, J. J. M. and Hewings, G. J. D. The Asian economy: trade structure interpreted by feedback loop analysis. *Journal of Applied Input-Output Analysis*, v. 2, n. 2, p. 24-40, 1995b.
- Thorbecke E. and Field, A. J. A ten-region model of world trade. In: Sellekaerts, W. (ed.), *International trade and finance: essays in honour of Jan Tinbergen*. London: IASP, 1974.
- Todaro, M. *Economic development*. New York: Longman, 1994.
- World Bank. *The World Tables 1994*. Washington, D. C..

Rendimentos perdidos por trabalhadores em condições inadequadas de saúde

Ana Lúcia Kassouf§

RESUMO

Este estudo tem como objetivo principal estimar as perdas de rendimento dos homens adultos (18 a 65 anos) devido à ausência de condições adequadas de saúde, baseando-se nos trabalhos de Luft (1974) e Haveman (1995). São utilizados dados individuais fornecidos pela Pesquisa Nacional sobre Saúde e Nutrição (PNSN) de 1989. A estimação das perdas de rendimentos é obtida por meio da estimação das equações de participação no mercado de trabalho e rendimentos, corrigindo-se para um possível viés de seletividade amostral. As perdas de rendimentos dos indivíduos que estão atuando no mercado de trabalho foram superiores às perdas dos homens que estão fora do mercado de trabalho. Além disso, as estimativas demonstraram que as regiões Sudeste e a Nordeste e o setor Urbano foram os que apresentaram as maiores perdas de rendimentos. Os resultados permitem concluir que as perdas de rendimentos dos trabalhadores, devido às condições adversas de saúde, são elevadas no Brasil, atingindo 1,66 bilhões de dólares, ou 258 dólares por homem adulto por ano.

Palavras-chave: perdas de rendimento, condições de saúde, regiões do Brasil, mercado de trabalho.

ABSTRACT

The major objective of this study is to estimate the earnings losses by adult men (18 to 65 years old) due to the lack of adequate health conditions, based on Luft (1974) and Haveman (1995). Data from the national nutritional and health survey (PNSN) from 1989 were used. The estimates of the productivity losses are obtained through the estimation of labor market participation and earnings equations, correcting for possible sample selectivity bias. The losses of individual's earnings, for those in the labor market, were higher than for those out of it. In addition, estimates show that southeastern and northeastern regions and the urban sector present the highest earnings losses. The results permit to conclude that labor earnings losses, due to adverse health conditions, are high in Brazil, estimated in 1.6 billion dollars per year, or 258 dollars per adult man per year.

Key words: earnings losses, health conditions, Brazilian regions, labor market.

§ Professora da ESALQ, Universidade de São Paulo.

Recebido em abril de 1999. Aceito em junho de 1999.

1 Introdução

Ter boa saúde está relacionado não só ao bem-estar do indivíduo mas também a ganhos econômicos. É certo que melhores condições de saúde permitem ao trabalhador realizar suas atividades de maneira mais produtiva, evitando, do ponto de vista econômico, elevadas perdas para a sociedade. Os ganhos econômicos advindos de melhores condições de saúde ocorrem devido à diminuição dos dias perdidos em razão de doenças, maior produtividade, maior disposição ao trabalho, maiores oportunidades de obter empregos e salários melhores e mais anos de vida ativa para o trabalhador.

Condições adequadas de saúde beneficiam principalmente os pobres, que podem ter os recursos que seriam gastos com tratamentos liberados. A família de baixa renda é a mais afetada quando um de seus membros trabalhador adoece. Como em muitos países em desenvolvimento, é raro haver seguro-uesemprego ou contra incapacidade, a família torna-se desamparada ou forçada a ter outros membros sadios trabalhando mais para compensar a perda de renda.

Neste sentido, ressalta-se a importância de se estudar os efeitos de condições inadequadas de saúde na produtividade da força de trabalho por meio das perdas de rendimento e ausência ao trabalho no Brasil.

Diversos estudos mediram e mostraram a importância da saúde na participação dos indivíduos no mercado de trabalho e na produtividade. (Bowen & Finagan, 1969; Davis, 1972; Luft, 1974; Rice, 1985; e Haveman *et alii*, 1995, entre outros) Entretanto, no Brasil, estas pesquisas são escassas. Considerando o pequeno número de estudos que estimam as perdas de rendimento da força de trabalho ocasionadas por condições de saúde inadequadas no Brasil, procurar-se-á oferecer uma contribuição empírica para a melhor compreensão desse tema.

O presente estudo tem como objetivo geral estimar as perdas de rendimento dos homens adultos (18 a 65 anos) provocadas pela ausência de condições adequadas de saúde, utilizando os dados individuais da Pesquisa Nacional sobre Saúde e Nutrição (PNSN) de 1989.¹

Baseando-se principalmente nos trabalhos de Luft (1974) e Haveman *et alii* (1995), as perdas de rendimento ocasionadas por ausência ao trabalho e por baixa produtividade são

1 Saúde, neste estudo, será mensurada pelo Índice de Massa Corporal (IMC) para os adultos. Esse índice é obtido pela divisão do peso pelo quadrado da altura.

estimadas. Para tal, os dados são divididos em duas amostras: homens adultos saudáveis e doentes, sendo então estimadas as equações de participação no mercado de trabalho e de rendimentos para os trabalhadores saudáveis. Com os coeficientes estimados destas equações e as médias das variáveis extraídas da amostra de doentes, são calculadas as proporções de doentes que, mantidas as suas características socioeconômicas, trabalhariam e receberiam rendimentos maiores, como se fossem saudáveis. Para evitar um possível viés de seletividade amostral, causado quando somente indivíduos que recebem rendimentos entram nos cálculos, a equação é corrigida pelo procedimento de Heckman (1974, 1980). Com base nestas estimativas e outras informações da amostra de dados, é possível obter um valor numérico para as perdas de rendimentos, por região e setor no Brasil.

Estas informações permitem mensurar os danos causados por problemas de saúde na renda das famílias e da população em geral, fornecendo subsídios para o planejamento de ações governamentais na melhoria dos programas de assistência e prevenção a problemas de saúde.

2 Dados

2.1 A amostra utilizada

O presente estudo utiliza dados individuais da Pesquisa Nacional sobre Saúde e Nutrição (PNSN), realizada pelo Instituto Nacional de Alimentação e Nutrição (INAN), Fundação Instituto Brasileiro de Geografia e Estatística (IBGE) e Instituto de Planejamento Econômico e Social (IPEA), no ano de 1989. O objetivo desta pesquisa foi avaliar o estado nutricional e de saúde da população, utilizando medidas antropométricas de peso e altura, com base nos padrões de referência recomendados pela Organização Mundial de Saúde (OMS). A PNSN inclui 17.920 domicílios, com cerca de 63.000 indivíduos, em 363 municípios.

Os questionários da pesquisa foram divididos em características demográficas básicas e características de nutrição e saúde. As características demográficas básicas compreendem as informações sobre a população, condições de habitação e instrução e participação no mercado de trabalho. Já as características de nutrição e saúde englobam as informações referentes à antropometria e sintomas de doenças, acesso a serviços de saúde, histórico obstétrico da mulher e aleitamento materno. Ressalte-se que o nível de desagregação na PNSN é bastante amplo, o que possibilita a realização de análises detalhadas.

O presente estudo se limita a avaliar os homens adultos, isto é, de 18 a 65 anos de idade.

2.2 A utilização de medidas antropométricas para se captar o estado de saúde

De acordo com James *et alii* (1988), o Subcomitê de Nutrição e Saúde e o Grupo Consultivo Internacional de Energia e Dieta identificou a necessidade de uma metodologia para diagnosticar a proporção de pessoas com problemas nutricionais na população adulta. É certo que atualmente existem mecanismos precisos para se mensurar a gordura corporal. Entretanto, estes mecanismos são complexos e custosos, tendo aplicação para uso clínico e individual. Desta forma, ocorreu o desenvolvimento de um método indireto baseado na relação do peso (P) e estatura (E) do indivíduo observado.

Os estudos realizados pelo pesquisador francês Quetelet, no final do século passado, indicaram ser a relação entre o peso e o quadrado da altura (peso expresso em quilogramas e estatura em metros) um indicador que, isoladamente, permite a avaliação da deficiência crônica de energia dos indivíduos adultos, em nível epidemiológico. Este indicador é conhecido como Índice de Massa Corporal (IMC), o qual respeita alguns importantes critérios de qualidade. A sua precisão não varia ao se considerar pessoas baixas ou altas.

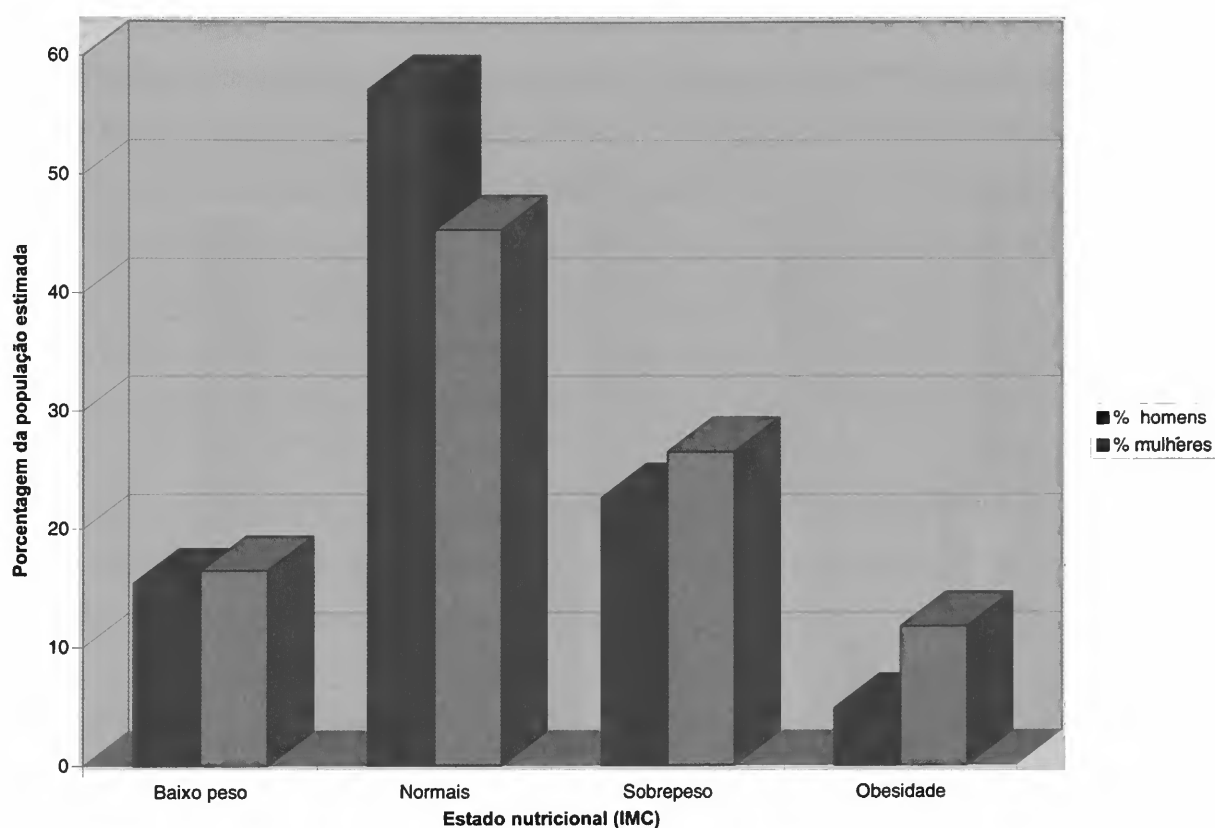
A conclusão de uma vasta revisão bibliográfica realizada por Anjos (1992) é a de que *“apesar do IMC não indicar a composição corporal, a facilidade de sua mensuração e a grande disponibilidade de dados de massa corporal e estatura, além da sua relação com morbi-mortalidade, parecem ser motivos suficientes para a utilização do índice como indicador do estado nutricional em estudos epidemiológicos.”* Segundo James *et alii* (1988), valores muito baixos do IMC refletem baixas reservas energéticas do organismo, resultando em menor produtividade e maior risco de morbi-mortalidade.

Vale ressaltar que a escolha do ponto de corte do IMC para a classificação de baixo e alto peso é matéria de amplo debate entre os pesquisadores internacionais. É certo, porém, que diversos estudos epidemiológicos (Jarret *et alii*, 1982; Lew, 1985; Sorlie, 1980), citados pelo INAN (1991), vêm indicando uma associação entre os valores extremos de peso relativo ($IMC < 20$ e $IMC \geq 30$) e a mortalidade em adultos, o que torna possível a utilização do índice também como indicador de saúde.

De acordo com Monteiro (1995), o Comitê de Especialistas da Organização Mundial de Saúde (OMS) concluiu que o percentual de indivíduos que seriam, por constituição física, magros, em uma população, representaria cerca de 3% a 5%. Ressalte-se que se um indivíduo apresentar um peso abaixo do ponto de corte, isto não significa, necessariamente, que ele esteja desnutrido. Muito provavelmente, este indivíduo encontra-se no limite do seu equilíbrio orgânico, podendo tornar-se desnutrido.

De acordo com a Figura.1, e tomando-se como referência os pontos de corte propostos pela OMS, a população brasileira adulta (18 a 65 anos) apresenta alta prevalência de baixo peso (IMC < 20), sobrepeso ($30 > \text{IMC} \geq 25$) e obesidade ($\text{IMC} \geq 30$), com uma taxa de 48,9% da população estimada. Somente 45,4% das mulheres e 57% dos homens são considerados normais ($20 \leq \text{IMC} < 25$).

Figura 1
Prevalência (%) de Baixo Peso, Padrão Normal, Sobrepeso e Obesidade na População Brasileira de 18 a 65 Anos, Segundo Sexo

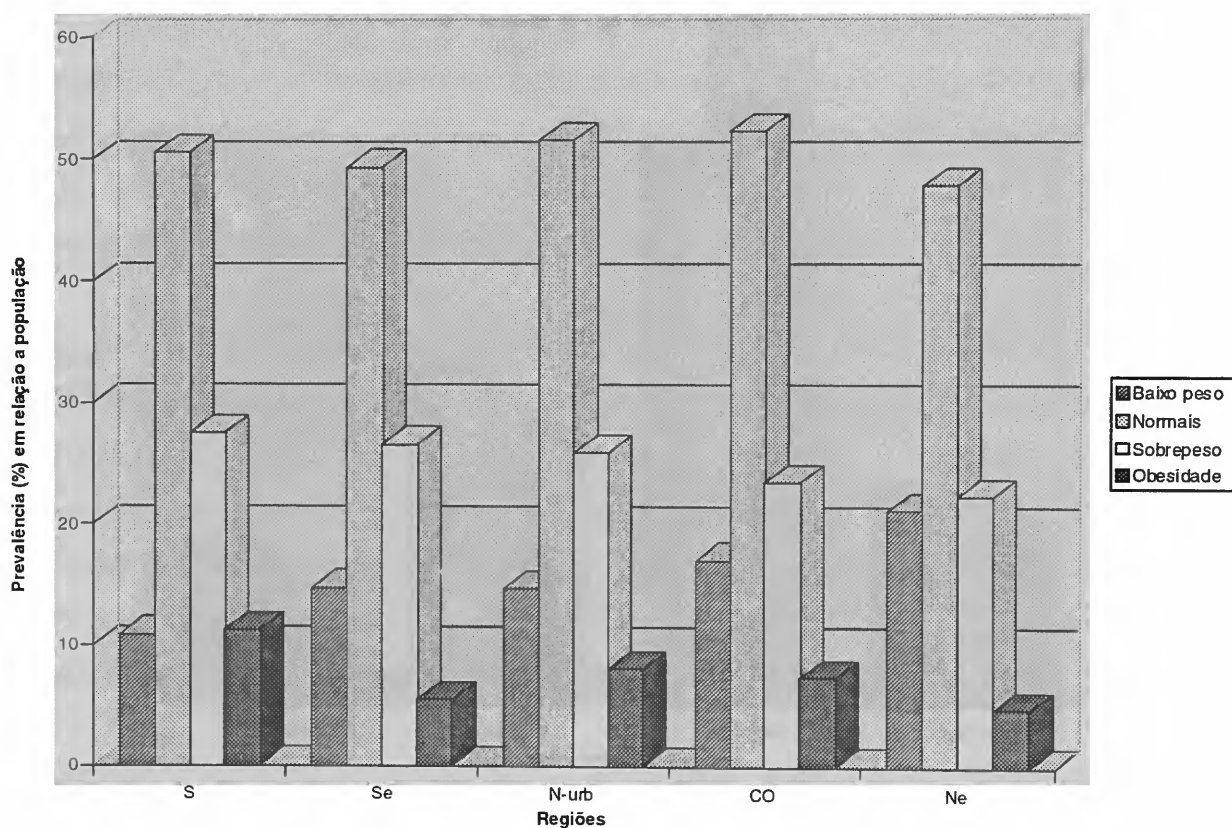


Aproximadamente 27 milhões de adultos brasileiros apresentam algum grau de excesso de peso. A obesidade, excesso de peso acentuado, atinge 8% da população, o que equivale a 6,8 milhões de indivíduos obesos. Os casos de sobrepeso e obesidade são mais frequentes entre as mulheres do que entre os homens. Os casos de sobrepeso e obesidade no sexo feminino atinge 38,2% da população estimada, cerca de 16 milhões de pessoas, enquanto que a participação masculina fica em 27,4% da população, equivalente a 11 milhões. A maior prevalência de excesso de peso em mulheres é tendência mundial.

A prevalência de baixo peso não apresentou diferença significativa entre os sexos. Aproximadamente seis milhões de homens (15,4%) e sete milhões de mulheres (16,5%) têm baixo peso, atingindo cerca de 13 milhões de adultos brasileiros

Quando a prevalência de baixo peso é analisada de acordo com a região onde reside os indivíduos, observa-se que a região Nordeste tem 21,7% dos adultos com baixo peso (Figura 2). A região Sul apresenta a menor frequência, tendo aproximadamente 10% de sua população de 18 a 65 anos. Se este problema for analisado em termos absolutos, constata-se que é na Região Sudeste que o problema se agrava, pois cerca de 6 milhões de pessoas têm baixo peso (14,7% da população).

Figura 2
Prevalência (%) de Baixo Peso, Padrão Normal, Sobrepeso e Obesidade na População Brasileira de 18 a 65 Anos, Segundo Regiões



O problema de excesso de peso apresenta-se em todas as regiões brasileiras. Em termos relativos a região Sul apresenta maior participação porcentual nos problemas de obesidade e sobrepeso, com 34% dos homens e 45% das mulheres, totalizando cerca de 5 milhões de adultos. As regiões Sudeste, Norte, Centro-Oeste e Nordeste possuem, respectivamente, 36%, 34%, 31% e 24% de seus adultos com problemas relacionados ao excesso de peso.

Ao se analisar a prevalência de baixo peso em relação à renda, o INAN/IBGE/IPEA (1989) cita que, para os jovens, esta variável parece não atuar de forma significativa. A diferença nos grupos de renda se acentua com a idade. O grupo de idosos de menor renda apresenta mais do que o dobro da prevalência de baixo peso quando comparado com aqueles cuja renda mensal excede mais de dois salários mínimos *per capita*.

3 Revisão bibliográfica

O desenvolvimento de estudos com o propósito de analisar o impacto das condições de saúde sobre o salário surgiu no início da década de 60.

O trabalho sobre determinação da renda realizado por Morgan *et alii* (1962), citado por Luft (1974), utiliza uma amostra de indivíduos adultos (com idade entre 18 e 64 anos) dos Estados Unidos no ano de 1967, estratificados por sexo, raça e estado de saúde. A partir deste conjunto de informações foram estimadas equações de salário, participação na força de trabalho, e emprego, entre outras, sendo que os coeficientes derivados da análise da amostra das pessoas em bom estado de saúde foram posteriormente utilizados para os valores observados da amostra das pessoas doentes, supondo que as pessoas doentes estivessem em boas condições de saúde. Neste estudo, os autores concluíram que o estado de saúde foi importante na estimação da participação na força de trabalho, horas trabalhadas por ano e taxa de salário horária, com grande impacto sobre a taxa de participação na força de trabalho. Os autores também procuraram estimar o custo da incapacidade. Entretanto, o pequeno número de observações presentes na amostra utilizada fez com que poucas pessoas fossem incluídas como incapazes, afetando os resultados do estudo.

Diversos outros estudos como, por exemplo, o de Parnes & Meyer (1971), citado por Luft (1974) e Davis (1972), indicaram que a saúde é importante para a participação na força de trabalho, semanas trabalhadas, taxa de desemprego e horas trabalhadas.

Bowen & Finagan (1969) explicaram os possíveis fatores que determinavam a participação na força de trabalho em grupos específicos da população, incluindo homens com idade entre 25 e 54 anos, mulheres casadas com idade entre 14 e 54 anos, pessoas idosas com idade acima de 54 anos e jovens com idade entre 14 e 24 anos. Os autores desenvolveram uma estrutura conceitual para determinar a taxa de participação na força de trabalho. De maneira geral, eles destacaram a importância do estado de saúde na determinação da participação na força de trabalho.

Pode-se dizer que uma das primeiras economistas que procurou estimar as perdas agregadas dos salários devido a doenças, por meio de indicadores de morbidade e mortalidade, foi **D. P. Rice**. Uma das dificuldades enfrentadas por ela foi a inexistência de estimativas, a não ser aquelas feitas pelo economista **J. N. Morgan**, dos efeitos da incapacidade sobre o mercado de trabalho.

Luft (1974) teve como interesse primário estimar os efeitos da saúde sobre o salário. O estudo considera as diferenças nos valores das variáveis dependentes para duas

populações definidas pelo estado de saúde (doentes e saudáveis). Quatro amostras categorizadas por raça e sexo foram divididas de acordo com o grupo de pessoas doentes ou saudáveis, utilizando uma pesquisa nacional domiciliar denominada *Survey of Economic Opportunity* (SEO). Ressalta-se que os coeficientes derivados da amostra dos indivíduos saudáveis foram combinados com os valores observados da amostra dos indivíduos doentes para estimar o comportamento destas pessoas se elas fossem saudáveis.

Os comentários sobre a participação da força de trabalho no mercado e a definição das variáveis dependentes elaborados no estudo de Bowen & Finegan (1969) influenciaram, de maneira decisiva, os procedimentos metodológicos realizados por Luft (1974).

Ao construir o seu modelo, Luft (1974) procurou determinar um grande número de variáveis explicativas, aproveitando a enorme gama de informações presentes na SEO. A variável estado de saúde, ponderada pelo peso de cada indivíduo na amostra, serviu para determinar a população de pessoas saudáveis e a população de pessoas inabilitadas, indicando a carência de saúde dos indivíduos. Rice *et alii* (1985) estimaram, para os Estados Unidos, a perda de produtividade causada por problemas de saúde, caracterizando-as como **perdas salariais**. O indicador proposto pelos autores tinha como propósito mensurar como o salário individual responderia à incapacidade/limitação de saúde, refletindo a preferência individual por trabalho *versus* lazer. A perda salarial estimada foi de 67,8 bilhões de dólares em 1980, ou aproximadamente 1,7% do PNB.

Outro importante e recente trabalho realizado sobre as estimativas das perdas de rendimento devido às condições de saúde é o trabalho realizado por Haveman *et alii* (1995), que propõe a utilização de um novo procedimento metodológico para determinação das referidas perdas. Os autores elaboram um novo conceito de perdas salariais denominado perda da capacidade de rendimento, que é a diferença entre o montante de dinheiro que as pessoas potencialmente poderiam ganhar se estivessem livres de inabilidade ou limitações de saúde e o que realmente ganham. Este estudo analisou a perda de capacidade de rendimento da população americana em 1973 e 1978. No ano de 1973 a perda de capacidade de rendimento ficou em 5,3% do Produto Nacional Bruto e no ano de 1988 a mesma perda foi estimada em 4,5%.

4 Método

Para o cálculo das perdas de rendimento, primeiramente são obtidas duas amostras distintas e, a partir daí, estimadas duas equações. Estas equações, que são de participação no mercado de trabalho e de rendimentos, englobam os indivíduos **saudáveis** (IMC \geq 20),

de acordo com o Índice de Massa Corporal (IMC), disponível na PNSN.² É importante ressaltar que as equações são ponderadas pelo fator de expansão da amostra

Optou-se por excluir os obesos da amostra de doentes e utilizar apenas o ponto de corte inferior para delimitar a amostra de saudáveis e doentes, já que a obesidade é um problema específico de vida sedentária e excesso de alimentação, na maioria das vezes relacionado a rendas elevadas. Desta forma, abordam-se os desnutridos, os quais compõem a camada mais carente da população, com baixa renda, falta de tratamento adequado de saúde, falta de infra-estrutura básica no domicílio e má alimentação. Já o obeso, na grande maioria, pode, com dieta e exercícios, perder peso e melhorar seu estado de saúde. A solução não é tão simples para os desnutridos. Ademais, a inclusão dos indivíduos com obesidade extrema ($IMC \geq 30$) como doentes carece de maior respaldo teórico e de estudos em nível internacional.³

A Tabela 1 apresenta a amostra e a população de homens saudáveis e doentes, de acordo com o índice de massa corporal acima ou abaixo de 20, em cada região e setor no Brasil. Observa-se que a maior porcentagem de homens saudáveis está na região Sul e a menor na região Nordeste. O setor urbano apresenta maior porcentagem de homens saudáveis comparativamente ao setor rural.

Tabela 1

População de Homens Saudáveis, População de Homens Doentes e Variação Porcentual por Regiões e Setores do Brasil (Exclusiva a População Rural da Região Norte)

Regiões/ Setores	Saudáveis			Doentes		
	População	Amostra	%	População	Amostra	%
Sudeste	14.451.490	2.951	82,5	3.067.439	713	17,5
Sul	5.335.475	3.106	88,2	711.981	431	11,8
Centro-Oeste	2.141.330	2.656	80,9	505.844	645	19,1
Nordeste	7.331.023	2.721	78,8	1.974.627	742	21,2
Norte*	938.480	1.625	84,6	170.445	297	15,4
Urbano	22.933.263	7.403	83,8	4.414.088	1.373	16,2
Rural	7.264.535	5.656	78,3	2.016.248	1.455	21,7
Brasil	30.197.798	13.059	82,4	6.430.336	2.828	17,6

* Inclui somente o Norte Urbano.

2 Apesar de existir informação sobre sintomas de doenças, considera-se que as medidas antropométricas são menos subjetivas e, portanto, mais confiáveis como medidas de saúde.

3 Os resultados não foram satisfatórios ao se incluir os homens obesos, além dos com baixo peso.

A Tabela 2 mostra as variáveis a serem utilizadas nas equações de participação no mercado de trabalho e rendimentos. É importante ressaltar que a maioria das variáveis propostas por Luft (1974) está presente nessa tabela. Isto comprova a qualidade das informações da PNSN.

Tabela 2
Descrição das Variáveis Explicativas e Dependentes no Modelo

Variáveis	Definição
Variáveis Explicativas	
Idade	Idade do indivíduo em anos
Idade2	Idade do indivíduo elevada ao quadrado
Educação	Número de anos de escolaridade
Idade x educação	Idade vezes Escolaridade
Renda não salarial	Renda não salarial <i>per capita</i> em dólares
Filhos2	Número de filhos de até 2 anos de idade
Filhos3-5	Número de filhos de 3 a 5 anos de idade
Filhos6-12	Número de filhos de 6 a 12 anos de idade
Filho>=13	Número de filhos com 13 anos ou mais
Filha>=13	Número de filhas com 13 anos ou mais
Norte	= 1 se o indivíduo reside na região Norte
Centro-Oeste	= 1 se o indivíduo reside na região Centro-Oeste
Sudeste	= 1 se o indivíduo reside na região Sudeste
Sul	= 1 se o indivíduo reside na região Sul
Urbano	= 1 se o indivíduo reside no setor urbano
Pardo	= 1 se o indivíduo é da raça parda
Bramarel	= 1 se o indivíduo é da raça branca ou asiático
Chefe	= 1 se o indivíduo é chefe na família
Cônjuge	= 1 se o indivíduo é cônjuge na família
Filho	= 1 se o indivíduo é filho na família
Outros	= 1 se o indivíduo é pertencente a outro parentesco
Variáveis Dependentes	
Trabalha	= 1 se o indivíduo possui rendimento positivo, e zero em caso contrário
Logrhr	Logaritmo do rendimento-hora em dólares

As variáveis explicativas procuram captar um amplo conjunto de informações, tais como características dos indivíduos, localização geográfica, cor e condição na família.

A participação do indivíduo na força de trabalho e o rendimento obtido são as variáveis dependentes. Será atribuído valor 1 à variável **Trabalha** quando o indivíduo receber qualquer rendimento positivo e zero em caso contrário.⁴ A variável **Logrhr** é o logaritmo do rendimento obtido por hora de trabalho.

As Tabelas 3 e 4 contêm as médias das variáveis provenientes das amostras utilizadas, respectivamente, nas equações de participação no mercado de trabalho e de rendimentos. Pode-se verificar que a idade média dos homens doentes é inferior à dos saudáveis em todas as regiões e setores. A proporção de filhos, em diferentes idades, também é superior para os homens saudáveis com relação aos doentes.

Ao se analisar a variável educação, constata-se que os anos de estudo dos homens saudáveis são, em todas as regiões e setores, superiores comparativamente aos indivíduos doentes. Este fato é esperado tendo em vista que quanto maior o grau de escolaridade melhor tende a ser a sua saúde. (Folland, Goodman & Stano, 1993)

Outra importante constatação é de que a renda não salarial dos homens saudáveis é superior, na maioria dos casos, à renda dos doentes.

5 Resultados

Conforme ressaltado no capítulo anterior, serão estimadas as equações de participação no mercado de trabalho e de rendimentos para o cálculo das perdas de rendimento. Pode ocorrer viés de seletividade amostral se forem estimadas equações de rendimento, por mínimo quadrados, apenas para os indivíduos que participam do mercado de trabalho. Visando solucionar este problema, é utilizado o procedimento de Heckman. Primeiramente é estimada a equação de participação no mercado de trabalho por meio do modelo próbite. A variável **Trabalha** assume valor um se o indivíduo tem rendimento positivo, e valor zero, caso contrário. Com base nos coeficientes estimados no modelo próbite obtém-se a variável **lambda**, que é em seguida utilizada na equação de rendimentos. (Kassouf, 1994) Desta forma, são corrigidos os problemas gerados pela não inclusão dos homens adultos que não recebem salários.

4 A variável **Trabalha** possui a limitação de não incluir os indivíduos que trabalham mas não recebem qualquer forma de rendimento.

Tabela 3
Médias das Variáveis Obtidas da Amostra de Homens Saudáveis e Doentes que Participam ou Não do Mercado de Trabalho, por Região e Setor

	CENTRO-OESTE		SUDESTE		SUL		NORDESTE		NORTE		URBANO		RURAL	
	Saudáveis	Doentes	Saudáveis	Doentes	Saudáveis	Doentes	Saudáveis	Doentes	Saudáveis	Doentes	Saudáveis	Doentes	Saudáveis	Doentes
Idade	35,53	32,66	36,88	33,55	36,44	34,51	35,70	33,09	35,25	32,28	36,46	32,32	36,10	35,80
Idade2	1416,	1247,	1524,	1302	1484	1385,	1441,	1283,	1401,	1216,	1490,	1214,	1472,	1483,
Educação	5,450	4,803	6,068	5,375	5,557	4,841	3,459	2,757	6,097	5,391	6,195	5,587	2,479	2,017
Idade x Educação	96,16	65,96	117,6	89,05	107,8	80,15	63,94	39,33	109,6	80,30	116,8	84,86	51,53	39,87
Renda não salarial	16,71	4,958	31,61	34,30	29,48	7,060	21,07	7,125	22,79	22,23	33,94	27,52	6,538	4,536
Filho2	0,2251	0,1848	0,1854	0,1783	0,2084	0,1517	0,2992	0,2735	0,2431	0,2127	0,1984	0,1781	0,2952	0,2672
Filho3-5	0,2379	0,1609	0,1963	0,1664	0,2114	0,1632	0,2987	0,2695	0,2863	0,1988	0,2122	0,1591	0,2843	0,2837
Filho6-12	0,5431	0,4670	0,4838	0,3967	0,5239	0,3677	0,6860	0,6055	0,6485	0,5110	0,5156	0,3908	0,6555	0,6312
Filho13	0,2941	0,2541	0,3093	0,2336	0,3082	0,2827	0,3698	0,2990	0,3262	0,2647	0,3077	0,2018	0,3723	0,3923
Filha13	0,2324	0,2237	0,2389	0,1606	0,2323	0,2268	0,2997	0,2123	0,2526	0,1928	0,2496	0,1549	0,2614	0,2655
Norte											0,0409	0,0386		
Centro-Oeste											0,0689	0,0805	0,0772	0,07456
Sudeste											0,5453	0,5690	0,2680	0,2757
Sul											0,1573	0,0986	0,2380	0,1373
Urbano	0,7382	0,7028	0,8653	0,8188	0,6760	0,6111	0,5869	0,4768						
Pardo	0,5397	0,5704	0,2573	0,2627	0,1141	0,1214	0,7630	0,7805	0,7157	0,7363	0,3598	0,3884	0,4812	0,5620
Bramarel	0,4266	0,4110	0,6617	0,6707	0,8578	0,8371	0,1918	0,1766	0,2718	0,2579	0,5804	0,5626	0,4685	0,3822
Chefe	0,7803	0,5500	0,7294	0,5600	0,7646	0,5902	0,7156	0,5715	0,7065	0,5584	0,7359	0,5392	0,7329	0,6248
Filho	0,1714	0,3698	0,2217	0,3740	0,1971	0,3613	0,2469	0,3734	0,2310	0,3697	0,2169	0,3909	0,2305	0,3305
OBS	2656	645	2951	713	3106	431	2721	742	1625	297	7403	1373	5656	1455

Tabela 4
Médias das Variáveis Obtidas da Amostra de Homens Saudáveis e Doentes que Recebem Rendimentos, por Região e Setor

Variáveis	CENTRO-OESTE		SUDESTE		SUL		NORDESTE		NORTE		URBANO		RURAL	
	Saudáveis	Doentes	Saudáveis	Doentes	Saudáveis	Doentes	Saudáveis	Doentes	Saudáveis	Doentes	Saudáveis	Doentes	Saudáveis	Doentes
Idade	35,15	32,81	36,45	33,16	36,14	35,13	36,30	34,13	35,59	33,46	36,06	32,12	36,83	36,95
Idade ²	1370,	1244	1476	1261	1444	1412,	1468	1336	1413	1273	1442	1182	1514	1553
Educação	5,572	4,858	6,127	5,465	5,668	4,759	3,471	2,605	6,160	5,226	6,302	5,652	2,426	2,037
Idade x Educação	98,78	68,22	116,9	90,55	109,3	82,90	66,44	39,54	112,9	84,20	117,5	87,08	53,51	42,84
Norte											0,0411	0,0377		
Centro-Oeste											0,0711	0,0845	0,0809	0,0804
Sudeste											0,5505	0,5873	0,2827	0,3008
Sul											0,1596	0,0983	0,2408	0,1355
Urbano	0,7429	0,6977	0,8649	0,8109	0,6855	0,6143	0,5963	0,4662			0,3547	0,3674	0,4770	0,5646
Pardo	0,5343	0,5617	0,2592	0,2478	0,1172	0,1271	0,7625	0,7998	0,7087	0,7152				
Branarel	0,4313	0,4165	0,6597	0,6810	0,8547	0,8293	0,1882	0,1598	0,2787	0,2811	0,5853	0,5808	0,4682	0,3778
Lambda	0,1742	0,2869	0,2033	0,3045	0,2051	0,3193	0,2886	0,4225	0,2084	0,3104	0,2128	0,3408	0,2480	0,3400
OBS	2361	528	2597	582	2681	328	2210	529	1422	221	6456	1040	4815	1148

5.1 Equações de participação no mercado de trabalho e rendimentos

As Tabelas 5 e 6 apresentam, respectivamente, os coeficientes da equação de participação no mercado de trabalho e de rendimentos, para os homens saudáveis, por região e setor.

Os resultados das referidas tabelas não serão detalhadamente discutidos, pois são cálculos intermediários, cujo objetivo é estimar os coeficientes para utilizá-los posteriormente no cálculo das perdas.

Tabela 5
Equação de Participação no Mercado de Trabalho para Homens Saudáveis, por Região e Setor - Modelo Próbite

Variáveis	Centro-Oeste	Sudeste	Sul	Nordeste	Norte	Urbano	Rural
Constante	-0,749 (-1,55)	-0,0561 (-0,14)	-0,386 (-0,93)	-0,702 (-1,86)***	-1,433 (-2,36)**	-0,724 (-2,89)*	0,0596 (0,21)
Idade	0,116 (5,28)*	0,0854 (4,44)*	0,0865 (4,65)*	0,0928 (5,34)*	0,125 (5,04)*	0,101 (8,53)*	0,0483 (3,75)*
Idade ²	-0,00181 (-7,04)*	-0,00130 (-5,84)*	-0,00141 (-6,56)*	-0,00127 (-6,14)*	-0,00162 (-5,52)*	-0,00152 (-11,07)*	-0,000753 (-4,97)*
Educação	-0,00715 (-0,49)	0,0218 (1,76)**	0,0332 (2,50)**	0,0194 (1,53)	0,0261 (1,67)**	0,0223 (2,96)*	-0,0391 (-3,03)*
Idade x Educação	0,00164 (2,43)**	0,000189 (0,35)	-0,000420 (-0,74)	0,000189 (0,31)	0,000161 (-0,21)	0,0000716 (0,22)	0,00253 (4,19)*
Renda não salarial	-0,000982 (-4,64)*	-0,00112 (-6,16)*	-0,000413 (-3,54)*	-0,000556 (-2,90)*	-0,00105 (-3,78)*	-0,000682 (-7,54)*	-0,00169 (-5,50)*
Filho ²	-0,0506 (-0,53)	-0,155 (-1,71)***	-0,0210 (-0,20)	0,105 (1,47)	0,00483 (0,04)	-0,0918 (-1,65)***	0,131 (2,40)**
Filho ³⁻⁵	0,0142 (0,15)	0,202 (1,11)**	0,280 (2,61)*	-0,0393 (-0,56)	-0,0489 (-0,48)	0,131 (2,24)*	0,0363 (0,67)
Filho ⁶⁻¹²	-0,0276 (-0,55)	-0,00340 (-0,07)	-0,0601 (-1,27)	0,0778 (1,97)**	0,0186 (0,33)	0,0155 (0,56)	0,0378 (1,27)
Filho ¹³	-0,0496 (-0,88)	0,0237 (0,48)	0,0351 (0,71)	-0,0235 (-0,53)	0,00144 (0,02)	0,0113 (0,37)	0,00545 (0,16)
Filha ¹³	0,0828 (1,18)	-0,0139 (-0,24)	0,143 (2,22)**	0,0331 (0,65)	-0,0511 (-0,66)	0,0499 (1,41)	0,00595 (0,14)
Centro-Oeste						0,282 (3,04)*	0,347 (3,83)*
Sudeste						0,226 (4,08)*	0,558 (8,68)*
Sul						0,210 (2,86)*	0,392 (5,65)*
Norte						0,185 (1,73)***	
Urbano	0,115 (1,38)	-0,0552 (-0,57)	0,0943 (1,36)	0,0842 (1,25)			
Pardo	-0,107 (-0,48)	0,0451 (0,36)	0,180 (0,86)	-0,267 (-1,66)***	-0,0595 (-0,15)	-0,0130 (-0,15)	-0,175 (-1,46)
Bramarel	-0,0233 (-0,10)	-0,0177 (-0,15)	-0,0269 (-0,14)	-0,319 (-1,85)***	0,0588 (0,15)	0,00540 (0,06)	-0,427 (-3,52)*
Chefe	0,691 (3,70)*	0,266 (1,56)	0,619 (3,84)*	0,531 (3,32)*	0,697 (4,04)*	0,412 (4,03)*	0,613 (5,05)*
Filho	-0,178 (-0,97)	-0,446 (-2,69)*	-0,438 (-2,87)*	-0,333 (-2,17)**	-0,0396 (-0,25)	-0,314 (-3,20)*	-0,409 (-3,42)*
Teste Razão Veross.	335,74*	233,72*	459,15*	338,28*	145,21*	659,97*	516,99*
OBS	2656	2951	3106	2721	1625	7403	5656

Obs.: Os testes t estão entre parênteses abaixo dos coeficientes.

* Significativo a 1%.

** Significativo a 5%.

*** Significativo a 10%.

Tabela 6
Equação de Rendimentos para Homens Saudáveis, por Região e Setor

Variáveis	Centro-Oeste	Sudeste	Sul	Nordeste	Norte	Urbano	Rural
Constante	-2,314 (-8,16)*	-3,20 (-13,62)*	-3,363 (-12,72)*	-2,545 (-9,04)*	-3,291 (-7,70)*	-3,143 (-17,61)*	-2,353 (-15,24)*
Idade	0,0565 (4,06)*	0,0960 (8,34)*	0,104 (8,41)*	0,0540 (4,05)*	0,131 (7,21)*	0,0993 (11,79)*	0,0453 (6,33)*
Idade2	-0,000558 (-3,04)*	-0,00107 (-7,35)*	-0,00113 (-7,21)*	-0,000599 (-3,67)*	-0,00152 (-6,94)*	-0,00112 (-10,32)*	-0,000471 (-5,41)*
Educação	0,0934 (13,48)*	0,106 (18,98)*	0,111 (16,63)*	0,0961 (12,28)*	0,0884 (10,15)*	0,104 (28,26)*	0,0878 (10,97)*
Idade x Educação	0,00166 (4,64)*	0,000796 (2,94)*	0,000539 (1,64)***	0,00154 (4,17)*	0,000863 (2,03)**	0,000962 (5,27)*	0,00107 (3,11)*
Centro-Oeste		-				0,312 (7,00)*	0,358 (8,00)*
Sudeste		-				0,257 (8,71)*	0,227 (6,80)*
Sul						0,168 (4,53)*	0,226 (6,11)*
Norte						0,306 (5,72)*	
Urbano	0,272 (6,50)*	0,382 (8,53)*	0,342 (9,64)*	0,250 (6,33)*			
Pardo	0,102 (1,07)	0,0965 (1,64)***	0,0107 (0,10)	0,148 (1,82)***	0,0709 (0,37)	0,0999 (2,29)**	0,0949 (1,86)***
Bramarel	0,278 (2,89)*	0,315 (5,68)*	0,262 (2,82)*	0,342 (3,80)*	0,330 (1,69)***	0,343 (8,13)*	0,194 (3,69)*
Lambda	-0,338 (-1,96)**	-0,0292 (-0,19)	-0,216 (-1,60)	-0,180 (-1,39)	0,0379 (0,17)	-0,122 (-1,15)	-0,119 (-1,59)
R2	0,37	0,43	0,39	0,37	0,32	0,41	0,21
Teste F	176,91*	242,14*	210,65*	163,61*	93,26*	408,19*	124,69*
OBS	2361	2597	2681	2210	1422	6456	4815

Obs.: Os testes t estão entre parênteses abaixo dos coeficientes.

* Significativo a 1%.

** Significativo a 5%.

*** Significativo a 10%.

Os coeficientes da equação de participação no mercado de trabalho foram estimados por máxima verossimilhança, utilizando-se o modelo próbite.

As variáveis idade e idade ao quadrado (idade2) refletem experiência e captam a depreciação do capital humano. Os resultados mostram coeficiente positivo para idade e negativo para idade2, isto é, à medida que a pessoa adquire mais experiência, aumentam as oportunidades de trabalho, até se atingir um determinado ponto, após o qual a

participação começa a decrescer, refletindo o envelhecimento do indivíduo e sua perda de habilidade no trabalho. (Berndt, 1991)

O aumento no número de anos de escolaridade (educação) aumenta a oportunidade dos indivíduos obter emprego. Em quase todas as equações os coeficientes foram positivos e significativos,

A variável renda não salarial é obtida somando-se todos as rendas com aluguel, pensão, aposentadoria etc., dentro do domicílio. Os coeficientes estatisticamente significativos e negativos mostram que quanto maior a renda não salarial menor é a participação do indivíduo no mercado de trabalho, como era de se esperar.

O número de filhos e filhas em diferentes idades foram incluídos na análise como determinantes da participação dos indivíduos na força de trabalho. Os resultados mostram que essas variáveis são quase todas não significativas e, quando significativas, não ocorreu um padrão para os sinais dos coeficientes.

Os coeficientes das raças (a variável preto foi omitida) dos indivíduos mostraram-se quase todos não significativos.

Com relação à condição do indivíduo na família, observa-se que os chefes participam mais do mercado de trabalho do que os filhos (a variável outros parentes foi omitida).

Os coeficientes da equação de rendimentos (Tabela 6) foram estimados por mínimos quadrados generalizados, utilizando-se o fator de expansão da amostra como peso.

Observe que o coeficiente da variável idade é positivo e o da variável idade ao quadrado é negativo, seguindo uma forma parabólica com um pico próximo à idade média de vida do indivíduo, como indica a teoria do capital humano. Segundo Mincer (1962), o capital humano deprecia-se com a idade e as pessoas ficam mais lentas e menos eficientes nas suas tarefas.

O aumento no número de anos de escolaridade leva a um acréscimo no salário dos trabalhadores. O coeficiente da variável educação, em todas as equações, é altamente significativo e positivo.

Homens trabalhando nas regiões Centro-Oeste, Sudeste, Sul e Norte recebem maiores salários do que os trabalhadores da região Nordeste, assim como os do setor urbano com relação ao rural. Este resultado é observado por meio dos coeficientes positivos e significativos das regiões e setor urbano. Fica constatado, também, que indivíduos da raça branca recebem maiores salários do que os da raça negra, indicando uma forma de discriminação contra os trabalhadores negros.

5.2 Cálculo das perdas de rendimento

Neste estudo obtêm-se as estimativas para dois tipos de perdas de rendimento. O primeiro tipo de perda inclui no seu cálculo apenas os doentes que estão fora do mercado de trabalho (Perda I). A segunda perda refere-se aos doentes que participam da força de trabalho, mas têm rendimentos inferiores aos saudáveis (Perda II).

A Tabela 7 mostra as etapas a serem seguidas para se obter a perda total de rendimentos (item 15), a qual é definida como sendo a soma do rendimento perdido por homens doentes devido à não participação na força de trabalho (item 8) e o rendimento perdido pelos homens doentes que participam da força de trabalho mas que recebem menos do que poderiam receber se fossem saudáveis (item 14).

A população de doentes (item 2) em cada região e setor do Brasil é obtida utilizando-se o fator de expansão da amostra selecionada, isto é, homens de 18 a 65 anos com IMC < 20. O item 3 fornece a porcentagem de homens saudáveis, isto é, com IMC \geq 20, que participam do mercado de trabalho, obtido diretamente da amostra. O item 4 é obtido multiplicando-se os coeficientes estimados da equação de participação no mercado de trabalho (Tabela 5) dos homens saudáveis pelas médias das variáveis obtidas da amostra de homens doentes (Tabela 3). Este método computa a porcentagem de indivíduos doentes que trabalhariam se fossem saudáveis.

A diferença entre a proporção de homens saudáveis trabalhando (item 3) e a proporção de homens doentes que trabalhariam se tivessem a estrutura dos homens saudáveis (item 4) fornece a diferença ajustada (item 5), que é a proporção de homens que não trabalham devido a problemas de saúde.⁵ Para se obter o número de homens que deixam de trabalhar no Brasil devido exclusivamente a problemas de saúde (item 6) multiplica-se a diferença ajustada (item 5) pela população de doentes (item 2).

O item 7 mostra o rendimento médio por hora recebido por homens doentes que trabalham, obtido diretamente da amostra. Portanto, a perda de rendimento que ocorre devido ao fato de os homens doentes não trabalharem (item 8), denominada perda I, é obtida multiplicando-se a população de homens que não trabalham por terem problemas de saúde (item 6) pelo rendimento que receberiam se trabalhassem (item 7).

O item 10 fornece a população de homens doentes participando do mercado de trabalho, obtida expandindo-se a amostra pelo fator de expansão enquanto o item 11

5 Utilizar simplesmente a diferença entre a proporção de homens saudáveis trabalhando e de doentes que não estão trabalhando, sem o ajuste, não estaria correto, pois as duas populações podem diferir em outras características, diferentes do estado de saúde.

mostra o rendimento médio em dólares por hora dos homens saudáveis, obtido diretamente da amostra. Já o item 12 é obtido multiplicando-se os coeficientes estimados da equação de rendimentos para a amostra de homens saudáveis (Tabela 6) pelas médias das variáveis da amostra de homens doentes (Tabela 4). O item 13, que é a diferença entre os itens 11 e 12, fornece o valor do rendimento que, devido exclusivamente a problemas de saúde, os homens que trabalham deixam de receber, isto é, mostra qual seria o rendimento, dado o comportamento dos indivíduos saudáveis e as características dos doentes. A perda de rendimento (item 14) é obtida multiplicando-se a diferença ajustada (item 13) pela população de doentes trabalhando (item 10), a qual é denominada perda II. A perda total de rendimento em dólares por hora é obtida da soma da perda I (item 8) com a perda II (item 14).

Tabela 7
Perdas de Rendimento por Hora em Dólares para os Homens Adultos do Brasil

	Centro-Oeste	Sudeste	Sul	Nordeste	Norte*	Urbano	Rural
1. Amostra total de doentes	645	713	431	742	297	1373	1455
2. População total de doentes	505.844	3.067.439	711.981	1.974.627	170.445	4.414.088	2.016.248
PERDA I - Perdas dos homens doentes que estão fora do mercado de trabalho							
3. Proporção de homens saudáveis que participam do mercado de trabalho	0,8889	0,8800	0,8632	0,8122	0,8751	0,8721	0,8513
4. Proporção dos homens doentes que trabalhariam se tivessem a estrutura dos homens saudáveis	0,8452	0,873	0,817	0,755	0,828	0,835	0,788
5. Diferença ajustada (3-4)	0,0437	0,0070	0,046	0,057	0,047	0,037	0,0633
6. Perda de força de trabalho em número de pessoas (2×5)	22.105	21.472	32.894	112.949	8.028	163.763	127.629
7. Rendimento médio dos doentes em dólares por hora de trabalho	0,62	0,7165	0,6058	0,35	0,802	0,72	0,34
8. Perda de rendimento em dólares por hora de trabalho (6 × 7)	13.708	15.385	19.927	39.611	6.439	117.925	43.266
PERDA II - Perdas dos homens doentes que permanecem no mercado de trabalho							
9. Amostra de doentes que permanecem no mercado de trabalho	528	582	328	529	221	1040	1148
10. População de doentes que permanecem no mercado de trabalho	411.890	2.462.799	544.168	1.401.779	128.337	3.400.585	1.548.388
11. Rendimento médio dos homens saudáveis em dólares por hora de trabalho	0,9146	1,0078	0,8728	0,4705	1,0063	1,0056	0,4169
12. Rendimento que os homens doentes receberiam se tivessem a estrutura dos homens saudáveis	0,7261	0,83	0,6878	0,3763	0,851	0,8149	0,3788
13. Diferença ajustada (11-12)	0,1885	0,1778	0,185	0,094	0,155	0,1907	0,038
14. Perda de rendimento por hora de trabalho em dólares (10×13)	77.641	437.886	100.671	132.048	19.931	648.492	58.994
15. Perda total de rendimento por hora de trabalho em dólares (8+14)	91.349	453.271	120.598	171.659	26.370	766.417	102.260

* Inclui somente o Norte urbano.

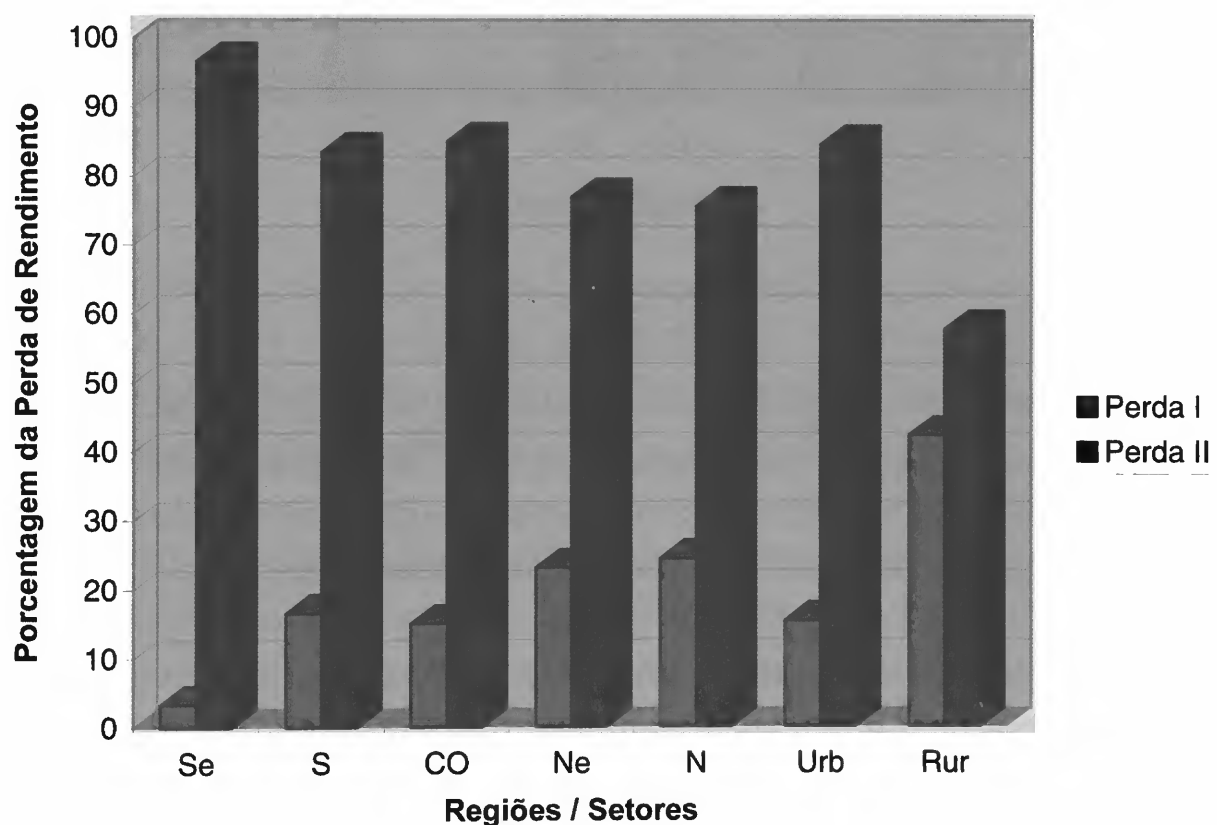
Os resultados apresentados na Tabela 7 indicam que as perdas de rendimento devido aos problemas relacionados à saúde são elevadas no Brasil.

A Figura 3 indica que as perdas das pessoas que estão atuando no mercado de trabalho (perda II) são superiores às perdas das pessoas que estão fora do mercado de trabalho (perda I). Nota-se que as regiões Norte e Nordeste apresentam as maiores perdas do tipo I, assim como o setor rural. Já as regiões Sudeste, Sul e Centro-Oeste e o setor urbano apresentam as maiores perdas do tipo II.

Da população total de doentes (item 2), a porcentagem mais elevada é a da população de doentes que permanece no mercado de trabalho (item 10), o que pode explicar a maior perda do tipo II.

Figura 3

Distribuição Percentual das Perdas de Rendimento por Hora de Trabalho do Tipo I e Tipo II Entre Regiões e Setores do Brasil

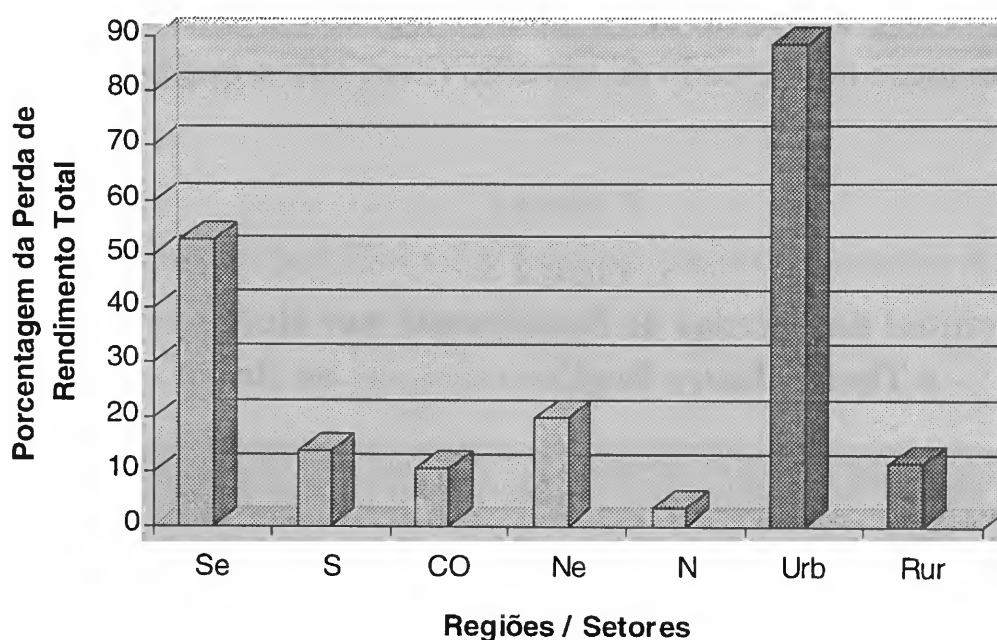


Considerando-se a perda de rendimento total, que é obtida por meio do somatório das perdas I e II, com relação à perda total nas cinco regiões, observa-se que as regiões Sudeste e Nordeste apresentam as maiores perdas percentuais com 52,5% e 19,9%, respectivamente. As regiões Sul, Centro-Oeste e Norte apresentam percentuais inferiores a 15% (Figura 4).

Quando a comparação é feita entre os setores, verifica-se que o setor urbano responde por 88,2% das perdas de rendimentos totais, sendo que o setor rural participa com apenas 11,8% das referidas perdas. Este resultado também é aceitável tendo em vista que o rendimento médio por hora de trabalho no setor rural é bastante inferior ao do setor urbano.

Figura 4

Distribuição Percentual das Perdas de Rendimento Totais por Hora de Trabalho Entre Regiões e Setores do Brasil



A região Sudeste apresenta elevada densidade demográfica e, conforme citado anteriormente, possui uma grande população de doentes. Já a região Nordeste possui elevados índices de pobreza absoluta. É de se esperar, portanto, que as perdas de rendimento devido a condições inadequadas de saúde sejam elevadas nestas regiões.

Para obter as perdas totais de rendimento por ano multiplicou-se as perdas de rendimento por hora de trabalho pelo número máximo de horas de trabalho semanal permitido pela legislação trabalhista (40 horas), pelo número de semanas em um mês (4 semanas) e pelo número de meses em um ano (12 meses). Portanto, para se obter a perda de rendimento por ano é necessário multiplicar o valor das perdas de rendimento em horas por 1.920 (Tabela 8).

Tabela 8
Perda Total de Rendimento por Hora e Perda de Rendimento por Ano em Dólares para a População Masculina do Brasil

Regiões/Setores	Perda Total de Rendimento por Hora	Perda de Rendimento por Ano em Dólares
Regiões		
Centro-Oeste	91.349	175.390.080
Sudeste	453.271	870.280.320
Sul	120.598	231.548.160
Nordeste	171.659	329.585.280
Norte	26.370	50.630.400
Total regiões	863.247	1.657.434.240
Setores		
Urbano	766.417	1.471.520.640
Rural	102.260	196.339.200
Total setores	868.677	1.667.859.840

A perda de rendimento por ano, para os homens do Brasil, considerando-se o somatório por região ou setor, atinge o valor aproximado de US\$ 1.660.000.000 (1,66 bilhões de dólares). Dado que a população de homens doentes é de 6.430.336, conclui-se que a perda de rendimento por homem adulto devido a problemas de saúde é de 258 dólares por ano.

Luft (1974), utilizando dados de 1966, obteve para cada homem americano, branco e preto, perdas anuais de, respectivamente, US\$ 2208,00 e US\$ 1716,00. De acordo com o autor, os rendimentos anuais dos homens adultos americanos sofreram redução de até 37% quando doentes. Ele também estimou em 23 bilhões de dólares a perda ocorrida em 1966 nos Estados Unidos devido a problemas de saúde dos trabalhadores. Haveman *et alii* (1995) estimou valores de perdas superiores ao de Luft nos anos 80, também nos Estados Unidos. Apesar de informações importantes, não é possível fazer qualquer comparação entre as pesquisas, pois existem diferenças de metodologia, dados, países, período de tempo, amostra utilizada etc..

6 Conclusões

Tomando-se como referência os pontos de corte propostos pela Organização Mundial da Saúde, a população brasileira adulta (18 a 65 anos) apresenta alta prevalência de baixo peso ($IMC < 20$), sobrepeso ($30 > IMC \geq 25$) e obesidade ($IMC \geq 30$), com uma taxa de 48,9% da população estimada. Somente 45,4% das mulheres e 57% dos homens são considerados normais ($20 \leq IMC < 25$).

Aproximadamente seis milhões de homens (15,4%) e sete milhões de mulheres (16,5%) têm baixo peso, atingindo cerca de 13 milhões de adultos brasileiros. A região Nordeste tem a maior porcentagem de adultos com baixo peso (21,7%), enquanto a região Sul apresenta a menor frequência, com aproximadamente 10% de sua população de 18 a 65 anos.

As perdas de rendimento dos homens que estão atuando no mercado de trabalho foram superiores às perdas dos homens que estão fora do mercado de trabalho. Além disso, as estimativas demonstraram que as regiões com maiores perdas foram a Sudeste e a Nordeste que, juntas, responderam por mais de 70% das perdas de rendimento totais. No que se refere aos setores, as estimativas indicaram que o setor urbano atingiu quase 90% das perdas de rendimentos totais.

O estudo mostrou ainda que as perdas de rendimento, devido às condições de saúde adversas ($IMC < 20$), são elevadas no Brasil, atingindo 1,66 bilhões de dólares. Dado que a população de homens doentes é de 6.430.336, a perda de rendimento por homem adulto devido a problemas de saúde é de 258 dólares por ano.

É certo que os serviços de saúde não possuem, por si só, capacidade de atenuar as diferenças regionais e setoriais existentes. Isto ocorrerá em consequência de um amplo processo de desenvolvimento econômico e da intensificação de políticas sociais, como aquelas direcionadas para as questões de saneamento básico e para a melhoria do estado nutricional da população.

Diversas empresas no Brasil estão cientes dos prejuízos acarretados por faltas e quedas de produtividade em decorrência dos problemas de saúde dos seus funcionários, e muitas já possuem programas preventivos, como o oferecimento de vacinas antigripais, campanhas antitabagismo e de prevenção de doenças cardiovasculares e sexualmente transmissíveis, além da realização de exames de prevenção.

O governo, entretanto, tem um papel fundamental na melhoria da qualidade de vida daqueles que, por problemas de saúde, estão fora do mercado de trabalho. É importante que

haja campanhas de prevenção de doenças e programas de suplementação alimentar, além de fácil acesso a serviços de saúde de boa qualidade, principalmente à população carente. Os retornos obtidos com uma população saudável, mais ativa e produtiva podem superar os gastos com programas de saúde e beneficiar a todos.

Referências bibliográficas

- Anjos, L. Índice de massa corporal como indicador do estado nutricional de adultos: revisão da literatura. *Revista de Saúde Pública*, v. 26, n. 6, p. 431-436, 1992.
- Barros, E. M.; Piola, S. F. & Vianna, S. M. *Política de saúde no Brasil: diagnóstico e perspectivas*. Brasília: IPEA, n. 401, 1996, 123p.
- Berndt, E. R. *The practice of econometrics: classic and contemporary*. New York: Addison Wesley, 1991
- Bliss, C. & Stern, P. Productivity, wages and nutrition. *Journal of Development Economics*, v. 5, p. 331-362, 1978.
- Bowen, W G. and Finagan, A. *The economics of labor force participation*. Princeton: Princeton University Press, 1969, 390p.
- Cooper, B. S. and Rice D. P. The economic cost of illness revised. *Social Security Bulletin*, v. 39, p. 21-36, 1976.
- Davis, J. M. Impact of health on earnings and labor market activity. *Monthly Labor Review*, p. 46-49, Oct. 1972.
- Folland, S.; Goodman, A. C. and Stano, M. *The economics of health and health care*. New York: Library of Congress, 1993, 726p.
- Haveman, R.; Wolfe, B & Buron, L. The loss of earnings capability disability/health limitations: toward a new social indicator. *Review of Income and Wealth*, v. 41, n. 3, 1995.
- Heckman, J. Shadow prices, market wages, and labor supply. *Econometrica*, v. 1, p. 679-694, 1974.
- _____ Sample selection bias as a specification error. In: Smith, J. P., *Female labor supply: theory and estimation*. Princeton University Press, 1980, p. 206-248.

- INAN/IBGE/IPEA. *Pesquisa nacional sobre saúde e nutrição (PNSN)*. Brasília, 1989.
- INAN. *Pesquisa nacional sobre saúde e nutrição: condições nutricionais da população brasileira*. Brasília, 1991.
- James, W. P. T., Ferro-Luzzi, A. & Waterlow, J. C. Definition of chronic energy in adults. *European Journal of Clinical Nutrition*, v. 42, p. 969-981, 1988.
- Jarret, R. J *et alii*. Weigh and mortality in the Whitehall study *British Medical Journal*, v. 285, p. 535-537, 1982.
- Kassouf, A L. The wage rate estimation using the Heckman procedure. *Revista de Econometria*, v. 14, n. 1. p. 89-107. 1994.
- Lew, E. A. and Garfinkel L. Variations in mortality by weight among 750,000 men and women. *Journal Chronical Disease*, v. 32, p. 563-576, 1979.
- Luft, H. S. The impact of poor health on earnings. *The Review of Economics and Estatistics*, 1974.
- Medice, A. C. *Perfil da saúde no Brasil*. Rio de Janeiro: IPEA, 1997, 189p.
- Mincer, J. On-the-job training: costs, returns, and some implications. *Journal of Political Economy*, v. 70, n. 5, p. 59-70, 1962.
- Monteiro, C. A. A dimensão da pobreza, da fome e da desnutrição no Brasil. *Estudos avançados*, v. 9, n. 24, p. 195-207. maio/ago. 1995.
- Morgan, J. N.; David, M. H., Cohen, W. J & Brazer, H. E. *Income and welfare in United States*. New York: McGraw-Hill, 1962.
- Parnes, H. S. & Meyer, J. *Withdrawal from the labor force by middle aged men, 1966-67*. Columbia, Ohio: Center for human resource research, Ohio State University, Jan, 1971.
- Rice, D., Hodgson, T. A. & Kopstein, A. N. The economic costs of illness: a replication and update. *Health Care Financing Review*, v. 6, n. 1, p. 61-80, 1985.
- Sorlie, P. Body build and mortality, The Framingham study. *JAMA*, v. 243, p.1828-1831, 1980.

Eficiência na produção agrícola paulista e seus determinantes^{*}

José R. Vicente[§]

RESUMO

O objetivo deste trabalho foi o de mensurar a eficiência na produção agrícola paulista, e a contribuição de seus fatores determinantes. Na análise empírica, empregaram-se funções de produção de fronteiras estocásticas com especificações Cobb-Douglas e Translog e dados seccionais dos anos agrícolas 1973/74 e 1988/89. Os resultados confirmaram o impacto positivo, sobre a eficiência na agricultura, de escolaridade, experiência e pesquisa agrônômica. O crédito rural e a política de garantia de preços mínimos aumentaram a eficiência na produção. Variáveis ambientais, como disponibilidade de água e qualidade das terras, também afetaram a eficiência. Pequenos imóveis rurais apresentavam maiores índices de eficiência geral, e o núcleo regional mais eficiente transferiu-se de regiões tradicionais para outras que empregavam mais intensamente tecnologias avançadas. Por último, as implicações políticas dessas descobertas empíricas foram discutidas.

Palavras-chave: função de produção de fronteira estocástica, capital humano, pesquisa agrícola.

ABSTRACT

The objective of this paper was to measure the efficiency in the agricultural production of São Paulo State, Brazil, as well as the factors that affected the efficiency indexes. Empirical analysis used stochastic frontier production functions - with both Cobb-Douglas and Translog production specifications and cross section data of the crop years 1973/74 and 1988/89. Results confirmed the positive impact on agricultural efficiency of schooling, experience, and agricultural research. Rural credit improved production efficiency, while guaranteed minimum prices for agricultural outputs boosted efficiency. Environmental variables, such as water availability and land quality, also influenced production efficiency. Small-sized farms tended to be more efficient overall, and over time the most efficient farms moved geographically from traditional agricultural areas to those deploying more advanced technologies. Finally, the policy implications of these empirical findings were briefly discussed.

Key words: stochastic frontier production function, human capital, agricultural research.

* Versão preliminar deste estudo foi apresentada no XXVI Encontro Nacional de Economia, Vitória, ES, 8 a 11 de dezembro de 1998. O autor agradece os comentários de Gabriel L. S. Peixoto da Silva, Cicely M. Amaral, Fernando B. Homem de Melo e Heron C. E. do Carmo, da FEA/USP, a uma versão anterior deste trabalho, que se baseia nos capítulos 4 e 5 de sua tese de doutorado. (Vicente, 1997)

§ Pesquisador Científico do Instituto de Economia Agrícola (IEA) e Bolsista do CNPq. E-mail: jrvicente@iea.sp.gov.br.

1 Introdução

O processo de modernização da agricultura brasileira, que se intensificou a partir da década de 70, resultou em considerável aumento da produção agrícola, da ordem de 60% na década de 70 e de cerca de 30% nos anos 80.¹ Essa modernização foi impulsionada por diversas políticas governamentais, como a de crédito rural, a de preços mínimos, e investimentos em pesquisa e extensão rural, que ocorreram predominantemente nos anos 70 e que seguiram uma estratégia delineada já a partir do Plano Estratégico de Desenvolvimento (PED) em 1968, e ratificada tanto no I Plano Nacional de Desenvolvimento (I PND, 1972-74) quanto no II PND (1975-79).

O processo de modernização, entretanto, não parece ainda haver sido estudado o suficiente para que se possa afirmar se ocorreram assimetrias importantes durante o mesmo, e se as condições em que se processa a produção tornaram-se mais ou menos homogêneas; análises empíricas, baseadas em modelos econométricos consistentes e em bases de dados adequadas, são raras. Por isso, também pouco se sabe a respeito dos efeitos de diversos fatores estruturais, conjunturais e ambientais, sobre a eficiência na produção agrícola. Os estudos que procuraram responder a essas questões foram geralmente efetuados com dados agregados, que impossibilitam análises sobre especificidades locais, de escala, e relacionadas à segmentação de mercados.

Com o processo de abertura da economia, implementado a partir dos anos 90, a agricultura brasileira passou a enfrentar concorrentes externos; no âmbito do MERCOSUL, por exemplo, o mercado brasileiro vem sendo alvo dos produtores argentinos, tradicionais exportadores de produtos agrícolas. Para enfrentar esse desafio é necessária eficiência em todas as fases da produção no complexo agroindustrial, a começar pelo processo produtivo dentro da propriedade. Portanto, cresce a importância de análises destinadas a identificar os determinantes da eficiência na agricultura.

Neste estudo pretendeu-se analisar a eficiência na produção agrícola paulista, buscando identificar diferenças entre regiões, distintos tamanhos de propriedades e grupos de produtos cultivados. Procurou-se, também, mensurar a influência de fatores estruturais, conjunturais e ambientais sobre as diferenças encontradas.

¹ Dados da década de 70 encontram-se em Vicente (1989); sobre os anos 80, ver Gasques & Verde (1990) e Silva (1991).

2 Metodologia

Os índices de eficiência foram obtidos por meio do ajuste de funções de produção de fronteiras estocásticas (Aigner, Lovell e Schmidt, 1977); esse método propugna, basicamente, ser possível mensurar a ineficiência das firmas individuais, que faz com que produzam em pontos abaixo da fronteira. Nos desenvolvimentos mais recentes, consideram-se os erros dos modelos compostos por duas partes: um componente simétrico (normal) representativo de efeitos aleatórios fora do controle das firmas, e um componente unilateral (não negativo), que captura os efeitos por elas controláveis (ineficiência).²

Formalmente (Greene, 1995):

$$y = \beta'x + v - u \quad (1)$$

onde y representa o nível de produção possível para a i -ésima firma, β é um vetor de parâmetros desconhecidos, x é a matriz de insumos - exógena (independente de u) -, v é um erro aleatório com média zero, associado aos fatores fora do controle da firma, u é uma variável aleatória, independente e identicamente distribuída, não negativa, relacionada a fatores específicos que contribuem para que a firma não atinja a máxima eficiência na produção.³

Como as pressuposições sobre a distribuição de u afetam os resultados, decidiu-se comparar os resultados das possibilidades permitidas pelo *software* utilizado: '**metade**' (*half*) *normal*, *normal truncada* e *exponencial*.⁴ (Greene, 1995)

Inicialmente, foram ajustadas funções do tipo Cobb-Douglas, que ainda é a mais freqüentemente empregada em estudos com a metodologia aqui utilizada; entretanto, são bem conhecidos os problemas restritivos a ela associados. Alternativamente, foi utilizada uma forma funcional flexível tipo Translog, cujos resultados foram cotejados com os provenientes da Cobb-Douglas. Não obstante, Maddala, citado por Saheli e Macedo (1998), argumenta que, como a produtividade advém da primeira derivada – que não é

2 A evolução dessa metodologia, desde a apresentação original de Farrell (1957), encontra-se em Vicente (1997); exemplos de aplicações em agricultura podem ser vistos em Battese (1992).

3 A escolha de uma determinada distribuição para u (ou e^u , que mede o grau de eficiência) condiciona os resultados das estimativas de máxima verossimilhança; entretanto, não existem razões definitivas *a priori* orientando tal decisão, somente as de conveniência estatística (Førsund, Lovell e Schmidt, 1980).

4 Detalhes sobre essas especificações, as técnicas de estimação e os cálculos dos valores correspondentes de u encontram-se em Greene (1995).

alterada pela forma funcional da função de produção – a questão da especificação não é relevante neste tipo de análise.

Os parâmetros foram estimados com dados de amostras probabilísticas, em nível de imóvel rural, de dois cortes seccionais: os anos agrícolas 1973/74⁵ e 1988/89,⁶ tendo como fonte o levantamento denominado objetivo, efetuado pelo Instituto de Economia Agrícola e pela Coordenadoria de Assistência Técnica Integral no Estado de São Paulo. O primeiro ponto - de meados dos anos 70 - representa o período conhecido como do 'milagre', com crescimento econômico acelerado e inflação relativamente controlada, enquanto que o segundo – do final dos anos 80 – retrata o auge da chamada 'década perdida' com estagnação econômica e beirando a hiperinflação; é de se esperar que fatores capazes de influenciar a eficiência econômica em situações tão distintas estejam, de fato, entre seus principais determinantes.

Como variável dependente foi utilizado o **Valor da Produção**⁷ e as variáveis explicativas incluídas nas funções foram representativas de **Capital** (Terra e Outras Despesas) e **Trabalho**.⁸

Após a estimação dos parâmetros - e de posse dos índices de eficiência - procurou-se explicar as diferenças de eficiência por meio de modelos de regressão, com variáveis independentes representativas de fatores estruturais, conjunturais e de condições ambientais.

Como variáveis explicativas associadas a **fatores estruturais** utilizou-se a área total do imóvel, para representar a disponibilidade de terra, a área cultivada e a proporção de área explorável efetivamente plantada com culturas anuais e perenes, buscando verificar influências da extensão e da intensidade da exploração sobre os índices de eficiência.

5 No ajuste dos modelos com dados de 1973/74 foram utilizados 5.215 observações (imóveis rurais), assim distribuídos pelas Divisões Regionais Agrícolas (DIRAs): 141 em São Paulo, 188 no Vale do Paraíba, 525 em Sorocaba, 579 em Campinas, 1.185 em Ribeirão Preto, 320 em Bauru, 885 em São José do Rio Preto, 308 em Araçatuba, 382 em Presidente Prudente e 702 em Marília.

6 Para o ano agrícola 1988/89, os modelos foram estimados com dados de 2.124 imóveis rurais, distribuídos pelas DIRAs da seguinte forma: 14 de São Paulo, 45 do Vale do Paraíba, 344 de Sorocaba, 263 de Campinas, 472 de Ribeirão Preto, 83 de Bauru, 336 de São José do Rio Preto, 116 de Araçatuba, 187 de Presidente Prudente e 264 de Marília.

7 Nesse caso, a variável aleatória u_i representa todos os tipos de ineficiência na produção, inclusive ineficiência técnica. Se a estrutura de preços for a mesma para todos os produtores, u_i medirá somente ineficiência técnica. Portanto, a rigor, empregando-se o valor da produção como variável dependente, u_i representará a ineficiência econômica da produção nos imóveis rurais. (Battese, Rambaldi e Wan, 1997)

8 Detalhes sobre a construção dessas variáveis encontram-se em Vicente (1997).

A existência de conhecimentos técnico-científicos foi representada pelo número de artigos científicos publicados referentes ao Estado de São Paulo e às culturas consideradas; essa medida foi utilizada, entre outros, por Evenson & Kislev (1973), Homem de Melo (1980), Silva (1984) e Vicente (1989). A quantidade de conhecimentos disponível para cada imóvel rural foi obtida ponderando-se o número de pesquisas acumulado, para cada cultura, pela área com ela cultivada no imóvel, no ano agrícola em análise.⁹

Ainda dentre os fatores estruturais foram considerados a educação e a experiência do produtor. A educação foi representada, em 1973/74, pelos anos de escolaridade do proprietário, de sua esposa, de seus filhos e do administrador do imóvel, e em 1988/89 pela escolaridade do proprietário.¹⁰ A experiência do produtor, disponível apenas para 1973/74, foi representada pelo número de anos de residência no imóvel rural atual.

No grupo de **fatores conjunturais**, procurou-se mensurar a eficácia das políticas de crédito rural e de preços mínimos. O crédito rural foi representado pelos valores totais dos financiamentos obtidos nos anos agrícolas 1972/73 e 1973/74, para as modalidades de custeio e de investimento, tomadas separadamente. Em 1988/89, como não constam informações sobre acesso a crédito nos questionários, foi necessário criar uma *proxy* denominada de probabilidade de acesso a crédito de custeio. Partiu-se do número de contratos de custeio por cultura, em 1988 e 1989, supondo-se que o percentual referente a São Paulo fosse proporcional à importância relativa das áreas das culturas no Estado. A probabilidade de acesso a crédito de custeio foi calculada pela razão do número estimado de contratos pelo número de imóveis produtores das culturas consideradas; a probabilidade em nível de imóvel foi obtida ponderando-se esses índices pelas áreas cultivadas com cada lavoura.

Como *proxy* para preços mínimos foi construído um índice de garantia, obtido pela razão preço mínimo/custo operacional de produção, para os anos agrícolas 1973/74 e 1988/89: isso foi feito para cada uma das culturas assistidas pelo programa à época: algodão, amendoim, arroz, feijão, mandioca, milho e soja. A cana-de-açúcar foi incluída entre esses produtos porque seu preço de venda era fixado pelo governo com grande antecedência, permitindo um conhecimento prévio desse componente da receita muito mais acurado do que o propiciado aos produtos contemplados pela política de preços mínimos,

9 As diferentes alternativas dessa variável que foram testadas, bem como as fontes do inventário de pesquisas utilizado, encontram-se em Vicente (1997).

10 Ressalte-se, entretanto, que os questionários não permitem correções para a qualidade ou a adequação da escolaridade. Os anos de estudo de um profissional de Ciências Agrárias que devem ter maiores efeitos sobre a eficiência na produção - pesam, portanto, o mesmo que a formação em qualquer outra área.

que nem sempre eram aderentes aos efetivamente recebidos pelos produtores. Um índice em nível de imóvel rural foi obtido ponderando-se o índice de garantia de cada cultura pela área com ela cultivada.

A conjuntura desfavorável aos produtos domésticos na década de 70¹¹ foi representada pela proporção de área ocupada com esses produtos em relação à área total com culturas. Espera-se que essa variável tenha captado os efeitos de outras não explicitamente consideradas, como, por exemplo, a relação entre os preços desses produtos e os de exportáveis e o risco econômico relativo. Nos modelos com dados do final da década de 80, a variável foi mantida.

Como a variável dependente foi o valor total da produção agrícola do imóvel rural, é razoável supor que condições favoráveis - ou desfavoráveis - de preços de determinados produtos estejam influenciando os resultados. Por esse motivo, para evitar o que poderia ser um importante viés de especificação, foi construída uma variável explicativa, obtida pelo quociente da receita líquida por hectare - deduzidos os custos operacionais de produção (Prognóstico, 1971-88) - dos diferentes produtos pela receita líquida, por hectare, obtida pelos produtores de milho, e ponderada pelas áreas cultivadas com cada produto nos imóveis rurais.

Os **fatores ambientais** foram representados por medidas usuais de clima e solo, características que, uma vez que determinam o rendimento potencial de cada técnica ou processo de produção, influenciam os rendimentos das culturas e os índices de eficiência.

Diversas variáveis climáticas vêm tendo sua influência sobre a produção agrícola enfatizada, como temperatura, precipitação pluviométrica, brilho solar, umidade do ar, geadas etc., com as duas primeiras sendo as mais frequentemente citadas como de maior importância. Todavia, a inclusão dessas duas variáveis diretamente em modelos nem sempre é uma boa alternativa, devido à conhecida interação entre ambas. Por esse motivo, foi efetuado o cálculo de uma variável derivada, a deficiência hídrica, que foi obtida por meio do cálculo de balanços hídricos sequenciais referentes ao período outubro-março.

Por intermédio de outra variável ambiental, procurou-se representar as condições do solo. Uma das *proxies* escolhida foi a aptidão agrícola das terras, mais especificamente, o percentual, em relação à área de cada DIRA, das terras aptas para lavouras, com aptidão

11 Essa distinção foi feita porque os preços e as decisões de plantio dos produtos domésticos dependiam somente da demanda local e das políticas governamentais, enquanto que os produtos exportáveis tinham também o mercado internacional como importante determinante da produção. (Barros, 1979) Embora essa classificação tenha perdido relevância com o processo de abertura de mercado levado a efeito nos anos 90, nas décadas de 70 e de 80 diversos fatores afetaram diferentemente esses dois subsetores, explicando seu crescimento distinto. (Homem de Melo, 1988)

boa e regular, e os níveis de manejo A (mais simples) ou, alternativamente, B(médio), conforme a classificação utilizada pelo Ministério da Agricultura. (Brasil, 1979) Como nessa classificação o nível máximo de desagregação possível é a DIRA, outra *proxy* foi tentada visando contornar esse problema, representando-se a qualidade das terras pelo valor do arrendamento ou, alternativamente, pelo preço médio das terras dos imóveis, itens que são levantados nos questionários. Essa medida, além dos quesitos relacionados à fertilidade dos solos, deve incorporar as facilidades de escoamento da produção, proximidade de mercados consumidores de produtos e distribuidores de insumos, de centros urbanos em expansão etc., estando mais próxima da noção ricardiana de produtividade diferencial das terras (Ricardo, 1982), e das análises dela derivadas, como o modelo de localização.¹²

Nove variáveis *dummies* representativas das regiões do Estado foram utilizadas para verificar especificidades locacionais e evitar possíveis vieses. Marília foi escolhida como DIRA base, em relação à qual eventuais diferenças foram medidas e, portanto, sem variável *dummy* específica.

3 Resultados e discussão

As diferentes especificações testadas apresentaram considerável variabilidade: para o ano agrícola 1973/74, as médias do índice de eficiência estiveram entre 0,562 e 0,703, e para 1988/89, entre 0,499 e 0,884. Como os modelos pouco diferiram segundo os critérios de avaliação utilizados, optou-se por analisar as medidas de eficiência com as menores amplitudes de variação, preferindo-se correr o risco de não apontar diferenças onde talvez existissem, do que indicar como distintas médias, na verdade, iguais. Tanto no ajuste da Cobb-Douglas como no da Translog, com dados de 1973/74, os modelos em que a forma de distribuição de u foi assumida como **metade (*half*) normal** foram os escolhidos. Já com dados de 1988/89, as menores amplitudes de variação ocorreram em modelos com distribuição de **u exponencial**.¹³

As discussões que se seguem são baseadas em resultados provenientes dos ajustes de funções Translog (Tabelas 1 e 2), já que as hipóteses de que as funções de produção poderiam ser do tipo Cobb-Douglas foram rejeitadas.¹⁴

12 Ver, por exemplo, Hayami & Ruttan (1988).

13 Os resultados de todos os modelos estão disponíveis em Vicente (1997).

14 Tal afirmação baseia-se em testes F, construídos com base nas somas de quadrados dos resíduos dos modelos completos (Translog) e restritos (Cobb-Douglas), e que levaram à rejeição das hipóteses iniciais.

Tabela 1
Principais Resultados do Ajuste de Função de Produção de Fronteira Estocástica,
Tipo Translog, Estado de São Paulo, Ano Agrícola 1973/74¹

Variável	Coeficiente	Desvio Padrão	T	Nível de Significância
Estimação por Mínimos Quadrados Ordinários				
Constante	-0,8760E-00	0,6877E-00	-1,274	0,2027
<i>ln</i> (Trabalho)	0,1315E-00	0,1471E-00	0,894	0,3715
<i>ln</i> (Terra)	0,9566E-00	0,8047E-01	11,888	0,0000
<i>ln</i> (Outras desp.)	-0,4233E-00	0,9249E-01	-4,577	0,0000
$\frac{1}{2} \ln^2$ (Trabalho)	0,4459E-01	0,1110E-01	4,016	0,0001
$\frac{1}{2} \ln^2$ (Terra)	-0,1098E-01	0,3553E-02	-3,090	0,0020
$\frac{1}{2} \ln^2$ (Outras Desp.)	-0,1144E-01	0,3715E-02	-3,079	0,0021
<i>ln</i> (Terra)x <i>ln</i> (Trab.)	-0,5664E-01	0,9510E-02	-5,955	0,0000
<i>ln</i> (Ter.)x <i>ln</i> (O. Desp.)	0,5772E-01	0,6392E-02	9,031	0,0000
<i>ln</i> (Tr.)x <i>ln</i> (O. Des.)	0,1336E-02	0,1154E-01	0,116	0,9078
R ²	0,832			
F _(9,5.205)	2.862			0,0000
Estimação por Máxima Verossimilhança ²				
Constante	-0,2736E-00	0,6725E-00	-0,407	0,6841
<i>ln</i> (Trabalho)	0,9701E-01	0,1388E-00	0,699	0,4845
<i>ln</i> (Terra)	0,9434E-00	0,6831E-01	13,809	0,0000
<i>ln</i> (Outras Desp.)	-0,3844E-00	0,7278E-01	-5,281	0,0000
$\frac{1}{2} \ln^2$ (Trabalho)	0,4627E-01	0,1019E-01	4,541	0,0000
$\frac{1}{2} \ln^2$ (Terra)	-0,9285E-02	0,2537E-02	-3,660	0,0003
$\frac{1}{2} \ln^2$ (Out. Desp.)	-0,8548E-02	0,2331E-02	-3,666	0,0003
<i>ln</i> (Terra)x <i>ln</i> (Trab.)	-0,5567E-01	0,8046E-02	-6,919	0,0000
<i>ln</i> (Terra)x(O. Des.)	0,5194E-01	0,4917E-02	10,563	0,0000
<i>ln</i> (Trab.) x (O. Des.)	0,3462E-03	0,9229E-02	0,038	0,9701
σ_U/σ_V	0,9574E-00	0,8932E-01	10,719	0,0000
$\sqrt{\sigma_V^2 + \sigma_U^2}$	0,8931E-00	0,2103E-01	42,467	0,0000

(1) A variável dependente é o logaritmo do valor da produção.

(2) Assumindo-se distribuição metade (*half*) normal para *U*.

Fonte: Resultados da pesquisa, a partir de dados básicos do IEA.

Tabela 2
Principais Resultados do Ajuste de Função de Produção de Fronteira Estocástica,
tipo Translog, Estado de São Paulo, Ano Agrícola 1988/89¹

Variável	Coeficiente	Desvio Padrão	t	Nível de Significância
Estimação por Mínimos Quadrados Ordinários				
Constante	3,9319	1,4550E-00	2,703	0,0069
\ln (Trabalho)	-0,5536	0,3328E-00	-1,663	0,0962
\ln (Capital)	0,7602	0,2305E-00	3,298	0,0010
$\frac{1}{2} \ln^2$ (Trabalho)	-0,1254	0,5795E-01	-2,163	0,0305
$\frac{1}{2} \ln^2$ (Capital)	-0,1815	0,3904E-01	-4,649	0,0000
\ln (Trab.) x \ln (Cap.)	0,2018	0,4260E-01	4,737	0,0000
R ²	0,672			
F _(5,2.118)	869			0,0000
Estimação por Máxima Verossimilhança ²				
Constante	4,0933	1,4500E-00	2,824	0,0048
\ln (Trabalho)	-0,8507	0,3046E-00	-2,793	0,0052
\ln (Capital)	1,1976	0,2329E-00	5,143	0,0000
$\frac{1}{2} \ln^2$ (Trabalho)	-0,0200	0,5196E-01	-0,388	0,6980
$\frac{1}{2} \ln^2$ (Capital)	-0,1405	0,3660E-01	-3,838	0,0001
\ln (Trab.) x \ln (Cap.)	0,1209	0,3915E-01	3,088	0,0020
θ	1,2746	0,9141E-01	13,944	0,0000
σ_v	1,0504	0,3441E-01	30,527	0,0000

1 A variável dependente é o logaritmo do valor da produção.

2 Assumindo-se distribuição exponencial para U .

Fonte: Resultados da pesquisa, a partir de dados básicos do IEA.

3.1 Diferenças de eficiência na produção agrícola paulista

De posse dos índices de eficiência, procurou-se inicialmente verificar se os grupos de produtos domésticos e exportáveis diferiam significativamente: as médias desses dois grupos foram iguais, respectivamente, a 0,599 e 0,648 em 1973/74, e a 0,580 e 0,676 em

1988/89. Submetidas ao teste não-paramétrico de Wilcoxon (Campos, 1976), verificou-se que diferem, nos dois pontos do tempo, a 1% de probabilidade ($W^*=15,62$ para 1973/74 e $W^*=22,06$ para 1988/89, com $z_{0,01}=2,33$). Portanto, pode-se concluir que a eficiência na produção agrícola paulista era maior entre produtores de exportáveis do que no grupo que cultivava predominantemente produtos de mercado interno.

Em seguida procurou-se verificar a ocorrência de diferenças entre distintos tamanhos de imóveis rurais: mini (até 20 ha), pequenos (de 20,1 a 100,0 ha), médios (de 100,1 a 500,0 ha) e grandes (mais de 500 ha). Já que, nesse caso, existem mais de duas médias a serem comparadas, foi utilizado o teste não-paramétrico de Kruskal-Wallis, complementado pelo procedimento não-paramétrico das Comparações Múltiplas. (Campos, 1976)

Os imóveis pequenos apareceram como os mais eficientes, nos dois cortes seccionais analisados. Em 1973/74, diferiam significativamente dos demais tamanhos de imóveis, e em 1988/89 eram significativamente mais eficientes do que os grandes (Tabela 3).

Tabela 3
Teste de Kruskal-Wallis e Comparações Múltiplas para Diferenças de Eficiência entre Estratos de Tamanho de Imóvel Rural, Estado de São Paulo

Teste de Kruskal-Wallis					
Ano Agrícola 1973/74			Ano Agrícola 1988/89		
H = 125,99 ¹			H = 26,63 ¹		
Comparações Múltiplas					
Ano Agrícola 1973/74			Ano Agrícola 1988/89		
Tamanho	Média das Ordens ²		Tamanho	Média das Ordens ²	
Pequeno	2.938,99	a	Pequeno	1.153,91	a
Grande	2.526,37	b	Mini	1.072,46	a,b
Médio	2.418,40	b	Médio	1.069,19	a,b
Mini	2.409,10	b	Grande	969,34	b

1 Valor crítico de χ^2 com 3 graus de liberdade, a 1%: 11,34.

2 As médias assinaladas pela mesma letra não diferem ao nível de 5%.

Fonte: Resultados da Pesquisa, a partir de dados básicos do IEA.

Por último, procurou-se examinar a ocorrência de diferenças significativas de eficiência entre as regiões do Estado (DIRAs). Em 1973/74, a DIRA de Sorocaba apareceu como a mais eficiente, não diferindo apenas do Vale do Paraíba, a segunda colocada, que, por sua vez, não diferia de Bauru (Tabela 4). Sorocaba era uma região de agricultura diversificada

e de crescente importância no Estado, embora abrigando uma das regiões de maior miséria em São Paulo, conhecida como o ramal da fome; na década de 70, a região passou a concentrar mais de 2/3 da produção estadual de feijão em lavouras, basicamente, solteiras. Já o Vale do Paraíba, aparecendo no topo da tabela, parece confirmar uma vez mais as colocações de Schultz (1965) sobre a possibilidade da agricultura tradicional ser eficiente, dados os recursos disponíveis. Bauru, em terceiro lugar, seria menos eficiente do que Sorocaba, não diferente do Vale do Paraíba, Ribeirão Preto, Marília e Campinas, e superior às demais. Ribeirão Preto, a quarta colocada, era superada por Sorocaba e Vale do Paraíba, igual a Bauru, Marília e Campinas, e mais eficiente do que as outras regiões. Tratando-se da região mais dinâmica e tecnificada do Estado, aparentemente o fato de não ser a mais eficiente também é consistente com as colocações de Schultz (1975) sobre o desequilíbrio temporário criado por novas técnicas.

Tabela 4
Teste de Kruskal-Wallis e Comparações Múltiplas para Diferenças de Eficiência
entre as Divisões Regionais Agrícolas, Estado de São Paulo

Teste de Kruskal-Wallis					
Ano Agrícola 1973/74			Ano Agrícola 1988/89		
H = 225,78 ¹			H = 100,04 ¹		
Comparações Múltiplas					
Ano Agrícola 1973/74			Ano Agrícola 1988/89		
DIRA	Média das Ordens ⁽²⁾		DIRA	Média das Ordens ⁽²⁾	
Sorocaba	3.266,23	a	Campinas	1.220,51	a
V. Paraíba	3.133,12	a,b	Rib. Preto	1.208,14	a
Bauru	2.766,08	B,c	Bauru	1.042,99	a,b
Rib. Preto	2.737,22	c	S.J.R.Preto	1.024,56	b
Marília	2.554,17	c,d	Sorocaba	1.017,60	b
Campinas	2.462,53	c,d	Marília	985,04	b,c
S.J.R.Preto	2.414,85	d	P.Prudente	962,36	b,c
P. Prudente	2.265,82	d,e	Araçatuba	947,71	b,c
São Paulo	2.161,94	d,e	V. Paraíba	597,73	c
Araçatuba	2.083,86	e	São Paulo	436,38	c

1 Valor crítico de χ^2 com 9 graus de liberdade, a 1%: 21,67.

2 As médias assinaladas pela mesma letra não diferem ao nível de 5%.

Fonte: Resultados da Pesquisa, a partir de dados básicos do IEA.

Para 1988/89, mudanças relativas foram observadas em nível das DIRAs, provavelmente como consequência da consolidação do processo de modernização acelerada ocorrido durante as duas décadas: o núcleo de regiões de agricultura mais

tecnificada do Estado - Campinas e Ribeirão Preto - aparece junto a Bauru com os índices de eficiência mais elevados. Já com relação às regiões de menor eficiência relativa, percebe-se as duas DIRAs de agricultura mais tradicional, São Paulo e Vale do Paraíba, apresentam os piores índices (Tabela 4).

3.2 Determinantes da eficiência na produção agrícola

Para a análise dos determinantes da eficiência foi necessário, inicialmente, escolher uma técnica para estimar os parâmetros, uma vez que o método de mínimos quadrados ordinários, aplicado a amostras estratificadas não equiprobabilísticas, pode levar a resultados enviesados. Estimadores de máxima verossimilhança, como o desenvolvido por Holt, Smith e Winter (1980), geralmente são de difícil obtenção a partir de resultados de “pacotes” estatísticos e econométricos; por isso, optou-se por seguir a segunda sugestão desses autores, que é o uso dos fatores de expansão da amostra como pesos, em um esquema de mínimos quadrados ponderados. Nas simulações efetuadas por Holt, Smith e Winter (1980) esse estimador funcionou bem, sendo não viesado, apesar de menos estável e de apresentar erro padrão maior do que o de máxima verossimilhança.

A existência de heterocedasticidade, com dados de 1973/74, foi descartada pelo teste de Pesaran & Pesaran (Matos, 1997), já que o valor do F obtido é não significativo a 5% ($F = 3,12$).

Com dados desse primeiro corte seccional, do grupo de variáveis conjunturais, a proporção de área cultivada com produtos domésticos teve sinal negativo e significativo, provavelmente em decorrência das condições desfavoráveis a esse grupo de produtos, à época.¹⁵ O montante recebido de crédito rural para custeio aparentemente não afetava significativamente a eficiência em nível de imóvel, enquanto que o valor dos financiamentos para investimento teve coeficiente negativo e significativo. Esse resultado, à primeira vista surpreendente, deve-se provavelmente à forma de medir o total de outras despesas na função de produção: já que os serviços de máquinas, instalações, benfeitorias e culturas perenes tiveram como base o valor declarado dos mesmos, certamente os imóveis que tivessem efetuado investimentos recentes entraram nessa parcela com maiores gastos (Tabela 5).

A medida empregada para representar a produção também deve ter contribuído para o sinal positivo e significativo encontrado para o índice de cobertura de custos pelos preços

15 Detalhes em Homem de Melo (1988).

mínimos; caso esses preços tenham influenciado os efetivamente recebidos, afetariam também o valor total da produção. Essa colocação é reforçada pelo efeito positivo e significativo detectado para o relativo receita líquida dos produtos/receita líquida do milho, sobre o índice de eficiência (Tabela 5).

Tabela 5

Principais Resultados do Ajuste de Modelos Tendo como Variável Dependente a Eficiência, Proveniente de Funções de Produção Tipo Translog, Estado de São Paulo

Grupos de Fatores / Variável	Estimação por Mínimos Quadrados Ponderados			
	Ano Agrícola 1973/74		Ano Agrícola 1988/89	
	Coeficiente	Nível Sig.	Coeficiente	Nível Sig.
Estruturais				
Escolaridade dos Dirigentes	0,10487E-03	0,084	0,71132E-03	0,090
Experiência do Produtor no Imóvel	0,25340E-03	0,003
Disponibilidade de Pesquisa	0,25425E-03	0,000	0,52882E-03	0,000
Área Total do Imóvel	-0,16831E-05	0,877	-0,17475E-05	0,884
Área Cultivada com Lavouras	0,88063E-04	0,015	0,27552E-04	0,180
Proporção da Área Cultivada	0,59275E-01	0,000	0,63077E-01	0,000
Conjunturais				
Crédito de Custeio	0,10847E-07	0,601	0,39565E-02	0,000
Crédito de Investimento	-0,90311E-07	0,002
Preços Mínimos	0,32968E-01	0,000	0,12978E-01	0,090
Proporção de Produt. Domésticos	-0,75101E-01	0,000	0,34790E-01	0,118
Receita Líquida Relativa	0,96104E-02	0,000	0,59360E-02	0,000
Ambientais				
Valor da Terra	-0,44528E-06	0,000	-0,33800E-07	0,000
Aptidão Agrícola das Terras	0,63637E-03	0,001	0,20742E-02	0,000
Deficiência Hídrica	-0,82691E-03	0,000	-0,31908E-02	0,000
Locacionais				
São Paulo	0,31624E-01	0,017	0,11548E-00	0,000
Vale do Paraíba	0,90563E-00	0,000	-0,42472E-01	0,093
Sorocaba	0,72275E-01	0,000	-0,57615E-02	0,626
Campinas	0,14105E-01	0,020	-0,44981E-01	0,005
Ribeirão Preto	0,22448E-01	0,000	-0,93822E-01	0,000
Bauru	0,63162E-01	0,000	0,18706E-01	0,239
Presidente Prudente	-0,47879E-02	0,418	-0,11819E-00	0,000
Constante	0,55054E-00	0,000	0,40803E-00	0,000
Indicadores da Qualidade do Ajuste				
R ²	0,440		0,396	
F	191,86 ^{a(1)}		54,85 ^{a(1)}	

1 Nível de significância: $\alpha = 1\%$.

Fonte: Resultados da pesquisa, a partir de dados básicos do IEA.

Entre os fatores ambientais, o valor das terras do imóvel apresentou sinal negativo, provavelmente também associado à medida utilizada para o uso desse fator nas funções estimadas, que era ponderado pelo valor do arrendamento. Nesse caso, outras variáveis - que não a fertilidade - formadoras do preço devem ter influenciado, e optou-se por inserir nos modelos a outra medida de fertilidade disponível, o percentual de terras aptas para lavouras nas DIRAs; essa variável apresentou o sinal positivo e significativo esperado, provavelmente representando melhor o efeito de solos mais férteis sobre a eficiência na produção. Por outro lado, a deficiência hídrica do período outubro-março apresentou coeficiente negativo e significativo em todas as especificações, mostrando a susceptibilidade da agricultura às condições do tempo (Tabela 5); esse resultado indica que se todo o Estado houvesse sido submetido ao mesmo *stress* hídrico a que esteve sujeita a DIRA de Araçatuba (46 mm, nível que não é incomum de ser atingido em anos de *déficits* hídricos elevados) a média do índice de eficiência de 1973/74 cairia mais de 3%.

Ainda para 1973/74, entre as variáveis representativas de fatores estruturais ligadas à escala e à intensidade de uso da terra a área total do imóvel não apareceu influenciando significativamente a eficiência na produção no ano agrícola 1973/74, enquanto que a área cultivada com lavouras anuais e perenes e a proporção de área do imóvel ocupada com lavouras tiveram coeficientes positivos e significativos (Tabela 5).

Tanto a educação como a experiência tiveram efeitos positivos sobre o índice de eficiência de 1973/74.

A disponibilidade de pesquisa científica apresentou, em todas as alternativas testadas, influência positiva e significativa sobre a eficiência, corroborando a pressuposição que embasou sua inclusão no modelo¹⁶ (Tabela 5); a cada 10 novos artigos publicados, poder-se-ia esperar aumentos de quase 0,5% no índice de eficiência de 1973/74, de acordo com o cálculo da elasticidade no ponto médio.

As variáveis *dummies* apresentaram sinais positivos para todas as DIRAs, exceto Presidente Prudente, que não diferiu significativamente de Marília, tomada como base.¹⁷

16 Além de definir a própria fronteira, considerou-se o conhecimento científico capaz de determinar o nível de eficiência de um produtor individual, uma vez que, quanto maior o conhecimento sobre as práticas agrônômicas ligadas a determinada cultura, maior a probabilidade de serem contemplados aspectos particulares, tais como condições de clima, solo, manejo de pragas e doenças, tratamentos culturais etc.

17 Para contornar problemas de multicolinearidade, as *dummies* de São José do Rio Preto e Araçatuba, que nunca apresentaram coeficientes significativos, foram excluídas dos modelos.

A análise dos resultados do modelo com dados de 1988/89¹⁸ mostra que pouco diferem dos obtidos para 1973/74 com respeito às influências significativas das variáveis independentes, com exceção da área cultivada com lavouras e da proporção de área dedicada a produtos domésticos, que deixaram de ser significativos, e do crédito de custeio, que passou a ser (Tabela 5).

Reforçando afirmações sobre a maior susceptibilidade da agricultura moderna às adversidades climáticas, a elasticidade estimada a partir do coeficiente obtido para a deficiência hídrica do período outubro-março mostra que se todas as regiões tivessem sido submetidas ao maior nível de *déficit* de água observado no ano (45 mm) o índice de eficiência de 1988/89 tenderia a cair mais de 13%.

As *dummies* representativas das DIRAs apresentaram sinais positivos para São Paulo e não significativos para Sorocaba e Bauru; as demais tiveram coeficientes negativos.

Os retornos provenientes da pesquisa científica aparentavam estar aumentando, uma vez que a cada 10 artigos adicionais publicados poder-se-ia esperar, de acordo com a elasticidade no ponto médio, aumento de quase 1% na eficiência média (mais precisamente, 0,854%), em 1988/89.

Os efeitos substanciais constados e a influência crescente da pesquisa agrícola brasileira sobre a eficiência na produção levam a refutar a posição de que é possível simplesmente importar o conhecimento necessário nessa área. Essa visão, próxima à de antigos modelos de difusão, postula que quaisquer problemas presentes e futuros já encontraram solução, propiciada pela inteligência externa.¹⁹

Romer (1993) argumenta que utilizar idéias desenvolvidas externamente pode ser uma escolha apropriada para países nos primeiros estágios do desenvolvimento (como o exemplo, bem-sucedido, das Ilhas Maurício). Entretanto, reconhece que as idéias, embora bens econômicos, não são bens privados convencionais, e que os mercados são inerentemente menos eficientes na produção e transmissão de idéias do que na de outros bens. Com a integração mundial, a qualidade das instituições públicas contribuirá para

18 A existência de heterocedasticidade foi descartada pelo teste de Pesaran & Pesaran ($F=3.69$, não significativo a 5%).

19 Por outro lado, pesquisadores ligados ao setor rural em geral reconhecem que um dos papéis importantes da política agrícola é o de promover a pesquisa, em especial a pública (ver, por exemplo, Alves, 1996 e Portugal & Contini, 1997).

diferenciar as áreas geográficas, e as melhor sucedidas serão aquelas com os mecanismos mais competentes para atender aos interesses coletivos, especialmente na produção de novas idéias.

Muitos dos modelos recentes de abertura tecnológica do comércio internacional defendem a posição de que, sob livre comércio, é vantajoso para os países menos desenvolvidos “emprestar” ou copiar novas tecnologias.²⁰ Para David (1993), nessa concepção caberia aos países desenvolvidos do hemisfério Norte a geração de conhecimento; ficariam encarregados das pesquisas básicas e aplicadas, das novas invenções e inovações, enquanto que os países em desenvolvimento ocupar-se-iam com imitações e difusão de tecnologia. Bell & Pavitt (1993) consideram que o nível de conhecimento tecnológico pode tornar-se uma fonte de vantagens comparativas, refletindo-se em técnicas de produção aprimoradas, nos projetos de bens de capital, e na capacidade de engenharia reversa e de reproduzir pesquisas desenvolvidas externamente.

Parece, portanto, que a idéia de utilizar tecnologia externa implica assumir como adequada ao País uma posição caudatária no desenvolvimento internacional.

A partir dos 80 - num processo que se agravou mais para o final da década - os recursos para pesquisa nos países em desenvolvimento diminuíram acentuadamente, levando a uma involução dos sistemas oficiais de pesquisa e comprometendo a eficácia de sua atuação.²¹ As perspectivas são de uma deterioração, em breve, da situação atual concernente à produção agrícola, em especial a de alimentos nos países em desenvolvimento, tendo como consequência um aumento de preços e de carências alimentares.²² As pesquisas que vêm sendo desenvolvidas no campo da biotecnologia - quase que exclusivamente em países avançados - não têm se preocupado com os problemas da agricultura dos países em desenvolvimento (Hazell, 1995); ao contrário, são numerosos os exemplos de substâncias sintéticas que buscam substituir seus produtos agrícolas típicos.

20 Como, por exemplo, Grossman & Helpman (1990), citados por David (1993).

21 No caso brasileiro, o orçamento da EMBRAPA cresceu entre 1973 e 1982, quando passou a apresentar tendência declinante. Em 1984, o montante de gastos com pesquisa agropecuária foi 37% menor do que o de 1982, e em 1990, 28% menor, também comparado a 1982. No Estado de São Paulo, a evolução desses investimentos aparenta ter sido ainda pior: considerando-se três dos mais importantes institutos de pesquisa - Instituto Agrônomo, Instituto Biológico e Instituto de Zootecnia - o máximo de recursos foi atingido em 1978. Em 1982, os gastos com pesquisa atingiram apenas 45% daquele máximo, e em 1990, 63%. (Mesquita, 1994)

22 Esses argumentos estão desenvolvidos em Dresrüsse (1995), Hazell (1995), Pardey & Alston (1995) e em Rosegrant (1995).

A postura de diminuição de subsídios e de serviços oferecidos à agricultura nos países em desenvolvimento - que vem fazendo dos sistemas oficiais de pesquisa uma de suas maiores vítimas - carece, ao menos nesse caso, de fundamentação econômica. Mesmo que se ignorem os resultados de estudos sobre os retornos dos investimentos em pesquisas desenvolvidas no Brasil, segundo análises do IFPRI esses retornos são quase sempre superiores a 20% ao ano, existindo casos de algumas taxas de 100% a.a. (Hazell, 1995; Pardey & Alston, 1995)

4 Conclusões e considerações finais

Os imóveis rurais pequenos (20,1 a 100 ha) apresentavam os maiores índices de eficiência, enquanto que os mini (até 20 ha) e grandes (mais de 500 ha) eram os menos eficientes, em 1973/74 e 1988/89, respectivamente. Portanto, se por um lado imóveis rurais pequenos podem produzir eficientemente - e de fato isso foi observado empiricamente -, por outro lado, assentar produtores em módulos com área total inferior a 20 hectares aparentemente significa condená-los à dependência de políticas especiais de amparo.

Entre as DIRAs, o núcleo mais eficiente era formado, em 1973/74, por Sorocaba - uma região em crescimento - e pelo Vale do Paraíba, de agricultura mais tradicional. Já em 1988/89, a eficiência foi maior nas DIRAs de agricultura mais dinâmica e modernizada: Campinas e Ribeirão Preto. Tanto em 1973/74 como em 1988/89 os imóveis que cultivavam predominantemente produtos exportáveis tinham níveis de eficiência superiores aos dos imóveis dedicados aos produtos domésticos.

Das variáveis estruturais escolhidas para representar o capital humano, tanto a escolaridade - de proprietários, seus familiares e administradores em 1973/74, e somente de proprietários em 1988/89 - como a experiência do produtor - disponível apenas para 1973/74, e medida pelo total de anos dedicados à agricultura - apareceram com influência positiva e significativa sobre a eficiência na produção. A disponibilidade de pesquisa científica condicionou os níveis de eficiência mensurados tanto em 1973/74 como em 1988/89

Entre as variáveis representativas da escala do empreendimento e da intensidade de uso da terra a área total do imóvel nunca teve coeficientes significativos, o montante de área do imóvel efetivamente cultivada com lavouras apresentou coeficientes positivos em 1973/74, e não significativos em 1988/89 e, para a proporção de área plantada com culturas anuais e perenes, os parâmetros estimados sempre foram positivos e significativos, aparentemente indicando que os índices de eficiência estariam mais relacionados a uma

certa especialização agrícola do que à disponibilidade de terra.

No grupo de fatores conjunturais, é provável que o crédito rural mais escasso e caro do final da década de 80, que apresentou coeficiente positivo, tenha sido empregado mais eficientemente do que o de meados da década de 70, não significativo sobre os índices de eficiência. O parâmetro negativo estimado para o crédito de investimento, em 1973/74, deve-se provavelmente à forma de medir as despesas. Já o motivo dos sinais positivos e significativos encontrados para o índice de cobertura de custos pelos preços mínimos e para o índice de receita líquida relativa deve ser a forma de medir a produção, por seu valor agregado. A proporção de área de lavouras cultivada com produtos domésticos, que era uma desvantagem em 1973/74, deixou de influenciar a eficiência em 1988/89.

Entre os fatores ambientais analisados, a deficiência hídrica sempre compareceu com coeficientes significativos e negativos, reafirmando a susceptibilidade da produção agrícola a esse fenômeno. A variável valor das terras do imóvel - com sinal negativo nos dois cortes de tempo - provavelmente captou efeitos de outros de seus componentes, que não a fertilidade; já os parâmetros estimados para o porcentual de terras aptas para lavouras sempre foram positivos e significativos, indicando que zoneamentos agrícolas precisos podem contribuir para a elevação da eficiência na agricultura.

Os resultados obtidos relacionados a algumas das variáveis explicativas da eficiência na produção merecem comentários adicionais. Em meados dos 90, os limites dos instrumentos clássicos de política agrícola, crédito rural e preços mínimos de garantia parecem ser claramente outros.

Especula-se que a política de crédito rural deverá caminhar para um esquema de redescontos, em que os bancos tenham acesso a recursos a juros inferiores aos de mercado, em montante igual a certa proporção dos empréstimos efetuados por suas carteiras rurais. Essa proporção, variável de acordo com a "qualidade" dos empréstimos efetuados, permitiria ao governo determinar modalidades, regiões, explorações e grupos de produtores prioritários. Considerando-se os resultados deste estudo, em que o crédito rural apareceu como importante determinante da eficiência na produção, e considerando-se também que as taxas de juros internas pagas pelos agentes econômicos estão atualmente entre as maiores do planeta, é imperioso que certos segmentos produtivos sejam protegidos, como os pequenos e - principalmente - os mini produtores.

A política de preços de garantia, por sua vez, parece ter como limites numa economia aberta, níveis próximos aos observados no mercado internacional, sob pena de acúmulo indesejável de estoques; entretanto, sobretudo para produtos de mercado interno, pode ser

conveniente manter esquemas plurianuais e garantia de cobertura ao menos próxima dos custos variáveis.²³

Esses dois instrumentos de política agrícola, se conjugados a sólidas bases técnicas, como zoneamentos agrícolas bem elaborados e atualizados, provavelmente contribuirão decisivamente para o aumento da eficiência, em parte dirigindo o quê, onde, como e quando plantar. Tal encadeamento permitiria também contornar os efeitos de um dos condicionantes da eficiência detectado nos modelos estimados, as condições do tempo.

Entretanto, apenas com essas medidas e o mercado como regulador, é provável que vejamos nos próximos anos consideráveis áreas de agricultura intensiva retrocederem, tornando-se novamente regiões de pecuária extensiva, de reflorestamentos, ou serem simplesmente abandonadas. É quase certo que esse processo não poderá ser de todo estancado, mas os limites da agricultura intensiva, eficiente e competitiva, podem - e devem - ser ampliados.

Essa ampliação de limites certamente passa por medidas que sequer foram tocadas neste trabalho: produção e distribuição eficientes de insumos modernos, infra-estrutura de transporte e armazenamento, reforma agrária e diminuição da defasagem cambial, entre outros. Todavia, passa também pelos fatores estruturais aqui analisados, como educação e conhecimento tecnológico.

A importância da educação para a produtividade agrícola vem sendo destacada há décadas por diversos autores em diferentes localidades, inclusive no Brasil; é interessante destacar que os resultados mostraram ser esse um fator limitante à capacidade de produzir com eficiência, mesmo no setor primário. Atualmente - é verdade que mais em função dos interesses na competitividade industrial do que na agrícola -, autoridades federais parecem decididamente empenhadas em elevar a média de escolaridade da população, e se esse esforço atingir o meio rural provavelmente a agricultura brasileira elevará seus padrões de eficiência produtiva.

Quanto ao outro fator, o conhecimento tecnológico,²⁴ aparentemente existe maior controvérsia; por isso, é necessário discutir as perspectivas do esforço de pesquisa nacional, mais especificamente o dirigido à agricultura. Existe ainda um papel a ser

23 Essa posição é melhor defendida por Homem de Melo (1991).

24 Embora discutidos separadamente, é importante frisar que investimentos em capital humano (educação) e em pesquisa são complementares; ver Redding (1996).

desempenhado pelo sistema estatal de pesquisas? Quais os desafios futuros para a pesquisa agrícola? Poderão parcerias com o setor privado viabilizar os recursos necessários ao sistema estatal de pesquisas? Suas prioridades serão as mesmas da sociedade como um todo?

Sobre os desafios futuros à pesquisa agrícola, pode-se argumentar com base no pensamento de Hazell (1995), que procurou resumir em três principais. O primeiro seria o de manter os ganhos de rendimento obtidos e estabelecer padrões ambientalmente mais adequados. Vencer a capacidade inata de doenças e pragas suplantarem a resistência das plantas e tornarem-se imunes a defensivos agrícolas é um trabalho constante e interminável; ao mesmo tempo, o desenvolvimento de alternativas biológicas aos insumos químicos atualmente utilizados é urgente e necessário.

O segundo desafio seria aumentar ainda mais os níveis de produtividade em áreas de agricultura intensiva, que empregam maciçamente insumos modernos; tal efeito só poderá ser conseguido com novas variedades, insumos de melhor qualidade e tratamentos culturais aprimorados.

O último - e talvez mais importante - é o de elevar a produtividade em sistemas de produção que empregam menos tecnologia, muitas vezes caracterizados por solos frágeis e pouco férteis, altamente susceptíveis a secas e pobres em infra-estrutura, fatores que dificultam a implementação de uma agricultura mais moderna, baseada no uso intensivo de insumos. Nesse caso, além de novas variedades mais resistentes a secas, a doenças e a pragas, a pesquisa agrícola deverá ocupar-se também de técnicas de combate à erosão, e de aumentar a eficiência de sistemas complexos, compostos por culturas consorciadas, intercaladas, e muitas vezes coexistindo com explorações animais.

Para atender a essas necessidades, especula-se que o caminho passa por uma crescente importância da pesquisa privada, ou via parcerias entre o sistema oficial de pesquisas e a iniciativa privada.

Em relação à capacidade do esforço de pesquisa privado atender a essas demandas, a literatura mais atual parece pouco animadora. Embora a privatização reduza os custos governamentais, o tipo de pesquisa desenvolvido em tais ambientes passa a ser somente o que permite ganhos substanciais e que podem ser protegidos, esquecendo-se completamente a transferência de conhecimentos aos agricultores, principalmente aos de menores condições financeiras e com dificuldade de gerenciamento de recursos. (Oram, 1995) O setor privado tem pouco ou nenhum incentivo em pesquisar problemas de pequenos

agricultores, em geral produtores de alimentos básicos destinados ao mercado interno, e tampouco em estudar aspectos relativos à degradação ambiental. (Hazell, 1995)

Sabe-se que as empresas multinacionais, em especial, concentram quase todo seu esforço de pesquisa nos países-sede²⁵ e, segundo Dresdüsse (1995), protegem vigorosamente suas descobertas. Pesquisa privada em países de menores níveis de renda é praticamente inexistente, e mesmo onde existe, o setor privado é capaz de responder por apenas uma pequena parte das necessidades; a maioria das pesquisas, com o maior potencial de benefícios para a sociedade, é melhor conduzida pelo setor público, já que as companhias privadas não podem se apoderar dos resultados o suficiente para garantir os investimentos necessários. (Pinsuup-Andersen, 1995)

As alternativas de parceria de pesquisas do setor público com o setor privado esbarram também em consideráveis dificuldades.²⁶ Dos atores envolvidos no complexo agroindustrial, os agentes privados com maior potencial de financiar pesquisas são, tradicionalmente, os produtores de insumos. Entretanto, como multinacionais, tendem a concentrar tais esforços onde se localizam suas sedes, como já citado; seus interesses passam também por todas as restrições de apropriação de resultados que acabaram de ser discutidas. O outro lado do complexo, em que se localizam as indústrias de transformação e beneficiamento de produtos agrícolas, infelizmente não parece um campo fértil para tais esforços, já que o montante de recursos que costumam destinar à pesquisa científica é inferior a 2% de sua receita.²⁷

Associações de produtores seriam outros parceiros em potencial; mas também neste caso é preciso lembrar que os setores organizados e capitalizados para tal são poucos. É improvável que tais organizações estivessem dispostas a financiar estudos que não os

25 Dados apresentados por Hirst & Thompson, citados por Batista Jr. (1996), mostram que apenas 10% a 30% da atividade tecnológica das grandes corporações dos países desenvolvidos acontece em subsidiárias estrangeiras. Patel & Pavitt, também citados por Batista Jr. (1996), demonstram que as firmas das principais economias do mundo - Alemanha, Japão e EUA - realizam menos de 15% das atividades de pesquisa e desenvolvimento fora do país de origem, e concluem que a produção de tecnologia constitui um caso importante de não-globalização.

26 Uma visão mais otimista sobre essa possibilidade encontra-se em Portugal & Contini (1997).

27 Detalhes em Martinelli Jr. (1997).

beneficiassem diretamente, e exemplos anteriores - como demonstrado por Silva (1986) - indicam que certas lavouras podem ter sua evolução prejudicada quando menos contempladas pelo esforço de pesquisa.

A complexidade desse problema é maior devido às suas peculiaridades, já que além das restrições econômicas, o tempo de maturação dos resultados da pesquisa científica aplicada à agricultura é longo: tipicamente de 8 a 10 anos para que sejam obtidas novas variedades, e freqüentemente de 15 a 20 anos desde os primeiros passos até que seu impacto sobre a produção seja percebido. (Hazell, 1995)

A produção agrícola atual, suas particularidades e sua eficiência provêm de pesquisas efetuadas há décadas atrás; portanto, é necessário ter em mente que a oferta futura, daqui a 20, 30 anos, será consequência das políticas desenvolvidas hoje.

Por último, com relação aos métodos econométricos utilizados, deve-se destacar que os modelos de função de produção de fronteira mostraram-se muito sensíveis - tanto à forma funcional escolhida, quanto à especificação assumida para a distribuição das medidas de eficiência -, fornecendo estimativas substancialmente diferentes.

Referências bibliográficas

- Aigner, D. J.; Lovell, C. A. K.; Schmidt, P. Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, v. 6, n. 1, p. 21-37, jul. 1977.
- Alves, E. R. A. Entrevista: desafios da pesquisa agrícola. *Economia Rural*, v. 7, n. 4, p. 2-6, out./dez. 1996.
- Barros, J. R. M. Política e desenvolvimento agrícola no Brasil. In: Veiga, A. (coord.), *Ensaio sobre política agrícola brasileira*. São Paulo: Secretaria da Agricultura, 1979, p. 9-35.
- Batista Jr., P. N. O mito da empresa transnacional. *Folha de S. Paulo*, 10 nov. 1996, p. 2-2.
- Battese, G. E. Frontier production functions and technical efficiency: a survey of empirical applications in agricultural economics. *Agricultural Economics*, v. 7, n. 3/4, p. 185-208, oct. 1992.
- Battese, G. E. & Coelli, T. J. A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*, 20, p. 325-332, 1995.

- Battese, G. E.; Rambaldi, A. N.; Wan, G. H. A stochastic frontier production function with flexible risk properties. *Journal of Productivity Analysis*, v. 8, n. 3, p. 269-280, aug. 1997
- Bell, M. & Pavitt, K. Accumulating technological capability in developing countries. *In: THE WORLD BANK ANNUAL CONFERENCE ON DEVELOPMENT ECONOMICS*. Washington, D.C., april 30 - may 1, 1992. *Proceedings*. Washington: The World Bank, mar. 1993, p. 257-281
- Brasil. Ministério da Agricultura. Secretaria Nacional de Planejamento Agrícola. *Aptidão agrícola das terras de São Paulo*. Brasília: BINAGRI, 1979. 114 p. (Aptidão agrícola da terras, 20)
- Campos, H. *Estatística experimental não-paramétrica*. Piracicaba: ESALQ/USP, 1976. 332 p.
- David, P.A. Knowledge, property, and the system dynamics of technological change. *In: THE WORLD BANK ANNUAL CONFERENCE ON DEVELOPMENT ECONOMICS*. Washington, D.C., april 30 - may 1, 1992. *Proceedings*. Washington: The World Bank, mar. 1993, p. 215-248.
- Dresrüsse, G. Declining assistance to developing-country agriculture: change of paradigm? *In: A 2020 VISION FOR FOOD, AGRICULTURE, AND THE ENVIRONMENT*. Washington, D.C., june 13-15, 1995. *Brief 16*. Washington: IFPRI, apr. 1995.
- Evenson, R. E. & Kislev, Y. Research and productivity in wheat and maize. *Journal of Political Economy*, v. 81, n. 6, p. 1.309-1329, nov./dec. 1973.
- Farrell, M. J. The measurement of productive efficiency. *Journal of Royal Statistical Society, A*, 120, Part 3, p. 253-281, 1957
- Førsund, F. R.; Lovell, C. A. K.; Schmidt, P. A survey of frontier productions functions and of their relationship to efficiency measurement. *Journal of Econometrics*, v. 13, n. 1, p. 5-25, may. 1980.
- Gasques, J. G. & Verde, C. M. V. Crescimento da agricultura brasileira e política agrícola nos anos oitenta. *Agricultura em São Paulo*, v. 37 n. 1, p. 183-204, 1990.
- Greene, W. H. *LIMDEP version 7.0 user's manual*. New York: Econometric Software, 1995. 850 p.
- Hayami, H. & Ruttan, V. W. *Desenvolvimento agrícola: teoria e experiências internacionais*. Brasília: EMBRAPA, 1988. 583 p.
- Hazell, P. Technology's contribution to feeding the world in 2020. *In: A 2020 VISION FOR FOOD, AGRICULTURE, AND THE ENVIRONMENT*. Washington, D.C., june 13-15, 1995. *Speeches Made at an International Conference*. Washington: IFPRI, aug. 1995, p. 80-82.

- Holt, D.; Smith, T. M. F.; Winter, P. D. Regression analysis of data from complex surveys. *Journal of Royal Statistical Society, A*, 143, Part 4, p. 474-487. 1980.
- Homem de Melo, F. B. Disponibilidade de tecnologia entre produtos da agricultura brasileira. *Revista de Economia Rural*, v. 18, n. 2, p. 221-249, abr./jun. 1980.
- _____. Um diagnóstico sobre produção e abastecimento alimentar no Brasil. *Agricultura em São Paulo*, 35, (T. especial), p. 115-156, 1988.
- _____. A questão da política de preços para produtos agrícolas domésticos. *Revista Brasileira de Economia*, v. 45, n. 3, p. 385-396, jul./set. 1991
- Martinelli Jr., O. *As tendências recentes da indústria de alimentos: um estudo a partir das grandes empresas*. Campinas: IE/UNICAMP, 1997. (Tese de Doutorado)
- Matos, O. C. *Econometria básica: teoria e aplicações*. 2ª ed. São Paulo: Atlas, 1997
- Mesquita, T.C. *Desempenho da agricultura brasileira e sua relação com alguns instrumentos de política econômica 1970/1990*. São Paulo: FEA/USP, 1994. 224 p. (Tese de Doutorado)
- Oram, P. The potential of technology to meet world food needs in 2020. *In: A 2020 VISION FOR FOOD, AGRICULTURE, AND THE ENVIRONMENT*. Washington, D.C., June 13-15, 1995. *Brief 13*. Washington: IFPRI, Apr. 1995.
- Pardey, P. G. & Alston, J. M. Revamping agricultural R & D. *In: A 2020 VISION FOR FOOD, AGRICULTURE, AND THE ENVIRONMENT*. Washington, D.C., June 13-15, 1995. *Brief 24*. Washington: IFPRI, Jun. 1995.
- Pinstrup-Andersen, P. Toward a consensus for action. *In: A 2020 VISION FOR FOOD, AGRICULTURE, AND THE ENVIRONMENT*. Washington, D.C., June 13-15, 1995. *Speeches Made at an International Conference*. Washington: IFPRI, Aug. 1995. p. 104-108.
- Portugal, A. D. & Contini, E. O público e o privado na pesquisa agropecuária brasileira. *In: CONGRESSO BRASILEIRO DE ECONOMIA E SOCIOLOGIA RURAL*, 35, Natal, 4 a 8 de agosto de 1997 *Anais*. Brasília: SOBER, 1997 p. 38-52.
- Prognóstico. São Paulo, Instituto de Economia Agrícola, 1971-1988. v. 0-17.
- Redding, S. The low-skill, low-quality trap: strategic complementarities between human capital and R & D. *The Economic Journal*, v. 106, n. 435, p. 458-470, Mar. 1996.

- Ricardo, D. *Princípios de economia política e de tributação*. São Paulo: Abril Cultural, 1982.
- Romer, P. M. Two strategies for economic development: using ideas and producing ideas. *In: THE WORLD BANK ANNUAL CONFERENCE ON DEVELOPMENT ECONOMICS*. Washington, D.C., april 30 - may 1, 1992. *Proceedings*. Washington: The World Bank, mar. 1993. p. 63-91.
- Rosegrant, M. Who will go hungry? Scenarios for future global and regional food supply and demand. *In: A 2020 VISION FOR FOOD, AGRICULTURE, AND THE ENVIRONMENT*. Washington, D.C., june 13-15, 1995. *Speeches Made at an International Conference*. Washington: IFPRI, aug. 1995. p. 29-37
- Saheli, S. & Macedo, P. B. R. Eficiência técnica das unidades federativas brasileiras – padrões e determinantes. *Economia Aplicada*, v. 2, n. 4, p. 647-679, out./dez. 1998.
- Silva, G. L. S. P. *Produtividade agrícola, pesquisa e extensão rural*. São Paulo: IPE/USP, 1984. 143 p.
- _____. *Pesquisa, tecnologia e rendimento dos principais produtos da agricultura paulista*. São Paulo: IEA, 1986. (Relatório de Pesquisa 12/86)
- _____. Transforming Brazilian agriculture. *In: CONGRESSO BRASILEIRO DE ECONOMIA E SOCIOLOGIA RURAL*, 29, Campinas, 28 de julho a 1 de agosto de 1991. *Anais*. Brasília: SOBER, 1991. p. 254-278.
- Schultz, T. W. *A transformação da agricultura tradicional*. Rio de Janeiro: Zahar, 1965. 208 p.
- _____. The value of the ability to deal with disequilibria. *Journal of Economic Literature*, v. 13, n. 3, p. 827-846, sep. 1975.
- Vicente, J. R. *Influência de educação, pesquisa e assistência técnica na produtividade da agricultura brasileira na década de setenta*. Piracicaba: ESALQ/USP, 1989. 193 p. (Dissertação de Mestrado)
- _____. *Determinantes da adoção de tecnologia e da eficiência na produção agrícola paulista*. São Paulo: FEA/USP, 1997. 224 p. (Tese de Doutorado)

Sazonalidade em séries temporais quadrissemanais – o caso do IMEC*

Eliezer Martins Diniz[§]

RESUMO

O objetivo deste artigo é utilizar o ferramental econométrico disponível para analisar a sazonalidade na série do Indicador de Movimentação Econômica (IMEC), calculado pela Fundação Instituto de Pesquisas Econômicas (FIPE). É mais plausível para essa série supor um modelo com transição suave na tendência e na constante do que uma raiz unitária. Com base nisso, procura-se encontrar algum tipo de comportamento sazonal típico intra-mês na série sem tendência. Há evidências de sazonalidade determinista significativa na série nas duas primeiras quadrissemanas, com um sinal positivo na primeira e negativo na segunda. Encontramos também indícios de comportamento periódico, mostrando uma resposta diferenciada da série dependendo da quadrissemana em que ocorre o dado.

Palavras-chave: transição suave, sazonalidade, periodicidade, indicador de movimentação econômica.

ABSTRACT

This paper uses the available econometric techniques to analyze the seasonality in the time series of the Indicator of Economic Movements (IMEC), measured by Fundação Instituto de Pesquisas Econômicas (FIPE), from São Paulo (Brazil). It is more plausible for this series to consider a model with a smooth transition in trend and intercept rather than the alternative hypothesis of a unit root. Taking these findings into account, we then searched for any type of typical seasonal behavior inside the month in the detrended time series. There is evidence of deterministic seasonality in the beginning of the month, with a positive impact in the first week and a negative impact in the second week. There is also evidence of periodic behavior, indicating a different response of the series dependent on the week that the observation occurs.

Key words: smooth transition, seasonality, periodicity, indicator of economic movements.

* Agradeço a Zeina Abdel Latif pelas informações prestadas a respeito da série estudada aqui. Agradeço também a um parecerista anônimo desta revista. Os erros são de minha inteira responsabilidade.

§ Professor da FEA-USP/RP. Doutor em Economia pela FEA-USP e Pós-Doutorando pela University of Oxford. Pesquisador associado do McKeever Institute of Economic Policy Analysis (MIEPA) em Berkeley, CA, USA.

Endereço para correspondência: Faculdade de Economia, Administração e Contabilidade - Universidade de São Paulo - Av. dos Bandeirantes, 3900 - Monte Alegre - 14040-900 - Ribeirão Preto - SP - Brasil. e-mail: elmdiniz@usp.br

Recebido em janeiro de 1999. Aceito em março de 1999.

1 Introdução

A sazonalidade é uma característica pertencente a toda série temporal, em maior ou menor intensidade. Os economistas, tradicionalmente, não dispõem um tempo razoável para analisar este componente da série. Costuma-se supor que a sazonalidade não guarda relação com a tendência e o ciclo, isto é, a série é formada de componentes aditivos independentes entre si que podem ser analisados separadamente. O máximo de refinamento que se encontra nas análises é a hipótese de componentes multiplicativos, o que recai no caso aditivo ao se aplicar o logaritmo com o fito de estabilizar a variância. Muitas vezes também se supõe que a sazonalidade possui um padrão constante ao longo do tempo. Por meio dessa simplificação, chega-se a uma estratégia para se atacar o problema, que consiste em captar a sazonalidade por intermédio de variáveis *dummy*, eliminando-a em seguida para se proceder a uma análise de um dos dois componentes restantes: tendência ou ciclo. Isto é, segundo esse paradigma o componente sazonal não forneceria nenhuma informação relevante para o estudo da tendência ou do ciclo.¹

A insistência no estudo da tendência está enraizada na própria teoria econômica. Uma vez que a tendência capta o longo prazo das séries, e que os modelos econômicos trazem pontos de equilíbrio que são atingidos no longo prazo, procura-se, mediante o estudo da tendência das séries, descobrir a plausibilidade empírica de tais resultados. Obviamente, há diversas formas para se estudar o problema. A mais satisfatória e promissora é a descoberta da cointegração entre conjuntos de variáveis, pois esta revela a relação de longo prazo entre as variáveis envolvidas. O ajuste rumo ao equilíbrio é captado por um modelo de correção de erros, que nada mais é do que a tradução econométrica das equações dinâmicas de ajuste rumo ao equilíbrio que aparecem em qualquer modelo com um mínimo de complexidade. A análise deste tipo de fenômeno é bastante significativa na literatura econômica.

Os ciclos revelam um comportamento de curto prazo da série que pode ser explicado por outros fatores que não os de longo prazo. As pesquisas de curto prazo em geral tratam dos ciclos de negócios (*business cycles*), importantes para o dia-a-dia da economia. Há ainda o estudo incipiente dos ciclos sazonais.

A sazonalidade pode ser de três tipos: determinista, que é completamente previsível, captada perfeitamente por variáveis *dummy*; a estocástica não estacionária, que não é

¹ A eliminação da sazonalidade também é feita muitas vezes por filtros do tipo X-11. O problema de qualquer procedimento que elimine a sazonalidade é a perda da informação que ela contém.

perfeitamente previsível mas cujos efeitos se dissipam ao longo do tempo; e a estocástica integrada, que apresenta mudanças significativas de longo prazo. A compreensão do que é a sazonalidade estocástica integrada se dá ao fazermos uma analogia com o caso da integração no longo prazo, onde uma tendência determinista não consegue descrever a série, que se comporta como se a tendência estivesse mudando ao longo do tempo. A sazonalidade integrada opera como se o padrão sazonal estivesse se modificando ao longo do tempo. Um raciocínio *a priori* nos mostra que intuitivamente não há papel nas séries econômicas para a integração sazonal, pois isto equivaleria, por exemplo, ao impacto do mês de dezembro ocorrer em março a partir de uma data, sem razão aparente. Pode-se dizer que é mais provável encontrar sazonalidade determinista, ou estocástica estacionária, ou uma combinação de ambas. No entanto, é possível que haja mudanças no ambiente econômico e nas preferências dos agentes que alterem o padrão de comportamento dos agentes econômicos. No entanto elas precisam ser significativas.

A questão ainda pouco explorada é a seguinte: qual a importância da sazonalidade para o economista? É necessário verificar se o comportamento sazonal realmente não apresenta efeitos no longo prazo (questionando a existência de um padrão sazonal fixo ou estacionário ao longo do tempo) e se os ciclos de negócios não são influenciados por fatores sazonais (questionando a independência entre ciclo e sazonalidade). As hipóteses de efeitos cíclicos e sazonais a curto prazo, que se dissipam ao longo do tempo, e efeitos de tendência, que permanecem no longo prazo, juntamente com a independência de ciclo e sazonalidade, tornam o fenômeno econômico muito mais fácil de ser modelado, uma vez que ciclo, sazonalidade e tendência passam a ser características isoladas, que podem ser estudadas independentemente. Se os componentes ciclo, sazonalidade e tendência não puderem ser adequadamente separados, então há uma repercussão também em nível teórico, pois torna-se impossível fazer uma teoria em que o longo prazo seja completamente dissociado do curto prazo.

Aqui nos concentramos no estudo de um índice de movimentação quadrissemanal. É necessário imaginar quais os determinantes do comportamento sazonal dentro do mês. O principal fator a ser considerado é o conjunto de **práticas de pagamentos**, pois nas datas em que ocorrem os pagamentos de **salários** e em que se concentram os pagamentos de **contas** deve ocorrer um maior nível de atividade, caracterizado por deslocamentos dentro da cidade, por compras em supermercados e lojas, por passeios, por idas aos bancos, e muitas outras situações que são de difícil enumeração.

Neste artigo a análise é feita utilizando-se um índice particular: o Indicador de Movimentação Econômica (IMEC), calculado pela Fundação Instituto de Pesquisas

Econômicas (FIPE) em São Paulo. Usamos a versão quadrissemanal do índice, o que apresenta algumas vantagens. Em primeiro lugar, há um grande número de dados em um período relativamente curto de tempo. Segundo, possui quatro dados em um mês, e por isso pode aplicar testes convencionais de raízes unitárias sazonais que verificam a plausibilidade de aplicação do filtro $\Delta_4 = (1 - L^4)$, onde L é o operador defasagem. Dessa forma seria examinada a sazonalidade **intra-mês**.² É possível também estudar a sazonalidade **intra-trimestre**, ao se verificar a adequação do filtro $\Delta_{12} = (1 - L^{12})$.³ Outra vantagem é a de que temos os dados originais e os dessazonalizados, dando maior flexibilidade à análise.⁴ Em nossa análise nos limitaremos ao estudo das séries originais. Temos também que não há na literatura estudos de séries quadrissemanais que chegaram a nosso conhecimento. Por último, os estudos de variáveis que envolvam alguma medida de atividade da economia utilizam dados anuais, trimestrais e, às vezes, mensais. Nosso estudo captará, portanto, movimentos **intra-mês** que não são captados em dados mensais, o que torna a nossa empreitada promissora. É, portanto, um estudo pioneiro.

A exposição segue a seguinte estrutura: a seção 2 faz uma exposição sucinta sobre o IMEC, a seguinte se destina a uma análise econométrica dos dados e a última terá as conclusões.

2 Características e metodologia do IMEC

O Indicador de Movimentação Econômica (IMEC) foi concebido pela FIPE como um sinalizador do nível de atividade da economia que se apóia em uma motivação diferente em relação aos índices comumente utilizados: ele procura captar o **deslocamento** dos agentes econômicos, sob a hipótese de que este sempre possui uma movimentação econômica. Por exemplo, a saída de uma pessoa de sua casa está ligada a uma motivação econômica quando ela vai trabalhar, ou fazer compras, ou usufruir do lazer (cinemas, restaurantes etc.). Raramente um deslocamento não envolve uma atividade econômica, logo, é possível encarar as exceções como um número insignificante. O IMEC envolve em sua

2 A única hipótese adicional necessária é a suposição irrealista de que cada quadrissemana contém o mesmo número de dias. Esse é um ponto considerado para a elaboração do índice. Supomos que essa aproximação seja válida para os propósitos do artigo.

3 Não perseguimos essa estratégia no artigo.

4 Os dados originais não são, na verdade, brutos. Segundo os responsáveis, foi eliminado todo vestígio de efeito calendário. Não foi possível obter os dados brutos.

maioria índices que revelam **movimentação** de agentes econômicos, mas possui alguns indicadores tradicionais. Para que o dado estivesse disponível rapidamente, optou-se por um grupo pequeno de variáveis em geral restrito à Região Metropolitana de São Paulo.

Os índices que compõem o IMEC são:

- 1) número de passageiros (entradas) no METRÔ de São Paulo;
- 2) número de passageiros de ônibus urbanos na cidade de São Paulo;
- 3) número de passageiros (saídas) de ônibus interurbanos no Terminal Rodoviário do Tietê;
- 4) fluxo de transporte rodoviário de passeio e de carga no primeiro pedágio do sistema Anhanguera-Bandeirantes;
- 5) números de passageiros (embarques e desembarques) no Aeroporto de Congonhas;
- 6) consumo de energia elétrica na Região Metropolitana de São Paulo e áreas adjacentes do interior;
- 7) consumo de gasolina, álcool e óleo *diesel* na Região Metropolitana abrangida pelas distribuidoras de Barueri, Guarulhos e Utinga;
- 8) vendas no comércio em cartão de crédito ou cheque.

Os fornecedores dos dados são, respectivamente: Companhia do Metropolitano de São Paulo; São Paulo Transportes S/A, SOCICAM Administração, Projetos e Representação Ltda; DERSA Desenvolvimento Rodoviário S/A, INFRAERO Empresa Brasileira de Infra-Estrutura Aeroportuária; ELETROPAULO Eletricidade de São Paulo S/A; PETROBRÁS Petróleo Brasileiro S/A; e Associação Comercial de São Paulo.

A movimentação de pessoas por meio de transporte coletivo na cidade de São Paulo está em boa parte captada pelos três primeiros indicadores. Só ficam de fora os passageiros que se utilizam de ônibus intermunicipais, controlados pela EMTU (Empresa Metropolitana de Transportes Urbanos), e aqueles que se utilizam de ônibus que partem dos Terminais Rodoviários Bresser e Jabaquara de menor porte, bem como de transporte coletivo fretado (para viagens e transporte de funcionários).

O fluxo de transporte rodoviário está representado pelo quarto índice, que revela a integração da cidade de São Paulo com o interior do Estado e até com Minas Gerais e Brasília por meio do sistema Anhanguera-Bandeirantes. A integração de São Paulo com o resto do País e o interior também é captada pelo quinto índice. Não são contados, no entanto, os vôos internacionais que saem do Aeroporto de Cumbica em Guarulhos.

Os últimos três índices são de caráter mais tradicional, revelando o nível de atividade econômica. Todos se restringem à Região Metropolitana de São Paulo.

Todos os índices obtidos passam por exclusão de sábados, domingos, feriados e dias atípicos (por exemplo, greves e feriados prolongados) para então serem utilizados para formar o IMEC. Os dados mais pobres são os relativos aos ônibus interurbanos e aos combustíveis, que só são disponíveis semanalmente.

Na construção do índice quadrissemanal utilizado aqui são seguidos os seguintes passos:

- 1) coleta dos dados;
- 2) construção de cada um dos índices com base nos dados;
- 3) divisão de cada série pelo fator de sazonalidade obtido aplicando-se o procedimento X-11 aos dados **mensais**;
- 4) subtração, dos dados obtidos no passo anterior, do efeito das *dummies*, eliminando qualquer efeito calendário;
- 5) construção dos números índices das séries dessazonalizadas;
- 6) montagem das séries original e dessazonalizada do IMEC, tendo como pesos aqueles obtidos pelo método de componentes principais para a série mensal.

A série do IMEC dessazonalizada é obtida aplicando-se os procedimentos de dessazonalização a seus componentes individualmente, e não ao número-índice final. A justificativa dada para se seguir o terceiro passo na dessazonalização é a necessidade de que os índices mensal e quadrissemanal caminhem juntos (aderência). Aplica-se, dessa forma, um fator de sazonalidade, que é a média da sazonalidade do mês. Portanto, concebe-

se a sazonalidade **intra-mês** como associada somente ao efeito calendário, uma vez que as *dummies* estão associadas a datas comemorativas, feriados, férias, catástrofes naturais e dias de pagamento. Nossa hipótese de trabalho é justamente a de que há outros tipos de motivação para decisões que não o efeito calendário, e que ao contrário deste não são perfeitamente previsíveis, refletindo as regras de decisão dos agentes que podem variar conforme suas expectativas se modifiquem. Logo, deve haver sazonalidade que não foi retirada nem tratada. Essa é a razão pela qual nosso estudo é promissor. Nosso trabalho não é uma crítica à metodologia utilizada para calcular o IMEC. Muito pelo contrário, a solução encontrada é operacional. Mas é uma tentativa de obtenção de um conjunto maior de informações a partir dos dados disponíveis.

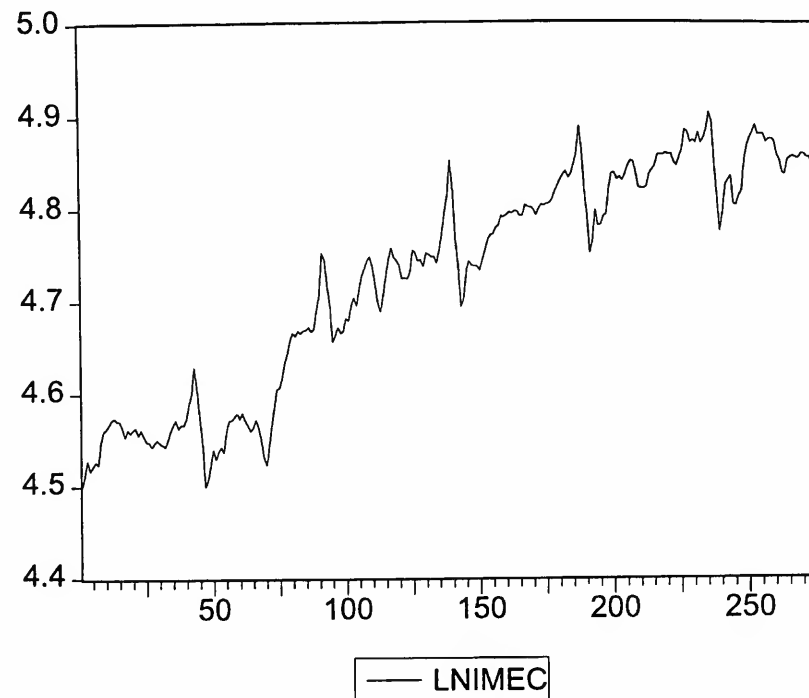
As referências bibliográficas originais principais estão citadas no texto. No Brasil há poucas resenhas e aplicações do ferramental econométrico utilizado aqui. Os de nosso conhecimento são Aguirre (1998), Diniz (1996) e Oliveira & Pichetti (1997).

3 Análise econométrica da série⁵

Utilizamos uma série do IMEC e seus componentes com 276 dados quadrissemanais (totalizando 69 meses) de fevereiro de 1993 a outubro de 1998 (inclusive). Os dados, como já colocamos acima, **não** estão dessazonalizados.

A série do logaritmo do IMEC está mostrada na Figura 1. A variável apresenta um comportamento que difere da distribuição normal (Jarque-Bera de 26,3).

5 Todos os resultados econométricos se verificam ao nível de significância de 1%, a não ser que expressamente indicado. Os valores críticos utilizados podem ser encontrados em diversos artigos e livros: Fuller (1976), Davidson & MacKinnon (1993), Dickey & Fuller (1981), Hylleberg, Engle, Granger & Yoo (1990), Ghysels, Lee & Noh (1994), Smith & Taylor (1998) e Leybourne, Newbold & Vougas (1998). Os resultados pressupõem que a amostra seja suficientemente grande de modo que os resíduos i.i.d. obtidos possam ser confrontados com as tabelas de valores críticos usuais (veja, por exemplo, Hamilton (1994) Tabela 17.1, p. 502). No entanto, para tornar os resultados mais robustos, foram utilizados também os valores assintóticos de acordo com a estratégia descrita em Davidson & MacKinnon (1993). O método de estimação foi o de mínimos quadrados ordinários, com as exceções apontadas no texto.

Figura 1 – Logaritmo do IMEC

Examinando as autocorrelações e as correlações parciais estimadas vemos que a série apresenta um comportamento auto-regressivo, onde há coeficientes nas defasagens 1, 2 e 5. De fato, o modelo estimado é dado por

$$\lnimec = 0,044619 + 1,425158 \lnimec(-1) - 0,581284 \lnimec(-2) + 0,146946 \lnimec(-4),$$

(1,44) (26,82) (-9,56) (5,97)

com AIC=-5,918498. Os valores entre parêntesis são as estatísticas t, que apontam para a significância de todos os coeficientes (exceto a constante a 15%). Os resíduos apresentam autocorrelações e correlações parciais estimadas com um padrão sazonal bem delineado (os valores significativos ocorrem sempre nos múltiplos de 4). Logo, os resultados apontam para uma provável raiz unitária e também para a existência de sazonalidade.

Pode-se questionar sobre os tipos de raiz unitária que a série apresenta. O teste HEGY aplicado a séries quadrissemanais consegue obter essa informação. Um sumário das estatísticas é encontrado na Tabela 1. O aumento da equação ocorre em todos os casos analisados nas defasagens 1, 2, 5, 6, 8, 44, 45, 46 e 48. Os resíduos de todas as equações estimadas se mostraram independentes e identicamente distribuídos (fato examinado pelos testes de Breusch-Godfrey de 1 a 4 defasagens e pelo teste ARCH com uma defasagem), o que torna a inferência a partir dos testes válida.

Tabela 1
Testes HEGY para a Série ln(IMEC)

MODELO	\hat{t}_1	\hat{t}_2	\hat{t}_4	\hat{F}_{34}	\hat{F}_{234}	\hat{F}_{1234}
1111	-1,61	-12,92*	-12,92*	175,76*	152,42*	121,19*
1110	-1,61	-13,00*	-13,01*	178,27*	154,48**	122,83**
1010	-1,62	-13,10*	-13,11*	180,83*	156,71**	124,60**
1000	-2,14	-13,17*	-13,32*	188,96*	162,41**	125,58**

Legenda: * 1%; ** 5%; embora os testes F estejam todos rejeitando claramente a hipótese nula a 1%, colocamos para alguns deles o nível de 5% porque é a tabulação disponível mais próxima. Modelo *abcd*: *a* para constante, *b* para *dummies* sazonais, *c* para tendência e *d* para *dummies* na tendência; cada um deles assume valor 1 se a variável a que se refere está presente na equação estimada, e 0 se está ausente.

Os resultados da Tabela 1 mostram, para todos os modelos estimados, que não há nenhuma raiz unitária sazonal e só há uma raiz unitária na frequência zero. A coluna com a estatística \hat{t}_1 avalia a hipótese nula de presença de uma raiz unitária na frequência zero (movimentos de longo prazo). Vemos que em nenhuma das equações houve rejeição da hipótese nula. A estatística \hat{t}_2 avalia a hipótese nula de presença de uma raiz unitária na frequência π , o que é rejeitado em todas as equações. As estatísticas \hat{t}_4 e \hat{F}_{34} avaliam a hipótese nula de presença de uma raiz unitária na frequência $\pi/2$, o que é rejeitado para todas as equações. O teste \hat{F}_{234} avalia a existência de raízes unitárias sazonais (frequências π e $\pi/2$), rejeitando a hipótese nula. O único resultado contraditório é dado pela estatística \hat{F}_{1234} , que avalia a existência de qualquer raiz unitária. Há rejeição da hipótese nula por causa da intensidade com que as raízes unitárias sazonais são rejeitadas comparativamente à não rejeição da raiz unitária na frequência zero. A tendência, as *dummies* sazonais de tendência e as *dummies* sazonais de intercepto também se mostram irrelevantes, embora testes mais rigorosos necessitem ser tabulados.⁶ A constante no modelo mais parcimonioso aparenta ser relevante. Um resultado que vá além da conjectura não é possível porque não há um teste que avalie conjuntamente a significância do intercepto e a presença da raiz unitária para o caso sazonal. No entanto, o resultado particular para nossa série mostra que só há uma raiz unitária na frequência zero, e por isso o teste ADF é o mais adequado. Logo, a presença de uma raiz unitária na frequência zero e a significância da constante devem ser reavaliados pelo teste ADF.

O teste de raiz unitária de Dickey-Fuller aumentado (ADF) revela a existência de uma raiz unitária. O modelo mais parcimonioso estimado com constante e tendência que produz resíduos independentes e distribuídos identicamente é dado por⁷

6 Semelhantes aos de Dickey & Fuller (1981), mas específicos para o caso sazonal.

7 De acordo com Davidson & MacKinnon (1993) o uso de valores assintóticos para os testes é o mais indicado quando os resíduos não são normais ou deixam de ser homocedásticos. Um conjunto de valores críticos fornecidos pelos

$$\ln imec = 0,088789 + 0,00000755 t + 0,981460 \ln imec(-1) + (\text{defasagens de } \Delta \ln imec),$$

$$(1,02) \quad (0,25) \quad (51,07)$$

onde o aumento do teste ocorre nas defasagens 1, 4, 5, 8, 9, 12, 13, 16, 17, 44 e 47. Este aumento produz os resultados desejados nos resíduos (homocedásticos e sem correlação serial) avaliados conforme os testes descritos na estimação do teste HEGY.

Utilizamos inicialmente uma estatística para avaliar se na presença da constante temos simultaneamente a tendência insignificante e uma raiz unitária. O resultado ($\hat{\Phi}_3 = 2,16$) mostra que a hipótese nula não pode ser rejeitada. Fizemos então um teste tipo t para verificar a presença de raiz unitária, obtendo que $\hat{\tau}_\tau = -0,96$, o que confirma a presença de uma raiz unitária. Fazemos agora um teste conjunto de significância do intercepto, da tendência e da raiz unitária. Obtemos $\hat{\Phi}_2 = 3,39$, o que, juntamente com os resultados anteriores, mostra que o intercepto não é significativo. Os resultados $\hat{\tau}_{\alpha\tau} = 1,02$ e $\hat{\tau}_{\beta\tau} = 0,25$ confirmam, respectivamente, que a constante e a tendência não são significativas. Logo, a série deve ser um passeio aleatório sem intercepto. Para confirmar os resultados, estimamos o mesmo modelo acima sem a tendência. Os resíduos se mostram satisfatórios como no caso anterior, e o teste de significância do intercepto, juntamente com a presença da raiz unitária, produz $\hat{\Phi}_1 = 5,08$ o que aponta para a insignificância do intercepto.⁸ Os testes t da raiz unitária e do intercepto produzem, respectivamente, $\hat{\tau}_\mu = -2,06$ e $\hat{\tau}_{\alpha\mu} = 2,12$, não rejeitando a presença de uma raiz unitária e a não significância do intercepto. Logo, toda a bateria de testes aponta para um passeio aleatório sem intercepto. Estimando o mesmo ADF sem constante e sem intercepto (com resíduos i.i.d.) obtivemos a confirmação da raiz unitária com $\hat{\tau} = 2,36$.⁹ Como alternativa ao teste ADF, fizemos o teste de Phillips-Perron

autores (Tabela 20.1, p. 708) foi confrontado com os resultados obtidos pela tabulação padrão, sendo que todas as conclusões foram mantidas.

8 A não significância do intercepto só pôde ser avaliada rigorosamente no contexto do ADF, juntamente com os testes de Dickey & Fuller (1981).

9 Davidson & MacKinnon (1993, p. 705) sugerem a utilização de *dummies* sazonais juntamente com o teste ADF e os mesmos valores críticos. Em nossa série, como era de se esperar, as *dummies* não se mostraram significativas (corroborando o resultado de não significância da constante). A utilização de uma variável de tendência ao quadrado nos testes ADF, apesar de ter diminuído as estatísticas t estimadas, não levou à rejeição da hipótese nula de raiz unitária. Pode-se verificar os dois resultados separadamente com os mesmos aumentos do teste acima.

(PP), o qual é menos sensível a resíduos com correlação serial.¹⁰ O teste aponta para a existência de raiz unitária em todas as alternativas ($z_{ct} = -3,28$, $z_c = -1,67$ e $z_{nc} = 1,31$).¹¹ Os resultados só vêm confirmar aqueles obtidos com o teste ADF.

Um ponto que merece destaque é a presença **ou** de observações aberrantes **ou** de um padrão comportamental diferenciado no início da série. Um exame da Figura 1 mostra que do início até ao redor da observação 90 o comportamento da série difere do restante. O gráfico aponta duas alternativas: **ou** a série é estacionária até ao redor da observação 80 e depois passa a ter um comportamento com tendência crescente **ou** há um padrão único para toda a série, que sofre algum tipo de modificação entre as observações 40 e 80. O período a partir da observação 90 aparenta ter o mesmo tipo de comportamento, isto é, ser gerado por um só modelo.

Vimos que os testes ADF, PP e HEGY produziram resultados semelhantes. No entanto, a não rejeição da hipótese nula de uma raiz unitária pode ocorrer em modelos com tendência não estocástica com uma quebra estrutural. As observações do parágrafo anterior mostram que a ocorrência de quebra estrutural na presente série é possível. Portanto, é necessário complementar o ADF com algum outro tipo de teste que permita escolher entre as alternativas de raiz unitária e de tendência determinista com quebra estrutural. Um modelo em que esse tipo de teste surge naturalmente é aquele que incorpora transições suaves na tendência e/ou intercepto. Essa classe de modelos é descrita em Granger & Teräsvirta (1993). Um teste com dois passos para se decidir pela existência de uma raiz unitária ou pela hipótese alternativa de um modelo com transição suave é sugerido por Leybourne, Newbold & Vougas (1998). O teste proposto é uma generalização daqueles sugeridos por Perron para distinguir uma quebra de tendência de uma raiz unitária na série.

Para se obter a resposta desejada para o nosso problema com base no novo teste estimamos a equação

$$\ln \text{imec} = 4,546796 + 0,000216t + (0,078314 + 0,000731t) S(0,270387, 0,279994) ,$$

$$(672,1) \quad (1,17) \quad (8,04) \quad (3,95) \quad (3,68) \quad (63,0)$$

com $AIC = -4,323$, e onde $S(\hat{\gamma}, \hat{\tau})$ é a função de transição suave dada pela função logística

$$S(\hat{\gamma}, \hat{\tau}) = [1 + \exp(-\hat{\gamma}(t - \hat{\tau}T))]^{-1} ,$$

¹⁰ Cf., por exemplo, Holden & Perman (1994, p. 66).

¹¹ Onde ct = constante e tendência, c = constante e nc = não há constante.

com as propriedades usuais. A estimação, feita por mínimos quadrados não lineares, utiliza o algoritmo de Marquardt. Os valores iniciais dos dois primeiros parâmetros são dados por uma regressão preliminar em mínimos quadrados ordinários envolvendo constante e tendência. Os dois valores iniciais seguintes são zero e os dois últimos são 1 e 0,5. A regressão confirma nossas suspeitas anteriores: até um determinado ponto a tendência determinista não é significativa, passando a ser a partir daí. Há uma transição suave para um novo intercepto e para uma tendência. A taxa à qual ocorre a transição é de 27%. A observação crítica da série é a de número 77 ($=0,279994 \times 276$), situada em 15 de setembro de 1994, certamente associada aos primórdios do Plano Real.¹² Logo, há uma mudança nos rumos da economia que se deve ao novo plano econômico, que conduz o País a uma trajetória de crescimento continuado. A taxa de crescimento dada pela tendência passa de 1,04% a.a., no momento inicial, para 4,64% a.a., que seria atingido no *steady state* na ausência de qualquer outro tipo de mudança. Como segundo passo do teste tomam-se os resíduos da regressão acima e executa-se um teste ADF sem intercepto nem tendência. O ADF com aumento na terceira defasagem apresenta resíduos i.i.d. (homocedásticos e sem correlação serial). A análise do ADF produz a estatística $\hat{s}_{\alpha\beta} = -5,94$, que rejeita a hipótese nula de uma raiz unitária em favor de um modelo com transição suave na tendência. Logo, a nossa suspeita quanto ao comportamento da série era válida.

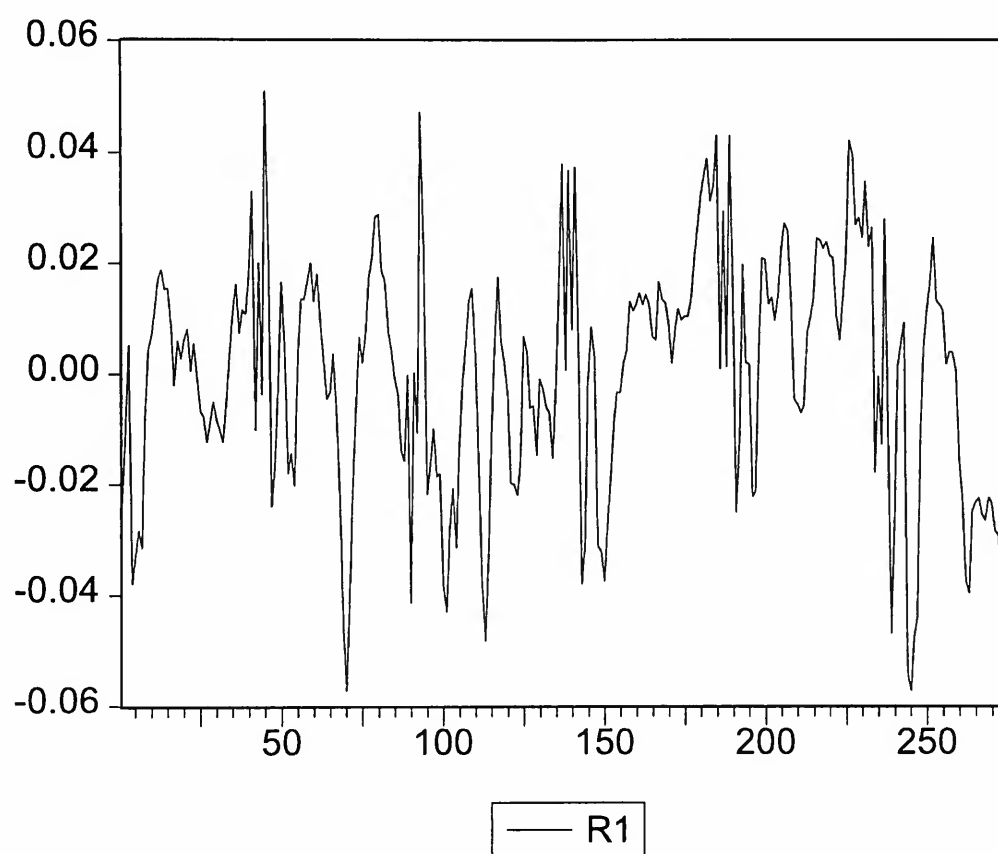
A opção entre um modelo com raiz unitária e um com tendência é importante porque enquanto no primeiro temos um passeio aleatório ao redor de uma constante ou uma tendência determinista (dependendo da significância do intercepto), no segundo, temos flutuações estacionárias ao redor de uma tendência determinista. Logo, a segunda alternativa é muito mais interessante para o governo, uma vez que o nível de atividade medido pelo IMEC apresenta flutuações em torno de uma tendência determinista, as quais não apresentam persistência. Se a alternativa de raiz unitária fosse a relevante para o nosso problema, teríamos que a trajetória do nível de atividade não seria previsível.

Resumindo nossa discussão até aqui, toda a evidência apresentada parece apontar para a existência de uma quebra na tendência e na constante para a série do IMEC. O resultado é útil porque uma série que aparentemente era integrada de ordem 1 passa a se mostrar estacionária com uma tendência determinista. A mudança no rumo da série, determinada endogenamente, ocorre no Plano Real, que colocou fim à estagnação e passou a introduzir uma tendência de crescimento na economia.

12 A moeda chamada Real foi introduzida a partir de 1º de julho de 1994, após um período de transição desde que o Plano Real foi proposto em dezembro de 1993. Uma exposição sobre o início do Plano Real pode ser encontrada em Baer (1996).

Para completar nossa análise, vamos examinar o que a nossa série tem a dizer sobre a sazonalidade intra-mês. Como vimos, análises utilizando o ferramental econométrico mais comum apontam para a conclusão de que não há nenhuma raiz unitária sazonal e que não há nenhum padrão sazonal determinista típico da série. Mas, uma vez que o modelo mais adequado para analisar a série é aquele com transição suave, necessitamos analisar o resíduo daquela equação (que está completamente despojado de componentes de longo prazo). Os resíduos dessa regressão estão na Figura 2.

Figura 2 – Resíduos da Regressão com Transição Suave



Em primeiro lugar, tomamos os resíduos da equação com transição suave e fazemos o teste HEGY para verificar se há sazonalidade na série. Os resultados estão sumariados na Tabela 2, com o modelo aumentado nas defasagens 1, 2, 3 e 4. Os resíduos são independentes e identicamente distribuídos.

Tabela 2
Testes HEGY para os Resíduos do Modelo com Transição Suave

MODELO	\hat{t}_1	\hat{t}_2	\hat{t}_4	\hat{F}_{34}	\hat{F}_{234}	\hat{F}_{1234}
1111	-4,27*	-5,33*	-10,01*	76,24*	83,14*	68,45*
1110	-4,29*	-5,36*	-10,07*	77,12*	84,14**	69,29**
1100	-4,33*	-5,38*	-10,08*	77,38*	84,52**	69,69**

Legenda: * 1%; ** 5%; embora os testes F estejam todos rejeitando claramente a hipótese nula a 1%, colocamos para alguns deles o nível de 5% porque é a tabulação disponível mais próxima. Modelo *abcd*: *a* para constante, *b* para *dummies* sazonais, *c* para tendência e *d* para *dummies* na tendência; cada um deles assume valor 1 se a variável a que se refere está presente na equação estimada, e 0 se está ausente.

Os resultados apontam para a inexistência de qualquer raiz unitária na série (inclusive sazonal). No entanto, a regressão estimada no modelo com constante e *dummies* sazonais aponta para uma constante significativa e a significância da *dummy* na primeira quadrissemana, com um valor positivo, e uma *dummy* negativa na segunda quadrissemana beirando a significância. A terceira *dummy* não foi significativa. Os resultados querem dizer que a resposta nas duas últimas quadrissemanas do mês é igual em termos estatísticos. Já a primeira quadrissemana apresenta um nível de atividade bem mais alto, e o da segunda é um pouco mais baixo (ou igual, se o critério de significância considerado for mais rigoroso).

A estimação de modelos para os resíduos serve para corroborar os fatos observados. Estimamos dois modelos alternativos. O primeiro deles é um ARMA(4,2), em que o termo MA da primeira defasagem é igual a zero. Temos

$$\begin{aligned} \text{RES} = & 0,001992\text{D1} - 0,001827\text{D2} + 0,719474\text{RES}(-1) - 0,646659\text{RES}(-2) + \\ & (6,78) \quad (-5,80) \quad (12,09) \quad (-8,98) \\ & 0,241363\text{RES}(-3) + 0,221192\text{RES}(-4) + 0,981675\text{MA}(2) \\ & (3,35) \quad (3,74) \quad (113,62) \end{aligned}$$

onde AIC=-5,813444 e SIC=-5,720647 Os resíduos do modelo são i.i.d. Os resultados mostram que as *dummies* são significativas e apresentam os sinais mostrados pelo teste HEGY. É importante também observar que o coeficiente estimado do termo MA(2) é bastante próximo da unidade.

É necessário verificar se a série pode ser descrita por um modelo que leva em conta o momento em que são feitas as decisões, uma especialização do AR tradicional. Nele a

resposta de uma variável pode ser diferenciada dependendo do período (quadrissemana em nosso caso) em que se encontra. Para isso precisamos testar se há indícios de autocorrelação periódica e de heterocedasticidade sazonal. A estratégia é a seguinte: escolhe-se o melhor modelo AR de acordo com um critério (por exemplo, o AIC). Toma-se o resíduo e faz-se uma regressão avaliando o papel do componente periódico da forma descrita em Franses (1996, p. 102). Em nosso caso, o melhor modelo foi o AR(10) com constante e *dummies* sazonais, que apresenta AIC=-5,675703 e SIC=-5,487098.¹³ Os resíduos não apresentam correlação serial de 1 a 4 defasagens, não há heterocedasticidade e os resíduos são normais. Obtivemos as estatísticas: $F_{PcAR,1-1} = 4,57$, $F_{PcAR,1-2} = 4,03$, $F_{PcAR,1-3} = 6,77$ e $F_{PcAR,1-4} = 5,56$. Todas elas são superiores aos valores críticos, rejeitando a hipótese nula de ausência de autocorrelação periódica. O teste de heterocedasticidade sazonal envolvendo os resíduos ao quadrado do modelo AR(10), a constante e as variáveis *dummy* produz $F_{SH} = 1,83$, o que não rejeita a hipótese nula de ausência de heterocedasticidade sazonal. Logo, os resultados apontam que deve haver espaço para se explorar os modelos periódicos neste caso por causa da presença de autocorrelação periódica.

O segundo modelo estimado é, de acordo com a discussão acima, do tipo auto-regressivo periódico (PAR), como descrito, por exemplo, em Franses (1996). A ordem de seleção é escolhida por meio de um teste tipo F de significância conjunta das quatro defasagens de uma variável. Partimos de 12 defasagens e fomos diminuindo o número de defasagens. A ordem do PAR deveria ser 9 ou 10.¹⁴ Escolhemos o PAR(10) porque seus resíduos se comportam de acordo com uma distribuição normal, o que faz com que seus estimadores sejam os de máxima verossimilhança. O que parece produzir os melhores resultados é um PAR(10) com buracos dado por

$$\begin{aligned}
 R = & 0,003350D1 - 0,002797D2 + 0,001128D3 - 0,000674D4 + 0,372570D1R(-1) \\
 & (2,39) \quad (-1,80) \quad (0,80) \quad (-0,45) \quad (2,74) \\
 & + 0,694915D2R(-1) + 0,722554D3R(-1) + 1,159165D4R(-1) \\
 & (7,97) \quad (7,70) \quad (10,56)
 \end{aligned}$$

13 O resultado do ARMA estimado acima é superior em termos do critério de informação. No entanto, tivemos que fazer o teste da forma proposta pelo autor, ou seja, com o melhor AR. Os coeficientes do modelo são mostrados em uma tabela a seguir, comparando os resultados com um modelo PAR a ser estudado. Também no AR temos as duas primeiras *dummies* sazonais significativas, com os sinais vistos acima.

14 O resultado poderia ser 9 ou 10 porque o teste F do PAR(10) beira os 10% de significância.

$$\begin{aligned}
&+0,315312(D1R(-2)+D2R(-2)+D3R(-2))+0,748191D4R(-2)-0,221433D1R1(-3) \\
&\quad (4,56) \qquad \qquad \qquad (3,54) \qquad \qquad \qquad (-2,30) \\
&-1,062095D2R1(-3)-0,220608D3R1(-3)+0,482960D1R1(-4)-0,324463D2R1(-4) \\
&\quad (-8,72) \qquad \qquad (-1,87) \qquad \qquad (4,79) \qquad \qquad (-1,68) \\
&-0,900908D3R1(-4)-0,458185D4R1(-4)+0,184025D1R1(-5)+1,246002D2R1(-5) \\
&\quad (-7,01) \qquad \qquad (-2,77) \qquad \qquad (1,51) \qquad \qquad (9,42) \\
&+0,738036D3R1(-5)-0,468184D1R1(-6)+0,234936D3R1(-6)-0,466051D1R1(-7) \\
&\quad (6,22) \qquad \qquad (-4,72) \qquad \qquad (2,39) \qquad \qquad (-6,89) \\
&+0,393934D1R1(-8)-0,296007D2R1(-8)-0,459279D3R1(-8)+0,430525D1R1(-9) \\
&\quad (4,66) \qquad \qquad (-2,24) \qquad \qquad (-3,60) \qquad \qquad (3,26) \\
&+0,737828D2R1(-9)+0,181276D3R1(-9)-0,285982D1R1(-10) \\
&\quad (4,76) \qquad \qquad (2,27) \qquad \qquad (-3,46)
\end{aligned}$$

com $AIC=-6,133512$ e $SIC=-5,729358$ e resíduos normais sem correlação serial e sem heterocedasticidade. Pelos dois critérios este modelo é superior ao anterior. Os quatro primeiros coeficientes estimados mostram que a *dummy* na primeira quadrissemana é significativa e a da segunda só se torna significativa a 10%. As demais são irrelevantes. Os sinais são iguais aos mostrados anteriormente (positivo para D1 e negativo para D2). Logo, o padrão sazonal determinista é o mesmo em todas as equações estimadas. Quanto aos demais parâmetros, vemos que a resposta em uma dada defasagem depende da quadrissemana em que ocorre. Por exemplo, nas duas primeiras defasagens vemos que a resposta da quarta quadrissemana é maior que a das demais. Logo, há campo para se explicar a sazonalidade não só em sua parte determinista quanto também em sua resposta periódica.

Um melhor entendimento do modelo PAR(10) comparado com o AR(10) anterior é feito na Tabela 3, onde são mostrados os coeficientes estimados. Nessa tabela, o PAR é desmembrado em quatro modelos, um para cada quadrissemana, o que torna o entendimento do modelo mais fácil.

Tabela 3
Comparação dos Resultados do AR(10) e do PAR(10) com Buracos

Coeficiente	R	D1R	D2R	D3R	D4R
C	0,000350 (0,19)	0,003350 (2,39)	-0,002797 (-1,80)	0,001128 (0,80)	-0,000674 (-0,45)
D1	0,006487 (2,44)	-----	-----	-----	-----
D2	-0,005946 (-2,26)	-----	-----	-----	-----
D3	-0,001799 (-0,68)	-----	-----	-----	-----
R(-1)	0,734200 (11,72)	1,159165 (10,56)	0,372570 (2,74)	0,694915 (7,97)	0,722554 (7,70)
R(-2)	0,183486 (2,37)	0,315312 (4,56)	0,748191 (3,54)	0,315312 (4,56)	0,315312 (4,56)
R(-3)	-0,323955 (-4,16)	-1,062095 (-8,72)	-0,220608 (-1,87)	-----	-0,221433 (-2,30)
R(-4)	-0,016146 (-0,20)	0,482960 (4,79)	-0,324463 (-1,68)	-0,900908 (-7,01)	-0,458185 (-2,77)
R(-5)	0,412790 (5,29)	-----	0,184025 (1,51)	1,246002 (9,42)	0,738036 (6,22)
R(-6)	-0,139343 (-1,78)	0,234936 (2,39)	-----	-0,468184 (-4,72)	-----
R(-7)	-0,277504 (-3,57)	-0,466051 (-6,89)	-----	-----	-0,466051 (-6,89)
R(-8)	0,129914 (1,70)	0,393934 (4,66)	-0,296007 (-2,24)	-0,459279 (-3,60)	-----
R(-9)	0,141748 (1,86)	-----	0,430525 (3,26)	0,737828 (4,76)	0,181276 (2,27)
R(-10)	-0,115209 (-1,87)	-----	-0,285982 (-3,46)	-0,285982 (-3,46)	-----

Nesta tabela estão os coeficientes do AR(10) e do PAR(10) com buracos. Foram eliminados os coeficientes não significativos que produziam ganhos no AIC. Os coeficientes dentro de um mês que não se mostraram significativamente diferentes por um teste F tiveram sua igualdade imposta na estimação. Os resíduos são n.i.d.

Alguns pontos podem ser destacados na Tabela 3. Vemos que os coeficientes dos modelos realmente variam bastante. Por exemplo, enquanto no AR(10) o coeficiente do termo R(-1) é 0,73, o valor correspondente aos dados da primeira quadrissemana é superior à unidade, enquanto os das demais é bem inferior. Logo, o modelo AR mascara o resultado individual para cada quadrissemana. Ademais, dependendo da defasagem, até os sinais variam, sendo o sinal da primeira quadrissemana diferente dos demais (veja as defasagens 4, 6 e 8). Por esses resultados vemos que o IMEC em uma data comparado com o da mesma quadrissemana do mês anterior leva a resultados diversos por parte dos agentes. Enquanto um resultado bom na primeira quadrissemana leva a uma perspectiva favorável na primeira quadrissemana dos dois próximos meses, temos que um resultado desfavorável em outra leva a uma perspectiva de melhora no dois meses seguintes (veja os coeficientes estimados

das defasagens 4 e 8). Os resultados das *dummies* sazonais são semelhantes nos dois modelos, apontando para um aquecimento da economia na primeira quadrissemana e uma diminuição na segunda relativamente às demais.¹⁵

4 Conclusões

Nosso estudo é pioneiro ao analisar dados intra-mês e procurar diversos padrões sazonais e periódicos em uma série econômica. Por meio de nossa análise conseguimos chegar a algumas conclusões. A série do logaritmo do IMEC não apresenta tendência estocástica, e sim uma tendência determinista apontando para uma quebra estrutural com transição suave. Somente por meio da descoberta deste fato é que tivemos condições de encontrar um padrão sazonal típico determinista intra-mês, em que a primeira quadrissemana apresenta claramente um aumento do nível de atividade relativamente às demais. A segunda quadrissemana pode apresentar uma diminuição do nível de atividade relativamente às demais, mas o resultado não é robusto, como pudemos perceber pelos dois modelos acima. A resposta da série dependendo da quadrissemana em que ocorre a defasagem é diferenciada. Não há nenhum indício de raízes unitárias sazonais, o que significa que não há padrões sazonais mutantes ao longo do tempo. Há, no entanto, uma resposta periódica diferenciada, que adequadamente tratada conduz a modelos superiores aos modelos tradicionais.

As causas verdadeiras do aumento do nível de atividade na primeira quadrissemana não podem ser verificadas, mas apenas conjecturadas. Pode ser, por exemplo, a conseqüência de pagamentos de salários. Há um aumento da oferta de moeda para acomodar os pagamentos, o que provoca a reação de curto prazo aumentando o nível de atividade. Se o Banco Central se preocupar no restante do mês em contrair a oferta de moeda, então o

15 Poder-se-ia postular uma análise utilizando modelos como os descritos em Harvey (1989), por exemplo. Há algumas observações a serem feitas nesse sentido. Em primeiro lugar, a detecção de raízes unitárias sazonais em um Modelo Básico Estrutural, descrito em Harvey, só será adequadamente feita se utilizarmos os aperfeiçoamentos contidos em Kawasaki & Franses (1996). Julgamos que, em face dos resultados sem ambigüidade obtidos anteriormente e do trabalho envolvido na análise e descrição dos novos resultados, o acréscimo de conhecimento seria mínimo e por isso não perseguimos tal alternativa. O segundo ponto, mais conceitual, é o de que os modelos do tipo descrito em Harvey pressupõem que seja possível separar os diversos componentes da série (tendência, sazonalidade, ciclos) e que eles não possuam interação entre si. Isto vai contra a evidência empírica descrita em Franses (1996, cap. 6), que aponta a possibilidade de interação entre os componentes. Por isso, a utilização de modelos PAR é mais geral, uma vez que eles permitem que aconteça essa interação sob determinadas condições. Para detalhes, veja Franses (1996).

comportamento do nível de atividade pode ser explicado pela oferta de moeda e sua relação de curto prazo com o nível de atividade.

O comportamento periódico dos agentes pode ser o resultado das preferências dos agentes, ou do modo particular como ele enxerga os fatos e processa as informações. Daí a razão de um resultado favorável em uma data provocar reações diversas nos dois próximos meses dependendo da quadrissemana em que ocorre.

Referências bibliográficas

- Aguirre, Antônio. The applied perspective for seasonal cointegration testing: a supplementary note. *Economia Aplicada*, v. 2, n. 4, p. 743-55, out.-dez. 1998.
- Baer, Werner. *A economia brasileira*. São Paulo: Nobel, 1996.
- Boswijk, H. Peter & Franses, Philip Hans. Testing for periodic integration. *Economics Letters* 48, p. 241-8, 1995.
- Davidson, Russel & MacKinnon, James G. *Estimation and inference in econometrics*. New York: Oxford University Press, 1993.
- Dickey, David A. & Fuller, Wayne A. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, v. 49, n. 4, p. 1057-72, 1981.
- Diniz, Eliezer Martins. *Oferta de moeda e preços no Brasil 1953-1985*. São Paulo: FEA-USP, 1996. (Tese de Doutorado)
- Franses, Philip Hans. *Periodicity and stochastic trends in economic time series*. Oxford: Oxford University Press, 1996.
- Fuller, Wayne A. *Introduction to statistical time series*. New York: Wiley, 1976.
- Ghysels, Eric, LEE, Hahn S. & Noh, Jaesum. Testing for unit roots in seasonal time series. *Journal of Econometrics*, 62, p. 415-42, 1994.
- Granger, C. W. J. & Teräsvirta, T. *Modelling nonlinear economic relationships*. Oxford: Oxford University Press, 1993.
- Hamilton, James D. *Time series analysis*. Princeton: Princeton University Press, 1994.

- Harvey, Andrew C. *Forecasting, structural time series models and the Kalman filter*. Cambridge: Cambridge University Press, 1989.
- Holden, Darryl & Perman, Roger. Unit roots and cointegration for the economist. *In*: Rao, B. Bhaskara (ed.), *Cointegration for the applied economist*. New York: St. Martin's, 1994, p. 47-112.
- Hylleberg, S., Engle, R. F., Granger, C. W. J. & Yoo, B. S. Seasonal integration and cointegration. *Journal of Econometrics*, 44, p. 215-38, 1990.
- Kawasaki, Y. & Franses, P. H. *A model selection approach to detect seasonal unit roots*. Econometric Institute Report EI-9670/A. Rotterdam: Erasmus University, 1996.
- Leibourne, Stephen, Newbold, Paul & Vougas, Dimitrios. Unit roots and smooth transitions. *Journal of Time Series Analysis*, v. 19, n. 1, p. 83-97. 1998.
- Oliveira, André Luís Rossi de & Pichetti, Paulo. The applied perspective for seasonal cointegration testing. *Economia Aplicada*, v. 1, n. 2, p. 263-79, abr.-jun. 1997
- Smith, Richard J. & Taylor, A. M. Robert. Additional critical values and asymptotic representations for seasonal unit root tests. *Journal of Econometrics*, 85, p. 269-88, 1998.

My life philosophy: policy credos and working ways^{*}

Paul A. Samuelson[§]

Ethics

Many economists - ALFRED Marshall, Knut Wicksell, Léon Walras, ... - became economists, they tell us, to do good for the world. I became an economist quite by chance, primarily because the analysis was so interesting and easy - indeed so easy that at first I thought that there must be more to it than I was recognizing, else why were my older classmates making such heavy weather over supply and demand? (How could an increased demand for wool help but lower the price of pork and beef?)

Although positivistic analysis of what the actual world is like commands and constrains my every more as an economist, there is never far from my consciousness a concern for the ethics of the outcome. Mine is a simple ideology that favors the underdog and (other things equal) abhors inequality.

I take no credit for this moral stance. My parents were “liberals” (in the American sense so the word, not in the European “Manchester School” sense), and I was conditioned in that general *Weltanschauung*. It is an easy faith to adhere to. When my income came to rise above the median, no guilt attached to that. Nor was there a compulsion to give away all my extra coats to shirtsleeved strangers: my parents would have thought me daft to do so, and neurotic to toss at night for not having done so. Some personal obligation for distributive justice liberals do expect of themselves: but what is far more important than acts of private clarity is to weight the counterclaims of efficiency and equity, whenever public policy is concerned, in the direction of equity. As my University of Chicago teacher and friend Henry Simons used to say, “*Any good cause is worth incurring some costs for. Everything should be pushed beyond the point of diminishing returns (else, why desist from pushing it still further?).*”

* Este artigo está sendo publicado com a autorização do próprio autor e da Cambridge University Press. Originalmente, esse texto, *My life philosophy: policy credos and working ways*, foi publicado em Michael Szemberg ed., *Eminent Economists. Their Life Philosophies*. Cambridge University Press, 1992.

§ Paul Samuelson is Institute Professor, Emeritus, Massachusetts Institute of Technology, and 1970 winner of the Nobel Memorial Prize in economic science.

Persons who will not volunteer to serve in the army can with good logic vote to pass a fair conscription law that will entail their being drafted with the same positive probability as any other persons. I have generally voted against my own economic interests when questions of redistribute taxation have come up. The fact that I have favored closing tax loopholes has not precluded seeking some advantage from those left in the tax code. But too avid an effort in that direction would seem not only unaesthetic but also a source of some discomfort and self-reproach.

Without exception all the economists I know regard themselves as humanitarians. This includes communists who toe the Stalinist line and Chicago-school zealots for laissez-faire. Yet we all pretty much know what to expect of each other when it comes to policy recommendations and judgments. It is not unanimity. If political economy were an exact, hard science, then more agreement on probable outcomes would occur. If economics were no science at all, only a tissue of value judgments and prejudices, then soliciting an opinion from an economist would tell the Prince or Parliament nothing about the merits or demerits of the proposal under deliberation but only give a reconfirmation that Economist Jones is a bleeding-heart liberal and Economist Smith a selfish elitist.

Political economy as we know it falls in-between. Economists do agree on much in any situation. Where Milton Friedman and I disagree, we are quick to be able to identify the source and texture of our disagreements in a way that non-economists cannot perceive. The disparity of our recommendations is not an unbiased estimator of the dispersion of our inductive and deductive beliefs. With my social welfare function (or, in Waldian statisticians' terminology, my "loss function") concerning the relative importance of unemployment bad business freedoms, I could disagree 180° with his policy conclusion and yet concur in diagnosis of the empirical observations and inferred probabilities. Yet such is the imperfection of the human scientist, an anthropologist studying us academic guinea pigs will record the sad fact that our hearts do often contaminate our minds and eyes. The conservative will forecast high inflation danger on the basis of the same data that lead the do-gooder to warn against recession. (Conscious of this unconscious source of bias, as the subsequent discussion will elaborate on, I make a special effort toward self-criticism and eclecticism - with what success, the record must testify to.)

An economist who has been preoccupied over the years solely with *Pareto optimality* wrote me long ago that I would be surprised to know how liberal he is. Indeed I would be. Reflecting on his writings, I wondered how he knew he had a heart: it had been so long since he had used it. Organs atrophy without exercise. "Use it or lose it" is nature's law.

It is not only the arteries that harden with age. Economists are said to appear to grow more conservative as they rise in seniority. This they often deny.

In my own case, I do not perceive that my value-judgment ideology has changed systematically since the age of 25. For a decade now mainstream economics has been moving a bit rightward. But I have not been tempted to chase it. What does tend to change with the accumulation of years and experience is one's degree of optimism about what is feasible and one's faith in good intentions alone. My enhanced skepticism about government ownership of the means of production or the efficacy of planning is not a reflection of ossifying sympathies and benevolence, but rather is a response to the testimony of proliferating real-world experiences.

I am conscious of one occasion in which my respect for the market mechanism took a quantum leap upward. This change had nothing to do with improved performing of the market system. Nor was it related to any new arguments brought forward by Hayek about generating and utilizing information, or to any old arguments about market efficiencies and freedoms by Adam Smith, Frederic Bastiat, or Frank Knight. Rather my changed viewpoint came from observing the communist witchhunting episode of the 1950s.

The McCarthy era, in my judgment, posed a serious threat of American fascism. I knew plenty of people in government and the universities whose civil liberties and careers came into jeopardy. I observed at close hand the fears and tremblings that the Harvard and MIT authorities experienced, and these were the bolders of the American academic institutions. As Wellington said of Waterloo, it was a close-run thing that Senator McCarthy was discredited: the Richard Nixon "enemy list" was a joke in comparison, and my being named on it only added to my fading credentials as a New Dealer. What I learned from the McCarthy incident was the perils of a one-employer society. When you are blackballed from government employment, there is great safety from the existence of thousands of anonymous employers out there in the market. I knew of people who got some kind of work in private industry, usually smaller industry since large firms tend to try to keep on the safe side of government. To me this became a newly perceived argument, not so much for laissez-faire capitalism as for the mixed economy.

How did free-market advocates among the economists score as defenders of personal freedoms and civil liberties? This was a subject of great interest to me and over several years I kept a quiet tally of the behavior and private utterances of scores of the leading American and Continental libertarians, almost all of whom I knew intimately. Like a visiting anthropologist I would ask innocent questions designed to elicit relaxed and spontaneous

views. If it was churlish to keep a record of private conversations, then I was a churl. The results surprised and distressed me. Worshippers of laissez-faire à la Bastiat and Spencer were insensitive and on the whole unsympathetic toward the rights and personal freedoms of scholars. Alone among the members of the Mt. Pelerin Society the name of Fritz Machlup stood out as one willing to incur personal costs to speak up for John Stuart Mill values. It is not the failure of people to be heroes that I am speaking about. There is little of the heroic in my own makeup and I have learned not to expect much of human nature. What my research found was a sad lack genuine concern for human values.

I was taught at the University of Chicago that business freedoms and personal freedoms have to be strongly linked, as a matter both of brute empirical fact and of cogent deductive syllogism. For a long time I believed what I was taught. Gradually I had to acknowledge that the paradigm could not fit the facts. By most Millian criteria, regimented Scandinavia was freer than my America - or certainly at least as free. When I used to bring up these inconvenient facts to my conservative friend David McCord Wright, he would warn: "*Just you wait. British and Swedish citizens, it is true, have not yet lost their freedoms. But it cannot last that the market is interfered with and people remain politically free.*" We have all waited for more than thirty years now.

Friedrick Hayek wrote his bestseller, *The Road to Serfdom*, at the end of World War II, warning that partial reform was the sure path to total tyranny. Cross-sectional and time-series analysis of the relationship between politics and economics suggest to me important truths.

1. Controlled socialist societies are rarely efficient and virtually never freely democratic. (There is considerable validity then for the non-novel part of Hayek's warning.)
2. Societies which resisted partial reforms have often been those over taken by revolutionary change. If it is the free market or nothing, often it has then had to be nothing. Indeed, after midcentury the finest archetypes of efficient free markets have often been quasi-fascist or outright fascist societies in which a dictatorial leader or single party *imposes* a political order - without which imposition the market could not politically survive. Chile with its military dictatorship cum-the-Chicago boys is only one dramatic case. Taiwan, South Korea, and Singapore are less dramatic but more representative cases.
3. I can nurture a dream. Like Martin Luther King, I have a dream of a humane economy that is at the same time efficient and respecting of personal (if not business) freedoms.

Much of producing and consuming decisions involve use of the market mechanism. But the worst inequalities of condition that result from reliance on market forces - even in the presence of equality of ex ante opportunity - can be mitigated by the transfer powers of the democratic state. Does the enhancement of equity by the welfare state take no toll in terms of efficiency? Yes, there will be some trade-off of enhanced total output against enhanced equality, some trade-off between security and progress. I call the resultant optimizing compromise *economics with a heart*, and it is my dream to keep it also economics with a head.

My methodology

It is some relief to move from the exalted realm of philosophical ethics to the mundane realm of scientific methodology. However, I rather shy away from discussions of Methodology with a capital M. To paraphrase Shaw: Those who can, do science; those who can't prattle about its methodology.

Of course I can't deny that I have a methodology. It's just that there seems little appeal in making it explicit to an outsider. Or for that matter, in spelling it out to my own consciousness.

I am primarily a theorist. But my first and last allegiance is to the facts. When I began study at the University of Chicago, Frank Knight and Aaron Director planted in me the false notion that somehow deduction was more important than induction. This was a confused tenet of Austrian methodology at the time, and I certainly do not mean by the word "Austrian" the logical positivism of the Vienna Circle. Rather, such direct and indirect disciple of Carl Menger as Ludwig von Mises, Friedrich Hayek, and Lionel Robbins seemed to put on their own heads the dunce caps of the classical Ricardians who believed that by thinking in one's study one could arrive at the basic immutable laws of political economy. I remember believing Director when he pooh-poohed Wesley Mitchell's empirical work on business cycles, claiming instead that the greatest breakthroughs in the subject were coming from Hayek's a priorisms on the subject.

I grew out of this phase fast. Once Lionel Robbins explained lucidly in the first edition of his *An Essay on the Nature and Significance of Economic Science* his claims for Kantian a priorism in economics, his case was lost. Logical positivism is now judged to be an oversimplified doctrine, but it was enormously useful in deflating the pretensions of deductionists. If one had to choose between the methodologies of the warring brothers -

Ludwig the economist and Richard von Mises the mathematical physicist - Richard would win hands down.

Let me not be misunderstood. I abhor the sins of scientism. I recognize that, as social scientists, we can have relationships with the data we study that the astronomers cannot have with the data they study. I am aware that my old friend Willard van Orman Quine, one of this age's greatest logicians, has cast doubt that anyone can in every case distinguish between "analytic" a prioriisms and the "synthetic" propositions that positivists take to be empirical facts. Furthermore, Wesley Mitchell's empiricisms on the business cycle do seem to me to have been overrated - not because they are empirical, but rather because his was an eclecticism that never had much luck in discovering anything very interesting, as the lifecycle profile of his post - 1913 career sadly reveals. Some of the skepticisms of Knight and Jacob Viner concerning the empirical statistical studies that their colleagues Paul Douglas and Henry Schultz were attempting, I readily admit, were well taken - just as some of Keynes's corrosive 1939 criticisms of Jan Tinbergen's econometric macrodomes were. But it is on *empirical* grounds that these empirical attempts have to be rejected or accepted, and not because deductive syllogisms can claim a primacy to vulgar fact grabbing. What was wrong with the German Historical School was not that it was historical, but rather that its sampling of the facts was incomplete and incoherent. The facts don't tell their own story. You can't enunciate all the facts. And if you could, the job of the scientists would just begin - to organize those facts into useful and meaningful gestalts, into patterns that are less multifarious than the data themselves and which provide economical *descriptions* of the data that afford tolerably accurate extrapolations and interpolations.

Whatever logical positivism's faults and superficialities are in science at large, it gets an undeservedly bad name in economics from being confused with Milton Friedman's peculiar version of positive economics. Much of what is in Friedman's 1953 essay on this topic is unexceptional and a story so old as to seem almost platitudinous. But what is novel in his formulation and commands most attention is that which I have called "the F twist" - the dictum that a scientific theory is none the worse if its premises are unrealistic (in the usual meaning of "unrealistic" as stating hypotheses that are false and/or far-from-true assertions about what obtains in the actual world), so long as the theory's "predictions" are usefully true. Thought suggests, and experience confirms, that such a dogma will be self-indulging, permitting its practitioners to ignore or play down inconvenient departures of their theories from the observable real world. A hypothesis's full set of predictions includes its own descriptive contents: so, literally understood, an unrealistic hypothesis entails some unrealistic predictions and is all the worse for those false predictions - albeit it is all the better for its (other) empirically correct predictions. We are left then validly with only the

prosaic reminder that few theories have all their consequences exactly correct; and it can be the case that a scientific theory is deemed valuable because we have reason to give great weight to those of its predictions that happen to be true and to give little weight to those that are found to be false. In no case is unrealistic falsity a virtue; and there is danger of self-serving Humpty-Dumptyism in letting the theorist judge for himself which of his errors he is going to extenuate or ignore.

Unpopular these days are the view of Ernest Mach and crude logical positivists, who deem good theories to be merely economical descriptions of the complex facts that tolerably well replicate those already-observed or still-to-be-observed facts. Not for philosophical reasons but purely out of long experience in doing economics that other people will like and that I myself will like, I find myself in the minority who take the Machian view. “Understanding” of classical thermodynamics (the archetype of a successful scientific theory) I find to be the capacity to “describe” how fluids and solids will actually behave under various specifiable condition. When we are able to give a pleasingly satisfactory “HOW” for the way of the world, that gives the only approach to “WHY” that we shall ever attain.

Always when I read new literary and mathematical paradigms, I seek to learn what descriptions they imply for the observable data. The paradigm’s full set of entailed descriptions is what is of interest and forms the basis for a complete judgment on it. My work in revealed preference, in *Foundations of Economic Analysis*, and in the several volumes of *Collected Scientific Papers*, consistently bears out this general methodological procedure.

I dislike being wrong. Long before knowing of Karl Popper’s writings, I sought to be my own strictest critic. Why give that fun to the other chap? All this explains why I am an eclectic economist. It is not because of inability to make up my mind. I am eclectic only because experience has shown that Mother Nature is eclectic. If all the evidence points to a single-factor causation, I have no internal resistance to accepting that. But there is a big “if” involved in the previous sentence.

Being prepared to be eclectic does not have to inhibit bold theory building. One creates boldly knowing that this does not commit one to exaggerated belief in the sole potency of one’s brain child.

We all have secret vanities. He prides himself on his good looks. She takes satisfaction in her sense of humor. I do delight in producing still another beautiful model that illuminates

important terrains of economics. But in my heart of hearts I nurture the claim that I have good judgment. Be wise, sweet maid, and let them who will be clever. My theories must run the gauntlet of my judgment, an ordeal more fearsome than mere peer review. (Of course one can have one's cake and eat it too by presenting a theoretical gem as an unpretentious mirror of some aspects of some corner of the economic terrain under observation.) Why let sagacity degenerate into well-informed nihilism? The mindless naysayer is no better than the mindless yeasayer. Neither adds anything to the silent scientist's cipher.

Joseph Schumpeter, who all his life whored after beautiful theories, just before he died testified at the 1949 National Bureau conference on business cycles: If he had to choose between mastery of mathematics and statistics, or of economic history, he would have to choose mastery of economic history. I won't disagree. But I deny the need for dichotomous choice. Give apes in the Widener Library a data bank of all that's there and you don't get a master economic historian. What you get back is the data bank and a curator.

Let me make a confession. Back when I was 20 I could perceive the great progress that was being made in economic *methods*. Even without foreseeing the onset of the computer age, with its cheapening of calculations, I expected that the new econometrics would enable us to narrow down the uncertainties of our economic theories. We would be able to test and reject false theories. We would be able to infer new good theories.

My confession is that this expectation has not worked out. From several thousands of monthly and quarterly time series, which cover the last few decades or even centuries, it has turned out not to be possible to arrive at a close approximation to indisputable truth. I never ignore econometric studies, but I have learned from sad experience to take them with large grains of salt. It takes one econometric study to calibrate another; a priori thought can't do the job. But it seems objectively to be the case that there does not accumulate a convergent body of econometric findings, convergent on a testable truth.

Does this mean that I belong to the camp which regards truth as in the eye of the beholder? Which denies the existence of an objective truth out there, in political economy as well as in astronomy and biochemistry? Which recognizes in the truth of mainstream economics only the class interests of the bourgeoisie, and in the truth of Marxian economics either the class interest of the nascent proletariat or the objective truth of the final classless and universal society?

No. Observing myself over fifty years and a vast number of scientists in various disciplines, I do recognize that truth has many facets. Precision in deterministic facts or in

their probability laws can at best be only partial and approximate. Which of the objective facts out there are worthy of study and description or explanation depends admittedly on subjective properties of the scientists. Admittedly, a given field of data can be described in terms of alternative patterns of description, particularly by disputing authorities who differ in the error tolerances they display toward different aspects of the data. Admittedly, observations are not merely seen or sensed but rather often are perceived in gestalt patterns that impose themselves on the data and even distort those data.

But still, having admitted all the above, as you observe scientists and study the developments of disciplines when schools evolve and paradigms are born and die, it is forced upon you that *what ultimately shapes the verdicts of the scientist juries is an empirical reality out there*. When a Marxist scores a triumph it is not by employing a useful alternative to $2 + 2 = 4$ logic, or cultivating a different Hegelian dialectic. We esteem a Pavlov, Lysenko, Haldane or Bernal, Landau or Baran for what they can or cannot accomplish with respect to animal experiments, plant breeding, hydrogen-bomb exploding or phase transitions, or insights into the observable paths of economic development.

When Thomas Kuhn's book, *The Structure of Scientific Revolutions*, came out in 1962, I made two lucky predictions: one, that in the physical and life sciences its thesis would have to be modified to recognize that there is a cumulative property of knowledge that makes later paradigms ultimately dominate earlier ones, however differently the struggle may transiently look; two, that Kuhn's doctrine of incommensurability of alternative paradigms would cater to a strong desire on the part of polemical social scientists who will be delighted to be able to say, "*That's all very well in your paradigm, but your white is black in my paradigm and who's to say that we'uns have to agree with you'uns.*" Kuhn has correctly discerned the warts on the countenance of evolving science. His readers must not lose the face for the warts.

How I work

As a theorist I have great advantages. All I need is a pencil (now a ball pen) and an empty pad of paper. There are analysts who sit and look vacantly out the window, but after the age of 20 I was not one of them. I ought to envy the new generation who have grown up with the computer, but I don't. None of them known to me sit idly at the console, improvising and experimenting in the way that a composer does at the piano. That ought to become increasingly possible. But up to now, in my observation, the computer is largely a black box into which researchers feed raw input and out from which they draw various summarizing

measures and simulations. Not having access to look around in the box, the investigator has less intuitive familiarity with the data than used to be the case in the bad old days.

I have been blessed with an abundance of interesting problems to puzzle out. Many artists and writers run into long fallow periods when new creative ideas just will not come. Luckily, that has not been my experience. Perhaps I am insufficiently self-critical to recognize when problems of lower quality are involved. In any case mine has never been the Carlylean view of Schumpeter that only the greatest ideas count, and only a few great men are important in history and in the development of science. One tackles the most important unsolved problem at hand. Then the next one. If that leads down the path of diminishing returns in the absence of dramatic new challenges and breakthroughs, so be it.

“What are you working on now?” This is a question I have been asked all my life. And never in my life have I known how to answer it. At any one time I have several balls in the air. And always there is an inventory of questions just below the threshold of my explicit attention. Some of these slumber in that limbo for two decades. There is no hurry: they will keep. Some morning (or at night in the dream) the evolving wheel of chance will turn their number up.

Poets testify that often their lines gush up from within. They merely write down what their muse is dictating. That sounds rather highfalutin, but there is something in it. When I was young I used to explore a topic; write down equations and syllogisms dealing with different aspects of it; then outline the final work. After that the final draft could be written out. Perhaps what I am describing is the optimal way to write a paper.

Increasingly after the age of 35 that is not how I have in fact operated. Instead I have often let the paper write itself. A problem is posed. One begins to solve it, writing out the steps in the solution, One development leads naturally to another, as one exposit in writing. Finally, what can be solved of the problem has been solved. The paper is finished. What has been finished is not something that has ever been envisaged, waiting only to be written down. All this is reminiscent of Franklin Roosevelt’s dictum, “*How do I know what I think until I hear myself saying it?*”

This means that some articles might be composed in half a day. Of course the first draft need not be the final draft. There may follow many hours of revising, involving additions, deletions, rearrangements, and corrections. Perhaps it would be better to follow the first draft with a completely new rewrite. But that is not my usual practice, as I trade some perfection against more time for new topics. This means I am a prisoner of my first drafts, and it is a source of exquisite pain if a manuscript is lost: my mind rebels at having to

reconstruct a lost argument, and impatience is likely to make a recollected version abridge some essential matter.

Prolific scholars are addicted to writing. A day spent in committee meetings is for me a day lost. After an interval of fasting, you are hungry. After an interval of doing no analytical research, there is so to speak a fluid inside you that wants to get free. I used to think that the unconscious mind, which Henri Poincaré described so beautifully as working away at specific puzzles the mathematician is interested in, was accumulating findings on the particular problems that routine duties prevented me from dealing with. But I have come to think that not to be quite correct. For *any* new topic can capture one's enthusiastic and fruitful attention after a period of deprivation. One snowy day in New England I was told at the airport gate that Washington was snowed in. A friend hearing me inquire, "*Can you go New York?*" asked, "*Are you just bound to go somewhere this day?*" That's exactly what it's like with the creative urge: It doesn't have to spend itself on the theory of capital that has been engaging the scholar's recent attention; it just wants to go about doing something creative, and its motors seem revved up to be effective in whatever direction it is pointed.

Reporters used to speak of a nose for news. What is important in scholarship is an aesthetic sense for what is an important problem. Otherwise the facile mind can spend itself on patterns that are merely pretty. For recreation I would rather play tennis than play chess, or read pedestrian detective stories than solve the mathematical conundrums that appear in the back pages of learned journals. My unconscious motivation, I suspect, is that chess and problems-solving involve the same energies as innovative scholarship does. They will usurp some of the limited supply of precious brainpower that might better go toward learning something new: and, involving use of the same workday muscles so to speak, those recreations do not provide as refreshing rest periods. I daresay that the powerful pure mathematician faces a different problem from the applied scientist. A great mathematician is only as great as his greatest deeds. The revolutionary idea that might lead to great deeds comes very rarely.

One marks time in between and one might as well mark time while keeping the brain tuned up in chess or bridge as in any other way. However, I do not have too much confidence in the distinction that I have just made. For it certainly does not cover the case of prolific mathematicians such as Poincaré or Euler. A mathematical snob like G.H. Hardy might judge that much of Euler and Poincaré could just as well have never been written. But even from the snobbish viewpoint, we must reckon with the fact that some of their best work would not have gotten done if it had not been as outgrowth of some of their less transcendental achievements.

I said that my working tools are only pen and paper, and that an airplane cabin provide as good an environment for research as a library study. That is true as far as analytical creativity is concerned. On the other hand to stay well informed on what it is that is important to be done, a scholar must have access to books and to learned journals. In this regard I have always been very lucky. Whatever works the MIT libraries have not had, the neighboring Harvard libraries can be counted on to provide. These are very few great scholars working of by themselves with paper and pen far from the centers of creative economic thought. Those who pride themselves on being most autonomous usually end up most idiosyncratic.

Long ago I set myself the grandiose challenge of not being merely subjectively original. More useful to science - and more truly fulfilling if you can bring it off - is to try to stay informed on what other scientists have done and to advance the frontier by your own quantum jumps. In terms of the old song: "*Good work if you can get it. And you can get it if you try.*"

Nicholas Georgescu-Roegen about himself*

Nicholas Georgescu-Roegen§

*Two things fill one's conscience with increasing wonder
and awe, the stars in heaven and the moral law in oneself.*

Immanuel Kant

Proem

It would be superfluous to tell the reader how high my spirits were lifted by Michael Szenberg's invitation to write an account of my life philosophy. But as I started to think about the task, a fear came to me, the fear that my acceptance would be taken as an implicit presumption that I am a philosopher like Plato or John Dewey, for example. Even though I have not been a professional philosopher, I certainly "have done philosophy" One does philosophy, I think, not when one reasons about the practical problems of a community or the mathematical structure of quantum mechanics, but only if one treats in a thoroughly free inquiry issues that cannot be rested at a workbench. It is in this sense that I claim to have done philosophy, probably more often than I have exercised any other intellectual expertise. However, although like all who have done philosophy I have asked questions about things, their nature, and their relations among themselves and with the human mind, I have gone further: I have also asked questions about questions.

Everyone's way of thinking is influenced by the events of his or her life. As some have maintained, it was Einstein's experience at the Berne Patent Office that developed his interest in how to test our ideas about things. Not ignoring the considerable difference of proportions, the observation has been especially true to my shattered life. In the country in which I was born and spent the most informative part of my life, Romania, I lived under four dictatorships and three wars, all in my backyard. That history instilled me with a kind of Paretoan view of human societies. Romania was at that time a struggling, overpopulated, peasant dominated culture and economy. And as I came to learn the economics professed in the capitalist world, I was struck by the claims of that discipline that it was a representative

* Este artigo está sendo publicado com a autorização da Cambridge University Press. Originalmente, esse texto, *Nicholas Georgescu-Roegen about himself*, foi publicado em Michael Szenberg ed., *Eminent Economists. Their Life Philosophies*. Cambridge University Press, 1992.

§ Nicholas Georgescu-Roegen is Distinguished Professor of Economics, Emeritus, Vanderbilt University.

guide not only for capitalism but for absolutely all economic conditions. I was evident to me that standard economics could not represent an agrarian economy, and hence could not be a guide for it. I thus acquired a special eye for issues ignored by the standard economic persuasion or by ordinary economic analysis. I learned “philosophy” from many consecrated philosophers, but my own philosophy sprouted from two great teachers of mine: Karl Pearson and Joseph A. Schumpeter. From Karl Pearson’s splendid *Grammar of Science* and from my listening to him for more than one year, I reached two philosophical tenets. First, contrary to the old epistemology, the stochastic form is not the peripheral but our only possible representation of natural laws. By implication I came to hold further that randomness, not just haphazardness, is an essential ingredient of phenomena. Second, given the human cosmological condition, I construed that for us nature consists of just what we can perceive. Beyond, there are only hypothesized abstractions about which a metaphysician may say anything with complete certitude, since nothing is controllable. Our laws of nature aim not at explaining phenomena, but at saving them, as Pierre Duhem calligraphed the philosophy initiated by Ernest Mach and embraced by Wilhem Ostwald and Karl Pearson. Even Albert Einstein’s philosophy was largely Machian: in 1936 he explicitly stated that “physics consists of a logical system of thought [that] can only be arrived at by free inventions. “ Not one word about “reality.” My philosophy it is spirit Machian: it is a particular kind of epistemology that is little concerned with the science of knowledge, or with the cognitive process itself, but mainly with the problem of valid analytical representations of the relations among facts.

What stirred me mainly in this direction was a frequent sin in mathematical economics. Choose any formula used by a physicist and ask him what it represents factually. He may invite you to his laboratory to witness the actual phenomenon described by that formula. However, in economics there is a vast and growing literature of purely mathematical exercises that correspond to absolutely no facts, not even to physical ones. If one starts only with mathematics, one is trapped inside it and cannot even think of the epistemological issues in my own sense.

A statistician turned into an economist

For my statistical dissertation at the Sorbonne, I followed the line of the smallest effort for a mathematician and thought up a mathematical method for discovering the hidden periodicities of time series. Since economic time series were then, and still are, enjoying great consideration, I felt (as I still do) that economic phenomena are not governed by a mathematical network. If the stock exchange market were governed by a mathematical system, I reasoned, no one would have thought of setting one up. I thus applied my method to

the annual rainfall in Paris.¹ After learning later about the three time series used by the famous Harvard Economic Barometer, I became curious to see it applied to them. I was overjoyed when the Rockefeller Foundation gave me an opportunity to visit that organization. But before I reached the United States, it had closed shop. Seeking some scope of activity, I naturally thought of contacting the professor who taught business cycles, whoever he might have been. This is how, quite unintentionally, I met Joseph A. Schumpeter, the man who directly and through his writings was to have an even greater influence on my thinking than Karl Pearson. Every one of his distinctive remarks were seeds that inspired my later works. In this way Schumpeter turned me into an economist - the only true Schumpeterian, I believe. My only degree in economics is from Universitas Schumpeteriana.

Because of my mathematical preparation, I naturally had to plunge first into the writings of Vilfredo Pareto, whom I learned to value as the greatest mathematical economist ever (with great accent on the noun, for Pareto's mathematics were not laudable: my first paper on mathematical economics was on one of his missteps). Being especially concerned with the valid representation of facts analytically, I felt that the neoclassical utility theory needed a fundamental postulate, listed as Postulate A in my 1936 *Quarterly Journal of Economics* essay: on a continuous move from nonpreference to preference with respect to a given basket, we must pass through a place indifferent to that basket. I affirmed for the first time the necessity of postulating the binary indifference (an idea that has been tacitly adopted by many later writers). But some immediately assailed me: "The postulate is totally superfluous. How can you move from nonpreference to preference without passing through indifference?" That objection helped me later move toward dialectics.

At that time an object of great agitation was a paradox created by a criticism of Pareto by Vito Volterra. As is well known, on a second round, Pareto argued that the same map as that based on binary choices could be derived from the reports of a sleuth who followed the individual in a great number of market situations. Clearly, those data serve to establish a total differential equation in the commodities space:

$$\sum a_i(x) dx_i = 0, \quad 1 \leq i \leq n. \quad (1)$$

It was about this point that Volterra committed a gaffe by countering that equation (1) is necessarily integrable if and only if it involves only two variables. The paradox of why Pareto's second method is valid in an economy of only two, but not in one of three,

¹ I should mention that my methods was communicated to the French Académie Sciences by Emile Borel, and the full version of my dissertation filled the whole October 1930 issue of *Journal de la Société de Statistique de Paris*.

commodities was still undefeated. After searching through my box of mathematical tools, I concluded that Volterra and all who after him worked on the issue of integrability completely ignored the real snag. I prided myself on having cut a Gordian knot by proving in the 1936 essay that, contrary to what Volterra and everyone else held, even when the integrals of (1) exist, they cannot be identified with the indifference varieties without some additional *factual* reasons. To explicate this point, I considered the integral curves of (1) on a particular two-dimensional space, namely, the budget plane with three commodities. Under simple conditions, those curves always exist (Volterra's point). Two diagrams proved my contention: in one, the integral curves were logarithmic spirals around a singular point, a focus; in the other, they were ovals around a center, also a singular point. The first corresponded to the nonintegrability case in three dimensions; the second, to the standard utility map. In the first case, *even though the integral curves were there, no preference scale can be constructed on their basis*. Fourteen years later (1954), on these two pictures Paul Samuelson based a parable, highly pleasing like all others of his pen.²

I further observed that, whether a utility map exists or not, the consumer always tends toward a point of saturation, either absolute or relative to the possibilities of the budget. The budget equilibrium, therefore, is always a point of saturation. Next, I pointed out that, whether or not equation (1) is integrable and whether or not saturation is at infinity, the direction from any point toward the saturation point is always one of preference. On this basis, in place of the principle of decreasing marginal utility or of decreasing marginal substitution I proposed the *principle of perseverance of nonpreference directions*; that is, once a direction becomes one of nonpreference, no good can come from persevering in it.³

Because of my tenet that random is an essential element of phenomena, I also initiated the idea of stochastic choice. That study led to several novel by-products, the most salient being that indifference is not necessarily transitive.

Romanian "exile"

If my literary activity during the two short Harvard years (1934-36) seemed rather unusual, it was probably because, as Schumpeter once said, being a novice I was "able to

2 It does not seem at all strange to me that Samuelson was prompted to reread carefully my 1936 essay by an article of H. Houthakker in which independently of my paper, Houthakker argued that no indifference curves may have a spiral form.

3 I expressed the perseverance principle by an inequality on which before long, Paul Samuelson based his epochal idea of revealed preference.

see aspects that trained economists refuse to see and unable to see others that they took for granted.” Be this as it may, that performance together with what I had published earlier must have been responsible for Schumpeter’s intention to write a treatise on theoretical economics with me, which in rum led to an offer to join the economics faculty. I am now unable to say exactly why, but I simply turned my back on the fantastic chance of being a co-author with Schumpeter and becoming “Georgescu-Roegen of Harvard” I left for Romania.

I looked forward to helping my native land become a happier place for all. The Parcae, however, had decided differently; for being an economist and a statistician as well, I was given the undistinguished, tedious task of organizing the economic statistics at the Central Statistical Institute, followed by that of colligating the daily statistics of foreign trade. (At the time even orthodox countries had foreign trade clearings.) But a truly great wringer was lying in wait for me at the end of the war, the arduous job as general secretary of the Armistice Commission. For about six months it meant long, tedious, and stressful discussions, often lasting the whole night, with the representatives of the USSR Control Commission.

Since the Yalta and Potsdam conferences ultimately shattered all my hopes of seeing the world reorganized on the principles for which Great Britain and the United States had entered the war, I had to flee Romania before I was thrown into a jail from which no one has ever come out alive. According to the Communists’ precepts, I was indeed guilty of three capital crimes: (1) being the servant of capitalists as a Rockefeller fellow and then president of the Romanian Association for Friendship with the United States, (2) being a “member” of the National Council of the Peasantist Party, and (3) being an ardent defender of Romania’s rights as the secretary general of the Armistice Commission. Since the safest way of escape seemed to be stowing away on freighter, together with my wife I stealthy entered the Constantza harbor in the middle of the night of February 13, 1948, surrounded by bribed smugglers. I then felt as if the past twelve years in Romania were scooped out of my life.

Before too long, my Harvard friends Edward S. Mason, and especially Wassily Leontief, found a means of bringing me back again. I arrived at Harvard in early July 1948 and what I found boggled my mind for days, for truly great avenues had been opened in economics during my “exile”

A disenchantfed neoclassical economist

Soon after my return to Romania in 1936, I entered into a wonderful friendship with Andrew Edson, the secretary Of the U.S. legation in Bucharest. One day Andy said softly: "Romania is economically underdeveloped because your institutions are silly. The legion of doormen who just sit at the door of every high functionary, public or private, produce nothing to motivate their pay." Andy, a strong believer in the neoclassical dogma, then opened my eyes to a violation in my own backyard of the sacrosanct neoclassical principle of marginal pricing. That icy shower on my religious confidence in mathematical economics started me worrying and thinking and thinking. The solution, when it hit me, was that marginal pricing does not maximize the national product proper - an idea that would undoubtedly strike a standard economist as a ridiculous product of some economic ignoramus. Yet the fact is that only in the lands of plenty does the marginal principle maximize a complex of product proper *and* chosen leisure. In the lands of scarcity, however, people must work as long as they can, to the point of zero marginal productivity of labor, as illustrated by the splendid institution, not too old, of the gleaners. In conditions of scarcity, income distribution is made not according to marginal pricing, but according to some institutional rules (as within most families, yours too, I think). Even in the advanced countries, we should note, the consumer is not guided only by a quantitative set of commodities as standard economics claims. Individual behavior is also affected by how one can realize one's want work for a dollar, beg for it, or pick the cash register, actions judged according to the corresponding social matrix, not affiliated with a quantitative scale.

Another bestirring lesson also came from Romania. Because the peasants always looked to the townies for what to do, the Communists wanted to bring the urban masses to their knees. The plan was to provoke a runaway money inflation so that the decreased value of the bills would stop the peasants from bringing food to the towns. From what I then knew of standard economics, I judged the plan flawless. I cannot describe my surprise when, at a meeting of the National Council of the Peasantist Party, another member, a former village school teacher and an old Peasantist, smugly shot at me, "You do not know the peasants, my friend; they will still sell for any money because money has always been *summum bonum* for them." Nor can I describe my public shame when the developments proved him correct. The Communists then resorted to an unparalleled trick in history: they declared that, on August 15, 1947 all old money was no longer legal tender. Each person (supposedly) received new money of about one U.S. dollar in exchange value.

I had never felt any attraction for monetary theory, and the Romanian peasants convinced me to steer away from its unthinkable quicksands. Yet later, rubbing myself against the

disturbing facts of the monetary conditions of Brazil, I reached another heretical conclusion to which I still cling firmly. Contrary to the general tenet of the professional establishment that inflation is the best strategy for economic development, inflation is the most perverse way of governing. Another invisible hand, a Keynesian this time, picks the pockets of the masses who cannot borrow at a privileged interest rate now and pay later.

Epistemology of economics

Science without epistemology is in so far that it is thinkable at all primitive and muddled.

Albert Einstein

Shortly after my return to the United States I completed several papers consisting of significant results. I am saying this not to boast but to illustrate the special usefulness of epistemology in general.

Why include only one structural component in the analytical representation of a process when we know that there are many?

This was the epistemological question that struck me first when approaching Leontief's system. I then proceeded to find out what would happen if each industry could choose from a set of recipes satisfying Leontief's basic assumptions: labor is the only primary factor of production, and return to a scales is constant. The theorem that I presented at a seminar of the Harvard Economic Project (March 22, 1949) became known as the "substitution theorem" It states that in equilibrium each industry must operate with only one particular recipe out of its own technological horizon. As I showed by the transparent diagrams reproduced in a paper presented at the December 1949 meeting of the American Economic Association, there are also some singular cases in which, for the equilibrium recipe, the ratio of labor to output is in the limit zero, the catalytic labor that represents the futurists dream of unlimited technological progress.

An analytical representation of business cycle by nonsymmetrical waves

Trough discussions with Professor Schumpeter about my idea that business cycles are not "cyclical" another epistemological began pressing my mind: how can we represent them analytically? My answer was presented at another seminar (April 1949) in the paper "Relaxation Phenomena in Linear Economic Models" In it I debunked the idea that decumulation is the

reverse process of accumulation and proposed that the business turning points are relaxation phenomena when the law of one phase suddenly changes into the other. To represent this conception of business cycles, I used a sequence of two alternating phases", F_1 , for the upswing, F_2 for the downswing. Schumpeter, who always attended those seminars, left me breathless when he asked with which of the two phases the process began. He used this sort of question against the theories that explain depression as the product of prosperity and prosperity as that of depression. One must then know, Schumpeter used to say, whether business cycles began with an overproduction or an underproduction of apples in the Garden of Eden.

Vicit Pareto

Because Paul Samuelson believed that a Houthakker axiom that assumed away spiral formations provided the definitive liquidation of the nonintegrability puzzle, he keenly endorsed it (1954). Yet the snag of the singularity was still overlooked. To prove the ineffectiveness of the new axiom, I countered with an analytical example (1954) in which the integrals of (1) involved a singular point – a node, later renamed pole – which naturally impeded the establishment of an ophelimity index. Within such a map the consumer could move around the node and arrive at the same subjective state as the initial one, a possibility not denied by Houthakker's axiom. A feeling that Pareto was right in a deeper sense than I had shown until then began pressing me. But it was only at the symposium in Pareto's memory (1973) that I presented the ultimate analytical example in which (1) is derived from *community* demand schedules and is nonetheless completely integrable without any singularity whatever. Those integrals do look like the usual indifference map, but we know only too well that they cannot be associated with it. Therefore, to identify integrals with the indifference varieties we must know beforehand that an ophelimity index exists, as Pareto did. This conclusion vindicates Pareto (and by the same coup exposes the behaviorists' folly of rejecting all subjectiveness).

Dialectics versus arithmomorphism

Science began as a trove of propositions describing some observed phenomenon. It still consists of such a trove, though restructured under the continual pressure of the limit of human memory. At first, some propositions were classified in groups, as in the Hammurabi Code. Later, some great relief for memory came with writing and the convenient materials for doing so. Taxonomical classification of the kind we find in biology and even in chemistry was the next advance. Ultimately, some ancient land surveyors on the Nile discovered that if one has memorized

- (A) The sum of the angles of a triangle is two right angles, one need not also memorize
- (B) The sum of the angles of a convex quadrangle is four right angles.

That was the germ of the *theoretical* science (not of every science). In a theoretical science all descriptive propositions must be filed, not alphabetically (as in a directory), or taxonomically, but in a *logical order* as in geometry. Through an intricate logical sorting all known propositions can be divided into two classes:

- (a) Every α proposition follows logically from some β propositions,
- (b) No β proposition follows from any β propositions.

All we have then to memorize is the set (β), for by simple ratiocination we can rediscover all the others. The greatest advantage of logical filing is thus the economy of thought, a point brought up by Ernest Mach and Kari Pearson. For many, though, memorizing is much easier than ratiocinating; many a student prefers courses based mainly on memory. Why are there such sciences?

It was this question that set me on the path to dialectics, for I observed that logic, though a marvelous accessory for our thinking, has its limits set by its own power. Logic works only with a restricted class of propositions, such as

- (A) The hypotenuse is greater than a leg but is totally impotent when it comes to propositions such as
- (B) Culturally determined wants are higher than biological needs.

Worthy of special note is that all concepts in proposition (A) are as *discretely* distinct as any clear symbol, say, m , 2, or ∞ . Discrete distinction is the specific property of real members; a number retains its distinct individuality even within the arithmetical continuum. This is why I have proposed to call such concepts *arithmomorphic*. No arithmomorphic concept overlaps with its opposite. The boundary between the two is vacuous: *tertium non datur*. A vast number of concepts, however, overlap with their opposites. That is, A and non-A may be both true (which does not mean *tertium datur*). As Max Rheinstein once remarked, “Even the dictatorship of Hitler... had democratic features, and in the democracy of the United States we find certain dictatorial elements.” Proposing, to refer to concepts of this kind as *dialectical*, I have obviously followed Hegel, yet only for a short while. In my view concepts are means of expressing our thoughts, not legislators of nature and society as Hegel and Marx (in a switched way) claimed.

Arithmomorphic concepts are absolutely invariant: “square” meant the same thing to Euclid as it does to us today. If in our imagination we alter an angle of a square no matter how little, it is no longer a “square” But we can squeeze an “oval” a great deal before it is no longer an “oval.” Dialectical concepts, though not discretely distinct, are distinct. They are separated from their opposites by a dialectical penumbra that is in turn delimited by other dialectical penumbras. A baby will be old when he will be ninety; but no one can say when he will just become old. On this point as well as on similar others, Bertrand Russell argued that one can determine that event by associating it with a convenient number. What he proposed was to define a particular democracy as that of the United States at, say, March 15, 1896, at π o’clock p.m. But my epistemology faults him. In relation with facts we cannot use paper-and-pencil numbers, 1 or π . A pointer-reading belongs to dialectics. This issue recalls one of Schumpeter’s incisive protests: “There is no sense in our case in asking: ‘Where does that type [of entrepreneur] begin then?’ and then to exclaim: ‘This is no type at all!’”

Entities that change qualitatively are necessarily dialectical. The epitome is “species” which is dialectical because, as Charles Darwin put it, “it includes the unknown element of a distinct act of creation.” The present temper that insists that “species” is an arithmomorphic concept is tantamount to a return to Lamarck, to species created once for all. Diehard logical positivists naturally would forever insist on exclusive arithmomorphism. Yet these apostles are utterly mistaken, for not even they could plead any case without using more dialectical concepts than arithmomorphic ones. Is “a sufficiently large sample” or “verifiability,” for instance, arithmomorphic? Eminent scholars - like Bertrand Russell and Percy Bridgman - who made a point of honor in combating vagueness offered us the best proof that reasoning with dialectical concepts is not only quite possible, but also indispensable.⁴ Only it is far more difficult than doing algebra when, as it often happens, the tip of the pencil may move faster than the writer’s mind.

This brings to mind Blaise Pascal’s immortal dichotomy *esprit géométrique* and *esprit de finesse* of which the widespread arithmomania of our epoch would rather not hear. Even the bland way in which I put it at a 1955 symposium - “There is a limit to what we can do with numbers, as there is to what we can do without them” - was anathema to the worshipers of the Almighty Arithmomorphic Concept. At the famous David Novick symposium organized by Seymour Harris (1954), Lawrence Klein proclaimed that “nonmathematical contributions to economics [are] fat, sloppy, and vague.” A verdict on Adam Smith, Schumpeter, or Simon Kuznets? And because at that time crime and drug

4 For completion, I should add that a still greater economy is achieved by introducing some thought-up propositions (ω), which added to (β) propositions shift many more of these to the (α) set.

addiction were not on the rise, Robert M. Solow could get off unscathed for having reasoned that mathematical economics must be really good because everyday there is more, not less, of it. But the actual crux was exploded later when Salim Rashid produced the document of the time: because of the views of the economic establishment, junior economists, he said, must grind papers by the mathematical engine, lest they perish.

Dialectics and similes

There is justice in the positivist objection that communication with dialectical concepts cannot be precise. With this point in mind, I tried to mirror dialectical concepts by some analytical pseudo-images to which I have appropriately referred as *similes*. While struggling with modern utility theory, my epistemology took offense at the absence of any mention of wants or dislikes - the real movers of our actions. I was thus delighted to discover that wants of all forms had formed the pillars of the older consumer theory propounded by T. C. Banfield and Carl Menger, now fallen from favor. Standard economists have chosen just to putter with the second differential of a nominal, opaque blanket named "utility," a term with which Jeremy Bentham himself was unhappy to the end. The neoclassical rationalization was that want cannot be defined precisely. To be sure, want is a dialectical concept. If want had been a rigid arithmomorphic element, the human species would not have been able to survive under the radically different environments of its long past. Wants even form a dialectical hierarchy: above those that respond to biological needs (which are common to all human) come those that correspond to social propensities (common to all members of the same community), and above these, the purely personal, disordered whims. On the basis of this hierarchy, we can justify two of the most essential propositions about us. First, the principle of marginal utility is just shorthand for the law that any human satisfies his wants in their hierarchical order. And, second, contrary to the most unfortunate fallacy of standard economics, most wants are interpersonally comparable. All humans - the Rothschilds and Hollywood stars included - will spend their *only* taller for quenching thirst and assuaging hunger. Of course, interpersonal comparison between the upper wants of two rich people, one enjoying a motorboat, the other a villa, makes no sense. Since standard economics is a discipline of the lands of plenty, Lionel Robbins' famous theses of interpersonal noncomparability fits in place there, but only there.

After seeing that Carl Menger's table of wants cannot explain how one distributes a given income with given prices among one's various wants, I proposed a simile diagram in which to every want there corresponds a domain delimited by ordinary fines and located according to the general hierarchy. It was by this factual analysis that I proved for the first time the necessity of lexicographic order for economic theory.

Similes had been occasionally used earlier in other respects. One of the most interesting cases is the theory of probability. The Laplacean, the frequentist, the betting coefficient, and all other definitions tried out are all arithmomorphic similes of probability, which is a dialectical notion in the truest Hegelian sense: it starts and ends with itself. It is because of this dialectical nature of probability that all the mentioned similes have ended in contradictions. The probability associated with natural phenomena is dialectical because its backbone, randomness, is a dialectical notion, for randomness implies irregularity, yet unlike the desultory haphazardness that irregularity is regular. I have captured all this in the following proposition:

If A is a random event and f_n is an observed relative frequency, there exists an associated number p such that, for any positive, ε and δ , there is an integer N such that

$$1 > \text{Prob}[|f_n - p| < \varepsilon] > 1 - \delta \quad (2)$$

is true for any number of observations $n > N$.

That p is the probability of A .

The double-barreled production function

In another paper at the Harvard Economic Project (March 22, 1949), I pointed out a serious epistemological discrepancy between the two production models in vogue at the time, Neumann's and Leontief's. Because the first takes into account only stocks, it hides what may have happened between the beginning and the end of the process. Because Leontief considers only flows, he does not allow one to know which of two processes is more efficient. Intrigued, I turned to the standard production function, the definition of which (very strange) has remained in the same vapid form in which Philip Wicksteed introduced it almost one hundred year ago (1894): "*The product being a function of the factors of production we have $P = f(a, b, c, \dots)$* " He said nothing about the kind of "function," or about the nature of the "factors," and no economic luminaries have ever questioned that diction. Some just said that the representation of a process involves only quantities:

$$Q = F(X, Y, Z, \dots); \quad (3)$$

others, that it involves only rates of flow per unit of time:

$$q = f(x, y, z, \dots). \quad (4)$$

Yet no one seems to have been bothered by this double-barreled view, not even Ragnar Frisch, who used both definitions on the same page (*Theory of Production*, p. 43). Using the elementary identities $X = tx$, $Y = ty$, $Z = tz$, ... for any t , I proved that if both (3) and (4) are equivalent representations of a process, the two functions must be identical $F \equiv f$, and, moreover, homogeneous of the first degree. Hence absolutely *all production processes are indifferent to scale!* I presented this astounding result in a formal paper at a 1965 conference of the International Economic Association. According to their rules, only the contre-rapporteur should present the highlights of the author's paper. My contre-rapporteur, Don Patinkin, stated that he could not introduce a paper vitiated by a "fundamental mathematical error" and simply sat down.⁵ This high-handed attitude of an asservative economist was an emphatic proof of how incredible my theorem could then be judged by standard economists. Did not Joan Robinson claim with her usual feeling for what economics should be that standard theory of production is an economic miseducation?

Analytical process: flows and funds

"Process" is one of the most frequently used, and also the most abused, term in science. We find no definition of it even in Alfred North White-head's famous opus *Process and Reality*. Its detailed description runs against several paradoxical tangles that are circumvented by a jump in the dark, from dialectical to arithmomorphic (or analytical). What is involved epistemologically is, first that an analytical process is *identified* by a vacuous boundary of double nature: a spatial boundary and a temporal boundary that must not begin or end at infinity. Second, even though inside the boundary things happened in every location - too many to be listed - in analysis what a process does is described only by what crosses the boundary. The definitions "input" equals what you put in, and "output" equal what is put out can now be made analytical: input is what crosses the spatial boundary from *outside*; output is what crosses is from *inside*.

5 While writing on this issue for my *Analytical Economics*, I said in passing that a new Aristotle might set dialectical reasoning on as solid a basis as the traditional logic. By a strange coincidence, L. A. Zadek ("Fuzzy Sets" *Information and Control* 3 [1965], 338-53) had just claimed to have achieved this. But the claim, endorsed by legions, is spurious. The entire construction, beginning with the membership function $f_i(x)$, is purely mathematical; hence, it has nothing to do with dialectical concepts.

There are only three distinct and exhaustive cases: (1) factors that go in and never come out, (2) factors that come out although they have never gone in, and (3) factors that come out unchanged just as they have gone in. There is also a fourth logical category - factors that do not cross the boundary at all. They are *internal* flows illustrated, say, by “payments of business to business” which smuggles a dialectical concept into analysis. This confusion, also committed by Karl Marx, necessarily ends in errors if not in a paradox.⁶

The three Production factors just described correspond to the classical Ricardian land, capital equipment and labor power. According to my epistemological view, they are the *agents* that transform the inflows into the outflows. I have proposed to refer to them as *funds* conceived as agents of constant efficiency.⁷ Clearly, excepting Ricardian land, everything changes with time. For the purposes of analysis, however, capital may be assumed constant, as Karl Marx first proposed. A separate process the household, aims at maintaining the vital ability of people in order.

What a process does is then analytically represented by a set of functions of time t from $t = 0$, the beginning of the process, to $t = T$, the end of the process. Each such function represents the cumulative amount up to time t of a flow, entered or exited, or of *service* provided by a fund. This new way of representing a production process is a vector of functions, that is, a functional,

$$[R_i(t), I_i(t), P_i(t), W_i(t); L_i(t), K_i(t), H_i(t)]_0^T \quad (5)$$

where the alphabetization indicates in sequence: natural resources, intermediate products, products, waste, and next, land, capital, and labor power. Each production process is thus represented not by a timeless vector in the commodity space as in standard theory, but by a *curve* in the same space. A critical difference that I thus introduced is the inclusion in (5) of natural resources and waste, inevitable but totally overlooked factors of any process.⁸

6 To be sure, shortly after the conference Doussinakin realized that he had been wrong (probably because he was not aware of the difference between *identify* and *equation*) and requested that his criticism not be published in the *Proceedings*.

7 When Leontief first presented his input-output system, he repeatedly insisted that all diagonal coefficients should be zero, which meant that the matrix should include no internal flow. However, in his later applications he did include them. To make the absurdity of the internal flow clearer, I devised a multiprocess matrix in which there is no empty box into which one could inscribe a coordinate for internal flow. That matrix also enables us to dispense with the helplessly intractable flow diagrams that overlay the recent ecological monographs (in which the frequent use of the “loop” should have exposed the limp concept of internal flow).

8 My concept of fund should not be confused with that of stock. The role of stock is to receive or to generate flows. And contrary to some opinions, the flow-fund model is essentially different from the flow-stock model encountered in the economic literature.

Standard theory makes intensive use of isoquants - *geometrical* curves - to represent the substitution of factors without changing the output. But since neither capital equipment nor laborers can be quantified as needed for the isoquant, if anybody uses an isoquant, one must inevitably admit that one's own framework is essentially dialectical. Some have indeed likened capital to clay or to putty. *I know of no other scientific discipline in which dialectics is as indispensable as in most sectors of mathematical economics.*

Patterns of production processes

My epistemological search has also led me to the new fact that production takes place in several and entirely different types of processes. The simplest type is that which not only is represented by (5) - all are - but can be represented only by it. It is the process of a single craftsman's workshop, where at any time work is applied to only one unit (or one batch) of the product and units are produced *in series*. And if we tease out any other kind of process, we find that all consist of some arrangement of such simple processes, which I *called elementary* processes. For an agricultural product within a uniform climate, the elementary processes are arranged *in parallel*.

One point now deserves unparsimonious attention: in any elementary process virtually all funds are idle over certain periods, and this idleness cannot be completely eliminated by technology. Think of the plough in the temperate zone or of the saw of a cabinetmaker. If in the latter's shop an additional craftsman is brought in, the two could use the same tool alternately and thus decrease the idleness of each fund. The production would thus be speeded up many times more, which would require an equally increased intensity of the demand, a finding that analytically vindicates Adam Smith.

In the industrial sector we also find that production - of one spacecraft, for instance - is an elementary process. But that sector is dominated rather by the factory system, so common a view, yet so totally disregarded. Like money, that process is a purely economic invention, not a technological discovery. A theorem I proved states that any set of commensurable tasks that would constitute an elementary process can be arranged in a pattern that would eliminate all idleness, a commonplace illustration of which is any assembly line. This is the superiority of the factory system, which has still another economic advantage. Its production needs no waiting. If Bali Island (where uniform climate would permit it) agriculture used the factory system, it could be said that the Balinese eat the rice sown that very moment. This peculiar property is due to a specific capital item, goods in process or *process fund* (my preferred term).

Production by factories, though, needs waiting too: building the plant itself and priming it may take years. Contrary to Piero Sraffa's celebrated thesis, production of commodities in general needs not only commodities as such, but factories too. And because there is no factory to produce factories, some waiting is irreducible in the case of growth.

Curiously, the simplest analytical representation is that of the factory process for which the arguments of functional (4) are simple linear homogeneous functions of t . We may thus put in a generic form $A(T) = aT - A^o$ for every element of (5), the notation A^o standing for quantities and the lowercase a for rates with respect to time. For the representation by quantities, there is a hitch: one hundred pounds of nails says nothing about T , which might have had almost any value. For this reason, in passing from the functional to its degenerated form - the quantitative vector - we must include T in the new form

$$(R^o \ I^o \ Q^o \ W^o; L^o \ Q^o \ H^o; T). \quad (6)$$

By contrast, the representation by a vector of rates need not explicitly contain the time coordinate:

$$(r, i, q, w; L, k, h). \quad (7)$$

Formulas (6) and (7) dissolve immediately the paradox of the double-barreled production function. The epistemology of the propounder of production function (2) failed to realize that quantities are not time dimensional. Another common epistemological fumble is the indiscriminate use of "flow" in saying, with Marxians and the legion of energetists, that the sewing needle, not only the cloth, flows into the pants.

After examining the blueprints L, K, H of a factory, specialists could say (1) how large would be the output q and (2) what production flows would be necessary for it. The factory process must be portrayed by two similes:

$$q = \Phi(L, K, H) \quad \text{and} \quad q = \Psi(r, i, w). \quad (8)$$

Factory production, therefore, is affected by a strict limitationality: it cannot produce more shirts by increasing only the sewing machines or only the input of fabric. A production function of only funds and flows, therefore, is total nonsense. Yet with the recent discovery of resource scarcity, numberless economists have used the function $q = F(H, r)$ for selling the newest economist's conjuring trick. If that formula were epistemologically valid, we could at whim substitute capital equipment for iron ore while increasing even the production. Another fictional function is $q = F(K, H; t)$, where t supposedly represents

technological progress. Several standard economists have used the partial derivative of F with respect with t , completely ignoring when they were trapped the fact that neither the function F nor the arguments K , H , are the same in 1980 as in 1960: $F_{1980}(K^{80}, H^{80}) - F_{1960}(K^{60}, H^{60})$ is not a difference on which the derivative is based. This fumble proves that even in mathematics we cannot do without epistemology.

Growth versus development

Ever since John Maynard Keynes' *General Theory*, growth has been conceived as a purely monetary phenomenon sustained by the century's monetary witchery: if government spends more for itself, all people will grow economically. For the simplicity of the diagram with the 45° line, Keynes became the darling of economists and, before long, of the politicians, who could now rationalize moving mountains without increasing taxes.

In spite of my uneasiness in approaching monetary facts, my epistemological penchant found, nonetheless, some delight in the problem of growth because of Neumann's and Leontiefs models of production. The mathematics of even the generalized Leontief system are rather simple (and I thought to have them all included in my article "Some Properties of a Generalized Leontief System" presented at the memorable conference in Chicago [1949] by the Cowles Commission and reproduced in my *Analytical Economics*, Chap. 9). They tell us that Leontief's system in which labor is the only primary fund needs only a definite amount of labor power, L_i , for the production of an additional unit of commodity C_i . Is this not a secret of growth? Reluctantly, I must disappoint the adherents of this viewpoint: the principle is true mathematically but not operationally.

A point I can hardly overemphasize is that *in the initial Leontief system, just like in Marx's labor theory, labor is necessary but not sufficient for production*. A pesky question therefore confronts would-be planners: given that coal is necessary to produce iron and iron is necessary, too, to produce coal, where does any growth plan based on Leontief's system get the necessary amounts of these commodities? An authentic story pinpoints this antinomy: a cookbook advised cooks to prepare stock A with some of stock B and the latter with some of the former. Even the king's cooks could not prepare anything.

It is very simple to calculate the matrix in which all net outputs are increased, yet how to pass from the initial to the last matrix is one of the most stubborn economic problems. Even Karl Marx, who considered both a stationary and an expanding state, never showed how to pass from the former to the latter.

Of course, instantaneous growth can be achieved at will if we increase the working hours of the shift. But this trivial solution is inapplicable in Leontief's system, which ignores that coordinate. And it would also be unacceptable to the temper of this era, which refuses to recognize the primary truth upon which I have repeatedly insisted: the economic progress of the West was fostered by a very long working day.

On paper, we may start by saving some of one commodity and, by ramification, stage by stage reach a higher net income for all commodities. In Leontief's dynamic system even growth by stages is beset by several snags. In that system, because growth also requires increased funds, an awful snag is that either bulldozers are consumption commodities or yogurt is a producer good. The morale cling to the idea that all facts involve irreducible structure.

Another strong epistemological uneasiness of mine concerns the use of differential equations to show how growth can be implanted in a poorly growing economy by fitting it to a growing differential system. Of course, this is the acme of self-deception.

For a final word, we should not delude ourselves that mere accretion poses no intricate epistemological issues. Remember that biologists had long been tormented by the problem of how the accretion of simple cells occurs. Only the double helix, by being its own negative template, cleared the mystery. Yet notwithstanding the claims of eminent biologists, development has still not been explained. My epistemology prompted me to dissent: the DNA of a zygote warrants only its identical self-reproduction, not its metamorphosis into, say, a nerve cell. How much more mysterious, then, should economic development be? I was taken to task for having said that the most valuable clues for that issue come only from economic history, but Professor Schumpeter saved my soul by avowing his own confidence in history in *History of Economic Analysis*.

Feasible recipes versus viable technologies

I came to realize the indissoluble dependence of the survival of humankind on scarce resources by the combined influence of two sources: Emile Borel's monograph on statistical mechanics (alias thermodynamics), which as a student of statistics I read in the 1920s, and the problems of an overpopulated agrarian economy, of which I became fully aware during my Romanian exile. I then became convinced that nothing can solve the problems of an exploding population except, as Maithus argued, the population itself. The thought that, even if the population stopped growing, its predicament would still remain came to me one day as I watched a big Romanian river running in its bed furiously and with a chocolate color. There goes, I said, our daily bread of tomorrow.

Homo sapiens, the exosomatic animal

As a witness of the political events during my Romanian exile, I realized that Romania could not remain neutral (as virtually everybody wished) in World War II. From history I also learned that this was true for World War I. The only impediment was Romania's possession of great reserves of oil, which neither Germany nor Russia was willing to let the other control. Other animals, however, do with just what they have. The quintessence of this view was recently expressed by Andrew Rooney, who watching a chipmunk at work observed that it never stopped to go to the hardware store to buy a tool, as humans must do. In all activities of life, all animals (humans included) use their organs with which their somata have been endowed by birth - *the endosomatic* organs, a term coined by the unusually perceptive biologist Alfred Lotka. And it is through changes of the endosomatic organs that every animal becomes better (or even less) adapted to life. But this mode of evolving is extremely slow. The human species alone found a far speedier way. Even some ancestors of *Homo* began using and finally making detachable *exosomatic* organs: first, stone hammers to bit harder; recently, airplanes to fly higher and faster than any bird. This does not justify the popular definition of humans as the only toolmakers. But as Henri Bergson first observed, humans are the only animals to use tools to make tools or, as Schumpeter used to say, to make machines, to make machines, to make machines, on end. The exosomatic human can do things that could not be done before. But exosomatism, I should stress, is not an unadulterated blessing. It is the root of inequalities within the same society as well as among societies. Because the production of exosomatic organs has to be planned and supervised, human societies have been divided into those who work (the ricksha men) and the governing individually (the mandarins). And, not to forget, exosomatism has also made us thoroughly addicted to the exosomatic comfort - hence almost completely dependent on the finite mineral dowry of our abode.

Thermodynamics and economic scarcity

While men of science were still interested in celestial affairs and in explaining them by the laws of mechanics, Sadi Carnot, a young French officer, published in 1824 an immortal memoir on the efficiency of the steam engine, the seed of the science of thermodynamics. For these reasons I argued that Carnot was the first genuine econometrician and that thermodynamics is in essence a physics of economic value.

Thermodynamics has had an agitated history, and its theoretical structure is now beclouded by a swarm of mathematical fantasies. My own struggles with the vacillating literature have led me to a very simple and clear conception of it. One of its four laws, the

entropy law, as engendered endless controversies. Yet its content boils down to a commonplace known from the time when man was not yet *Homo sapiens*. It simply states, as Rudolph Clausius put it in 1856, that heat always goes *by itself* from the hotter to the colder body, never in reverse. Or as I put it in a more homely way, it is the hand touching a hot stove that is burned, not the stove. Since such transfer of heat cannot be prevented between bodies in contact, it follows that everywhere in the universe hot things *continuously and irrevocably* become colder and cold things, hotter. Clausius cast this fact into "The entropy of the universe tends toward a maximum." But what exactly is entropy?

My view is that the entropy law, like thermodynamic laws, reflects basic limitations of all living creatures. Let me begin by noting that a formulation of the entropy law that denies the possibility of converting the energy from a single source of uniform temperature into work, although well established in the literature, is not true; a piston and cylinder, by absorbing heat from such a source, can do it, as happens during the first expanding phase of a Carnot cycle. Yet we cannot take advantage of this splendid "engine" because of our limitation *in space*. Even after a small course, the piston must be brought back to its initial position. For this we must spend the same amount of energy as the work gained when the piston moved forward. There remains no surplus work for us. The solution to the impasse is one of Carnot's keen propositions: bring the piston back by a colder route than the first and you will get some surplus work. All steam engines must therefore work in cycles between a hotter and a colder temperature. In general, we can obtain work only from a source that involves a difference of temperature, of electrical or chemical potential. Only such energy is *available* (useful) to us as humans, homogeneous energy is *unavailable* (useless) to us. These fundamental thermodynamic concepts are clearly anthropomorphic. They justify my earlier contention that thermodynamics is a physics of economic value.

Any direct conduction of heat from hot to cold therefore robs us of available energy. Another robber is friction. However, friction does not produce heat if the motion is infinitesimally slow, in which case any movement would take a virtually infinite time. That possibility is off bounds for us, because we are limited *in time* as well.⁹

Those who have an unrestrained confidence in the power of science to fix anything keep preaching that science would help us get rid of even the entropic degradation. As a simile for the entropic degradation, I once used an hourglass assumed not ever to be turned upside down. Paul Samuelson followed with the remark (*Economics*, 11th ed.) that "science can temporarily turn the glass over." Yet I would not advise anybody to settle in an aentropic

9 These results were presented twenty-five years ago when pollution had not yet hit us in the face, nor had the embargo of 1973 made us aware of nature's niggardliness.

world (if one existed); for if one took a bath there, one might have the neck scorched and the toes frostbitten by a redistribution of heat. Not would I enjoy living in a frictionless world where I could not write or walk in the direction I wanted.

I have further argued, and very strongly, that matter, too, is subject to entropic degradation, that available matter in the bulk (say, the rubber of automobile tires) degrades irrevocably into the unavailable form of the rubber particles dispersed by friction on the pavement. We delude ourselves if we trust the popular belief that matter, unlike energy, can be completely recycled. What we can recycle is only *available* matter that is in an unusable form: broken glass, old papers, worn-out motors, and the like. I have stated this as the impossibility of perpetual motion of the third kind, defined as a closed system that could exchange only energy and would perform work at a constant rate forever. I have referred to it as the Fourth Law of Thermodynamics. The reaction to it has been thin and also very strange, for some have simply asserted that the law has long been known, while others have objected that it is not true. The latter critics have committed the same error as Galileo in claiming that the air offers no friction to the flying arrow on the ground that its existence could not be detected by any instrument of that time.¹⁰

The root of economic scarcity, hence of economic value as well, lies in the entropic degradation of energy and of matter in bulk. A different kind of scarcity is represented by Ricardian land, which is scarce because it sets a limit to the daily carrying capacity of the earth. The scarcity of mineral resources sets no reasonable limit to how much of them we can use during one day, but it sets a more dreadful limit, a limit on the survival of the human species on this planet. This was my message of twenty years ago, which, though pessimistic, did not spring from a pessimistic *Weltanschauung*, but from known facts. Yet a tidal wave of writers, new and old, have sought public admiration by opposing an ultra-optimistic battle cry to my pessimistic message.

10 A few critics, typically, represented by Carlo Bianciardi, Paolo Degli Espinoza, and Enzo Tiezzi "Ma la materia há una storia." *SE Scienza Esperienza* 4 (July 1986), 40-1. Argued that with a magnet it is possible to reassemble all iron filings into the original piece. But they did not specify the instrument that would guarantee that the iron particles dispersed during the proposed experiment, and *only these*, would be picked up by their proposed magnet. Naturally, I think my law is true, yet I would not oppose any scientific denial of it. As I told Ilya Prigogine during a symposium sponsored by the U.S. Department of Transportation, we must know whether my proposed perpetual motion is feasible or not: we have answered this question for other perpetual motions.

The promethean destiny of humankind

Confronted with the recent symptom of the scarcity of our environmental dowry, the 1973 oil embargo, economists in particular have reacted according to Disney's First Law, "Wishing will make it so," as William Miernyk aptly put it in his piercing contributions to the problem of the exhaustibility of fossil fuels. Economists have authoritatively advised us to go home and sleep tight in our beds assured that "come what may, we shall find a way" as we have done ever since the time of Tutankhamen. Robert M. Solow even declared in his Richard T. Ely Lecture that "the world can, in effect, get along without natural resources, so that ex-haustion is just an event, not a catastrophe." And legions have argued that "solar energy is here, we can use it now," as Denis Hayes, a very sound student of the problem, proclaimed in the *Washington Post* a few years ago. Standard economists have made a defensive circle around the dogma that the market knows best, that prices will take care of any economic turnabout. I have strongly dissented from this economic fantasy. Its advocates have completely ignored that we could not let the polluting driver pay: instead, we have enforced the use of the catalytic converter by law. Are not the whales on the way to extinction precisely because the price of their meat is right? The same also goes for the deforestation of all time, especially of that in Brazil now.

There now rages a fashion to fancy one or another alternative to the current technology. None (yes, none) is worth anything because none has taken account of the fundamental condition of a viable technology, to which I now turn.

The number of production recipes used by humans ever since they became exosomatic animals is so enormous that even a lifetime would probably not suffice to compile a complete list of them. However, surprising though it may seem, only three of that vast number have effectively pushed on our exosomatic progress. In chronological order they are husbandry, the mastery of fire, and the steam engine. I have called these recipes *Promethean* on remembering the old legend that Prometheus, a Titan, stole the fire from the gods and gave it to humans. With just the spark of a match we can set on fire a whole forest, nay, all forests. This property, although not as violent, characterizes the other two Promethean recipes. It is a commonplace that a seeded grain of corn will normally yield a surplus of a handful of grains.

The steam engine, however, needs further discussion along with the prodigious story of its invention. Helped by the Promethean fire, humans were able to keep warm, cook food, bake ceramics, and above all smelt metals. An era of vigorous technological progress thus began. But given human impatience, any recipe that increases our power over things is self-defeating. We would normally use it oftener and oftener so that its technology would spread at a fantastic speed. In this way, by the middle of the seventeenth century the technology

based on Promethean fire ran out of its fuel: wood. Deforestation was advancing so fast that even in Norway legal restrictions onto cutting trees had to be introduced. Coal had been known as another source of heat ever since the thirteenth century. But the energy of coal, *thought available*, was not profitably accessible. Below even a moderate depth, underground water floods any mine and must be drained off, which requires an appreciable amount of energy. At the time, this posed an insuperable problem. Operators of mines asked even Galileo for help. He advised them to use a pneumatic pump; nature, he explained, abhors vacuum. But after they reported back that, no matter how hard they pulled out the pump, the water would not rise above some ten meters, Galileo reflected that perhaps nature abhors a vacuum only up to that height. The situation was in all respects like that of today. Fate had it then, that Prometheus II - two mortals, Thomas Savery and Thomas Newcomen - saved the day by inventing the steam engine. This engine, too, is a Promethean recipe: with just a little coal under it we can drain the water completely from a mine and also bring out far more coal than that used by the engine, nay, enough to operate other mines as well. Yet the steam engine has its entropic limitations, too. What the energetist tyros seem to ignore is that absolutely no recipe can produce additional available energy or available matter. From what? I would ask. Let us assume that another earth would possess vast reserves of bituminous coal located 10⁷ feet below the surface. Since it would take more than the energy of one pound of coal to mine one pound, no steam engine could then be Promethean for that coal. The object lesson of this parable for the self-styled energetists is that a recipe that works well in the laboratory (as many do) may not necessarily support a viable technology.

From what I have said in this section, it is obvious that no viable technology can exist unless it is supported by a Promethean recipe. A new Promethean recipe, not just any fancy contrivance, is what the present crisis needs (a point totally ignored by those who exalt, individually or within the ever-growing number of global associations, one's own alternative solution). The stern question now is: will Prometheus III come in time to save our souls by a new Promethean recipe? Some claim that two Promethean recipes are already at hand. One is indeed: the breeder that produces more fissionable fuel than it consumes - hence its uniquely alluring name. Its hitch is the genetic danger, of its waste. The second (alluded to earlier) is the *direct* harnessing of solar energy. With sanguine hopes but with accomplished technological knowledge, a serious attempt was made by Solarex, Inc., to construct a solar "breeder" The result of that experiment was categorical: the amount of solar energy captured by a number of silicon cells does not suffice to reproduce them all even if all the necessary materials are obtained gratis from elsewhere. Automobiles and airplanes have been propelled by solar cells. But as my epistemological obsession forces me to observe, the energy that produced the cells, the automobiles, and the planes came

from nonsolar sources. At this time, harnessed solar energy is, like electricity, a parasite of other energies.

Faced with the present hovering crisis, what could humankind do? In strictest logic the answer is: practice "conservation", which, nor to remain just a word é must be examined epistemologically. Two important factors emerge. The first is the necessity of reducing consumption so as to slow down the depletion of our vital resources to the minimum compatible with a reasonable survival of our species. A prominent economist challenged me to set a number for the reasonable consumption rate, an inept challenge. Are there numbers *set* for the rates of taxation, hospitals, schools? Undoubtedly, we must adopt some austere program (nor to go back to nature, as some have wrongly read into my writings). Besides renouncing all kinds of instruments for killing ourselves, we should also stop overheating, overcooling, overlighting, overspeeding, and so on. Most important, we should cure ourselves of the morbid craving for extravagant gadgetry, such as the contradictory golf cart and two-garage cars. I think that we could stop following fashion, that disease of the human mind, as Abbot Fernando Galliani described it in 1750.

The austerity program should, of course, apply primarily to the lands of plenty, certainly not to economically wanting people, say, to Bangladesh. But nations with a growing overpopulation should make all efforts to stop growing in numbers. In a facile way we all speak of overpopulation without realizing that, if the United States were as densely populated as Bangladesh, its population would be just over 6 billion, the present population of the world! This is what overpopulation is.

Conservation would also allow more time for Prometheus III to emerge, and should he or she fail to come in useful time, we will be capable of sliding without social convulsions into a technology not identical with, but very much like, the old one based on wood. Unfortunately, two obstacles rise against this plan. First, no human would voluntarily give up luxuries or even conveniences to help some future generations have plowshares. The human species seems determined to have a short but extravagant existence. Second, conservation is not a program for a club, a town, or even a whole nation. It requires the participation of all in a world organization that would administer the use of the worldized (a word I coined after the manner of "socialized" and "nationalized") resources. But perhaps the human race will pass into extinction segregated economically. It cannot be ruled out that some of the last people should die in penthouses, the others in hovels. *Chi vivra verra.*

Bioeconomics and evolution

Economic life is a unique process that goes in historical time and in a disturbed environment.

Joseph A. Schumpeter

After learning of Alfred Lotka's idea that the role of our tools is analogous to our biological organs, I began thinking that a greater analogy exists between exosomatic and endosomatic organs. On that trail I saw that exosomatism was the fountainhead of the economic process. Since the exosomatic organs offer unique advantages to their users and also are detachable, they began being traded and being produced for trade. Production for trade ultimately led to large social organizations. This development brought down upon humans the irrevocable predicament alluded to earlier: the social conflict. Societies of other species do not know such a conflict. The periodic killing of drones by worker bees is a biological, natural action, not a civil war. The reason for the difference is the fact that the role of the individual in the latter societies is decided at birth. The ant doorkeeper, for instance, has a flat head with which it blocks the entrance of any foreigner to the gallery, and he would not like (as we can judge from experiments) to do anything else. But in our case would not a ricksha man, for example, prefer to be a mandarin? And would he not struggle to become almost one?

The exosomatic organs evolve just as the biological ones do, though much, much faster. And just like the latter, they may be deleterious to the species: enormous deer antlers and the automobile that "attains one hundred miles per hour before the cigarette lighter gets hot," as a topical advertisement praises it. Our incurable addiction to even futile exosomatic organs complicates further our existence with problems that belong to *bioeconomics*.

Bioeconomics reminds us of Alfred Marshall, who first envisioned the sisterhood of economics and biology. Although he repeatedly preached that biology, not dynamics, is the Mecca of the economist, he himself hardly practiced that teaching. The economist who developed a general framework to represent evolution everywhere, nor only in economics, was Professor Schumpeter. Let me explain this great contribution of his to science, for it still needs to be explained. Schumpeter's vision of development, as he termed it for the first time in opposition to accretionary growth, anticipated by some thirty years a salient idea thought up in 1940 by a prominent biologist, Richard Goldschmidt. Schumpeter's view was that economic evolution is constantly fostered by discontinuous innovations, the product of the continual inventing faculty of the human mind, whereas Goldschmidt contended that biological evolution fares primarily through successful monsters. Inspired by Schumpeter, in my bioeconomics I assimilated the emergence of palpable endosomatic changes with his

chain of innovations. Both are essentially unpredictable, not even randomly regulated, a point that exposes the fantasized attempts, such as Trygve Haalvelmo's, to equate evolution with an arithmomorphic mechanism.

In a firmer way than the biologists, Schumpeter maintained that economic evolution is irreversible just as the biological one is. But several biologists who believed in the supremacy of mechanics argued against irreversibility by pointing out that the color of the drosophila's eye changes back and forth constantly, like a pendulum. In this opposition there lies Schumpeter's piercing idea, which he relegated to a footnote (p. 81). There, Schumpeter excluded from innovations small changes, reversible changes akin to drosophila's changes of eye color. It was at that point that Schumpeter was confronted (for the first time, I believe) with an issue of a dialectical nature: What change is small? And there, as on other occasions, his answer was that you and I know when a change is small, although neither of us is able to say exactly when. By taking this position, Schumpeter implicitly opposed the neo-Darwinism account of evolution; for as he splendidly put it, "Add successively as many coaches as you please, you will never get a railway engine thereby." The railway engine compared with the horse cab is a monster, but such a successful one, as Richard Goldschmidt might have put it thirty years later. Of course, in an ocean of Darwinists and neo-Darwinists, Goldschmidt's thesis could not be accepted. Yet very recently, Stephen Gould, one of the most active minds in biology, has rehabilitated Goldschmidt's theory, adding, interestingly, that no explanation of evolution can dispense with dialectical reasoning. Economists, however, have failed to pay any attention to the greatness of the conception of evolution first thought up by one of them.

Against some current?

Don't be modest, you are not that great.

Golda Meir

The question brings up the relativity of motion; for one may feel one is moving against a current although one just stays put, as many men and women did because they could not do anything else to oppose the Nazi onslaught. And there is the symmetrical case in which one may feel movement against a current even though one is just moving in a placid milieu, in which one might hear the whisper "*Sh! vous reveillez Monsieur*", as happened to me on a few occasions.

To try to ascertain whether I have ever moved against an objective current is not smooth sailing, for it inevitably entangles me in what, in line with my epistemology, I prefer to call

the *sociology of scientists*. This term correctly describes the discipline now known as *sociology of science*, as Karl Mannheim called it first; for sociology necessarily refers to living individuals: humans, chimpanzees, bees, horses. It would be nonsensical to speak of the sociology of books or of differential calculus. As the unorthodox sociologist Florian Znaniecki argued, we can speak only of the role of people in acquiring and spreading knowledge. It is by reorienting Mannheim's view that Robert K. Merton set the "sociology of science" on a better track. There is, in particular, Merton's magnificent studies of the Matthew effect, of the multiple discoveries, or of plagiarism, copies germane to what scientists do rather than to what science is.

We may not *all* be aware of the most striking illustration of the Matthew effect; "*E pur si muove*" is ordinarily attributed to Galileo, although those words were the last ones uttered by Giordano Bruno on the burning stake! To descend to common people, my theorem of substitutability of Leontief's static system is usually not connected with my name but with Samuelson's, although Samuelson himself has always acknowledged my priority. (Maybe, Samuelson can be modest.) In the economic literature we also encounter a veiled plagiarism when an author lists only very recent works, two or three years old, avoiding any reference to Adam Smith, Karl Marx, Vilfredo Pareto, John R. Hicks, or others just as great. The aim is to place oneself within the tidal wave of pseudo-innovators. Newton thought this practice to be an academic crime of which he accused Galileo for failing to mention Kepler. When I receive one elegant flier after another about future large congresses organized by energetists who have never referred to my contributions, I always ask, "Why do they send these fliers to me?"

I should list now some of my strange ideas that have a connection with the question of my running against a current. For a start, I thoroughly deny that money is an economic factotum. By itself, it creates impediments for the customary international aid consisting only of money. More often than not, such aid has filled the pockets of the privileged with still more money and has developed the industry of luxury goods instead of much needed wage goods. This wrong is aggravated by the fact that the wanting people usually are toilers of the soil, using either inadequate methods or inappropriate tools. We could train industrial workers by bringing them in successive groups to a huge teaching workshop, but we could not do the same with people occupied in husbandry. Northeast Brazil is the strongest case in point. With this idea in mind, after a 1965 meeting on subsistence farming I declared to a Honolulu newspaper that the best way to help the undeveloped countries was to send not gushers of money, not a peace corps, but a peace army. Would sending a peace army instead of one fully armed be an inept idea?

When the UN General Assembly met in Stockholm in 1972 to consider the problems of the environment, I participated in the meeting of the Dai-Dong Association, the sole

organization acknowledged by the UN. As Tom Artin tells in his *Earth Talk*, a delectable report about the general events of that occasion, I offered several motions that immediately upset the other members. One motion was that all natural resources should be worldized. My aim was to preclude increasing scarcity from accentuating the extant international inequalities and from eventually fomenting wars. In an interview with the *New York Times* (December 1979), I insisted that, if the use of resources is still to be at the whim of the market, missiles will fly for the possession of the last drop of oil. What recently took place in Kuwait was, fortunately, only a rehearsal, but a rehearsal in full dress. My second tabled motion was to abrogate all passports for international travel. It was another bioeconomic idea to aid the people of undeveloped countries by allowing them to move freely where there is a much greater opportunity for the use of their hands, instead of resorting to the conventional, but extremely difficult operation of bringing capital equipment into their native countries. These ideas certainly were utopian, but I would plead guilty and with pride to that incrimination. There is hardly any social or economic practice of which we are proud now that was not a distasteful, though fully sensible, utopia once. Yet I did not feel that by the foregoing thoughts I was running against a current; there was no current opposing me. I just made my interlocutors conscious of their latent opinions, which happened to oppose mine.

In my earliest contributions I even ran with the current, which was then to expand the legitimate use of mathematics in economics, a program in which I have never ceased to believe and for which my exemplar is Sir John Hicks. My opposition is to the abuses of mathematics, although they have not caused the greatest harm. The greatest harm could come from the prevalent orientation that allowed as a leading item in the *American Economic Review* a paper about rats (which compelled me to resign from the American Economic Association).

If I finally realized that I was running against one current or another, it was not from any crossing of intellectual swords with my fellow economists, who have systematically shunned such an encounter, but from their personal attitudes toward me. I was a darling of the mathematical economists as long as I kept contributing pieces on mathematical economics. Several things radically changed their mood, especially that of the econometricians.

First, there was my contention that marginal pricing is the worst policy for an agrarian overpopulated economy. Soon after returning to the United States, I informally presented that idea at an after-dinner chat at the University of Chicago. How well I remember that there were absolutely no questions at the end! Those good friends wanted to spare me the embarrassment of being exposed as a neoclassical ignoramus. My position in the profession

worsened irreparably when, owing to the grace of George B. Richardson, my agrarian paper appeared as a leading item in *Oxford Economic Papers* (1960), not only for having thus touched the sacrosanct neoclassical dogma, but especially for pointing out that the much lauded proof by Kenneth Arrow and Gerard Debreu of the existence of a solution of the Walrasian system was irrelevant in practice because it was based on a fantastic premise; that every individual already had an income sufficient for life. My disclosure was hardly mentioned by subsequent writers, *et pour cause*. Yet it must have succeeded *in sotto voce* to alert others to the danger of breaking intellectual bread with Georgescu-Roegen. When quite recently I proposed collaboration on a significant agricultural project to a colleague, he turned me down explaining that he could not renege on his neoclassical testament.

Second were another series of irritating blunders. In *Analytical Economics* (1966) I stated that not all things can be made with the aid of numbers. And in a paper read at the meeting in honor of Corrado Gini (also in 1966), I dared to expose the ineptitude of predicting economic futures by econometric models. That was like signing my death sentence as a fellow of the Econometric Society (to which I had been elected in 1950 when election to fellowship was extremely selective). It was after expressing those articurrent ideas that I received identical treatment from two coeditors of *Econometrica*, E. Malinvaud and J. Dréze. Each sent me a paper critical of one of my articles. In their letters both stated categorically that they *had decided* to publish those papers and that I might, if I so wished, write a small reply (which I did). To my great surprise, both later sent me new versions with notes saying that, after seeing my reply, my critics had modified their initial versions. From Malinvaud I received even a third version together with a pronouncement that I had no proper right to a reply since my critic's paper was not aimed at my own work. After I pointed out that even in that relatively small third version my name appeared not less than twenty-two times, the strange tug of war had to end with the publication of my last reply, but, probably a unique case in the scientific literature, with an additional *replique* by that critic (1963). I am completely correct, I think, in believing that those two coeditors decided to publish the first critical versions because they thought that (without much care) they represented irrefutable blows to my scholarly reputation. But the greatest message of ostracization on the part of my fellow econometricians came on the occasion of my Richard T. Ely Lecture entitled the "Theory of Production" (1969). The Fellows of the Econometric Society scheduled their annual meeting at exactly the same hour as my feature, a machination that I dissected as a prelude to the lecture. This is just one symptom of the modern sociology of scientists.

Third, my idea that has irritate nor only the immense new crop of energetists, but especially most of the economists, was made known at a Distinguished Lecture at the

University of Alabama (1970). It was then that I raised my voice against the neoclassical dogmatic belief that the free mechanism of prices is the only way to ensure rational distribution of resources among, all generations. One pillar of that belief was (and still it) that the interests of future generations are taken care of by the fact that we care for our children, our children for their children, and so forth and so on. Our economic interests have been taken care of (so it seems) by this algorithmic sequence from the time of, say, Julius Caesar - nay, much earlier. Yet none of those propounds thought of asking whether the relation "take care of" is transitive.

I firmly believe in the philosophical idea that our understanding in any domain (including, yes, mathematics) needs both dialectical and arithmomorphic concepts. I cannot even get near the irascible reductionism - everything can be reduced to numbers - that especially dominates the thought of this century. Naturally, I cannot see in a computer anything other than a device to *calculate with numbers* (please, mark those words well) much, much quicker than our brain. About the time I was writing *The Entropy Law and the Economics Process*, a big din was being made about a computer that calculated 1 million decimals of π in eight hours.

As I was writing the present essay, another computer printed out 1 billion decimals! Besides greater speed, nothing has fundamentally changed. In both cases, I believe, the computers used Leibniz's infinite series for $p/4$. And as I said in my volume, if Leibniz had had to calculate by paper and pencil just 1 million decimals, it would have taken him thirty thousand years. How much ink, how much paper, how many quills? Now I wonder whether even the presupposed life of the universe would have sufficed Leibniz for calculating 1 billion decimals. But I am certain that the discovery of any new important theorem, Gödel's, for example, will remain the appanage of the human brain.

Today, "artificial intelligence" is a name so dressed up as to make us easy believers in the fantasy. In my 1971 volume, in considering the claim of that marvelous brain of A. M. Turing, that one day we will no longer be able to determine whether an interlocutor hidden by a screen is a human or a computer, with the proper apology I said that reading Turing's paper convinced me that it may have been written by a computer, that Turing only signed it. I recently sent the same punch to the editor of *Scientific American* in connection with an overenthusiastic article by a staff member. They naturally did not publish it: apparently, the press is free but only for those who own it.

I also contend that the impossibility of relating every function of the brain to some digital or chemical phenomenon is salient proof that we cannot do everything with numbers. The extraordinary experiment by the famous brain surgeon W Penfield pinpoints the mystery.

When Penfield told a patient under brain surgery not to raise his arm if Penfield touched his brain with an electrode, the patient just used the other arm to keep the impeded one down. Surprised, Penfield then asked what electrode caused the second arm to move. We still wait for a nonfantasized answer.

My epistemological addiction is the reason I am against arithmomorpha. I have only words of protest for the typical assertion of a physicist that it is not necessary to explain phenomena before dealing with them mathematically. If one starts only with mathematics, one is likely, as I said, to be trapped inside it. A superb illustration is the theorem of some mathematical economists that the market tends to an equilibrium even if the traders are more numerous than the continuum power. Being trapped, they could not even dream of asking what actual space could have room for so many actual traders.

This has been the story about my claim that I have indeed run against a current, why and how. Other scholars and philosophers have also run against a current. To my knowledge they are Isaiah Berlin, Paul Feyerabend, and Gunnar Myrdal. By comparing their conditions with mine, after long years I have concluded that for the results of one's struggle the place from which one runs against a current matters enormously.

Relevant contributions

“Leontief's System in the Light of Recent Results” *Review of Economics and Statistics*, 32 (August 1950): 214-22.

“Toward Partial Redirection of Econometrics: Discussion”. *Review of Economics and Statistics*, 34 (August 1952): 206-211.

“Mathematical Proofs of the Breakdown of Capitalism”. *Econometrica*, 28 (April 1960): 225-43.

“Measure, Quality, and Optimum Scale” *Sankhya*, Ser. A, 27 (March 1965): 39-64.

Analytical Economics: Issues and Problems. Cambridge, Mass.: Harvard University Press, 1966.

The Entropy Law and the Economic Process. Cambridge, Mass.: Harvard University Press, 1971.

Energy and Economic Myths: Institutional and Analytical Economic Essays. Elmsford, N.Y.: Pergamon Press, 1976.

“Methods in Economic Science”. *Journal of Economic Issues* (June 1979): 317-27; (March 1981): 188-93.

- “The Measure of Information: A Critique” In: J. Rose and C. Bileiu (eds.), *Proceedings of the Third International Congress of Cybernetics and Systems*. New York: Springer Verlag, n.d., 3: 187-217
- “Feasible Recipes Versus Viable Technologies”. *Atlantic Economic Journal* (October 1983): 21-31.
- “Hermann Heinrich Gossen: His Life and Work in Historical Perspective” In: R. C. Blitz, tr., *The Laws of Human Relations and the Rules of Human Actions Derived Therefrom*. Cambridge, Mass.: MIT Press 1983, 11-140.
- “An Epistemological Analysis of Statistics: The Science of Collective Description and of Rational Guessing” In: M. C. Demetrescu and M. Iosifescu (eds.), *Studies in Probability and Related Topics*. Montreal: Nagard, 1983, 221-59.
- “Man and Production”. In: M. Baranzini and L. Scazzieri (eds.), *Foundations of Economics*. Oxford: Blackwell Publisher, 1986, 247-80.
- “Interplay between Institutional and Material Factors”. In: J. A. Kregel, Egon Matzner, and Alessandro Roncaglia (eds.), *Barriers to Full Employments*. London: Macmillan Press, 1988, 297-326.
- “Closing Remarks: About Economic Growth – A Variation on a Theme by David Hilbert”. *Economic Development and Cultural Change*, 36, Suppl. 3 (April 1988): S291-S307

