

Determinação de tarifas em empresas de utilidade pública*

José Ricardo Nogueira**
José Carlos Cavalcanti**

Sumário: 1. Introdução; 2. Visão geral dos argumentos a serem desenvolvidos; 3. Estruturas tarifárias do tipo segundo melhor; 4. Aspectos distributivos da determinação de tarifas; 5. Considerações finais. Palavras-chave: tarifas; utilidades públicas; eficiência.

O principal objetivo deste trabalho é discutir o problema da determinação de preços para as utilidades públicas à luz da moderna teoria econômica. Tomou-se o princípio de determinação de preço pelo custo marginal como referência e assumiu-se a abordagem do segundo melhor como a mais adequada para a discussão. A distinção entre tarifas lineares e não-lineares foi estabelecida em relação a vários esquemas alternativos de preços. Finalmente, as questões distributiva e da estrutura ótima de preços públicos foram incorporadas na análise.

The main purpose of this paper is to discuss the pricing problem for public utilities in the light of modern economic theory. The principle of marginal cost pricing is taken as a reference and the second best approach is assumed to be the most adequate for the discussion. The distinction of linear and non-linear tariffs is established under several alternative pricing schemes. Finally, distributional equity and the optimal structure of public prices are also incorporated into the analysis.

1. Introdução

O sancionamento da Lei de Concessão dos Serviços Públicos à Iniciativa Privada, elemento fundamental para uma retomada de investimentos neste setor no país, em 13 fevereiro de 1995, ocorreu num contexto caracterizado por dois importantes problemas. Em primeiro lugar, muito pouco se conhece, pelo menos na literatura econômica nacional, sobre a estrutura e a dinâmica da infra-estrutura dos *serviços de utilidade pública*, tais como energia, transportes, telecomunicações, saneamento etc. Em segundo lugar, essa lei emerge num ambiente econômico marcado, de um lado, pelo preocupante estado da saúde financeira das empresas estatais que atuam na provisão dessa infra-estrutura, e, de outro, pela pouca experiência do setor privado nacional em atuar em segmentos econômicos tradicionalmente providos pelo governo.

No que diz respeito à superação das dificuldades enfrentadas pelo setor público para que o mesmo reconquiste a capacidade de investir no segmento de infra-estrutura, é possível argumentar que qualquer tentativa de reforma deve passar necessariamente por uma reavaliação da *política tarifária* até agora seguida e pela elaboração de alternativas à mesma. A razão é simples: o setor, para se modernizar, precisa aprender a gerar recursos próprios para sustentar a operacionalização de seus serviços.

* Artigo recebido em jul. e aceito em nov. 1995. Este trabalho faz parte de um conjunto de pesquisas que vêm sendo realizadas pelos autores com apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), da Fundação de Amparo à Ciência e Tecnologia de Pernambuco (Facepe) e do Instituto de Pesquisa Econômica Aplicada (Ipea), aos quais os autores são gratos. Agradecimentos também são devidos a dois pareceristas anônimos pelas sugestões propostas. Os erros e omissões remanescentes são de inteira responsabilidade dos autores.

** Professores doutores do Departamento de Economia da UFPE.

Assim, este trabalho, que não tem a pretensão de oferecer prescrições de natureza normativa, tem como principal objetivo contribuir para uma melhor compreensão do *problema da tarifação* à luz da teoria econômica, tentando esclarecer alguns dos aspectos fundamentais para sua análise a partir de bases mais sólidas. Como a tarifa é, ou tende a ser, o elemento básico do processo de garantir o equilíbrio econômico-financeiro das empresas de utilidade pública, é plenamente justificável uma análise mais detalhada e rigorosa da questão tarifária. Esta análise também se justifica pelo ângulo do setor privado, uma vez que, num contexto em que lhe é concedido o acesso para a provisão de serviços públicos, torna-se fundamental entender como se dá a formação de preços neste segmento econômico.

Para tanto, o estudo está dividido, além desta introdução, em mais quatro seções. Na segunda, estabelece-se o princípio do custo marginal como o ponto de referência de qualquer análise de tarifação e defende-se a abordagem de *segundo melhor* como a mais adequada para a análise em questão. Na terceira seção, a abordagem de segundo melhor é apresentada em relação aos critérios de eficiência econômica e de financiamento. A distinção entre tarifas lineares e não-lineares é discutida, e várias alternativas de estruturas tarifárias são consideradas.

A quarta seção incorpora a questão distributiva na abordagem de segundo melhor, refletindo o fato de que algumas utilidades públicas são vistas como provendo um serviço “de grande interesse social” e onde prevalece o princípio da *universalização do acesso aos serviços*. Finalmente, a quinta seção apresenta os comentários finais.

2. Visão geral dos argumentos a serem desenvolvidos

O conceito de eficiência econômica tem sido a pedra angular do desenvolvimento da teoria econômica, constituindo o critério básico utilizado pelos economistas para escolher entre alocações alternativas de recursos dentro de um sistema econômico. Em termos bem gerais, eficiência econômica significa que o sistema econômico está maximizando a utilização de seus recursos escassos ou minimizando os custos de suas atividades.

A teoria econômica mostra, rigorosamente, que uma condição necessária para garantir eficiência máxima é que os preços dos diversos bens e serviços sejam estabelecidos iguais aos seus respectivos custos marginais. No entanto, na discussão de preços de utilidades públicas há uma tendência de se construir controvérsias ao redor do princípio do preço igual ao custo marginal. Na primeira metade deste século se desenrolou um caloroso debate acerca da propriedade ou não de se utilizar aquele princípio nas empresas de utilidade pública.¹

O ponto-chave da controvérsia se centrava no problema da cobertura dos custos fixos de produção. Pela regra do preço igual ao custo marginal, apenas os custos variáveis seriam cobertos pela receita obtida pela empresa, ficando em aberto a questão do financiamento da parte fixa dos custos totais. Consequentemente, afirmavam alguns autores, seria inadequado aplicar aquela regra em empresas — como as que ofertam utilidades públicas — que apresentam um grande componente de custos fixos em seus custos totais.² Enquanto isso, aqueles que

¹ Não é intenção aqui rever a chamada “controvérsia do custo marginal”. Para uma reconstituição clássica de tal debate, ver Ruggles (1949/50) e Blaug (1990).

² Isto se deve à própria natureza da oferta de utilidades públicas, as quais requerem grandes doses iniciais de investimento (por exemplo, em reservatórios, estações de tratamento de água e redes de distribuição, no caso do setor de saneamento) de modo a garantir por um prolongado período o atendimento da demanda.

advogavam a aplicação da regra em empresas de utilidade pública a defendiam como sendo a menos ruim das soluções, com os custos fixos podendo ser cobertos através de impostos.

Todavia, a moderna teoria microeconômica mostra que a questão principal não é se os preços das utilidades públicas devem ser igualados ou não aos seus custos marginais. É amplamente reconhecido que, devido a certos fatores a serem expostos a seguir, os preços devem divergir dos custos marginais. A questão importante a ser estudada é o nível dessa divergência.

O custo marginal funciona, desta maneira, como o referencial básico da determinação das tarifas. Não há que se questionar aquela determinação baseada nos custos marginais, mas sim analisar como construir um esquema tarifário que inclua a informação sobre custos marginais e atenda ao critério de eficiência econômica.

Ou seja, dado que existem certos elementos estruturais que desaconselham o estabelecimento da regra de preço igual ao custo marginal, qual seria, então, o esquema ótimo de determinação de tarifas de serviços públicos que garantisse o atendimento do critério de eficiência econômica?

A ênfase em relação ao critério de eficiência econômica é um dos aspectos que distinguem a nova visão sobre as utilidades públicas da velha visão. Esta última, prevalecente até o final da década de 60, argumentava que, em face das características especiais das empresas de utilidade pública, particularmente o fato de serem ofertantes de serviços “essenciais” de alta relevância social, não seria conveniente ter-se eficiência econômica como o critério principal de avaliação daquelas empresas. Em outras palavras, dada a natureza social dos serviços ofertados, não se poderia impor sobre as empresas de utilidade pública a exigência de obedecerem a um critério que se aplicaria muito mais ao setor privado da economia. Não apenas isso, mas que tal imposição interferiria no atendimento de critérios superiores, como o de justiça social.

Com o desenvolvimento da moderna teoria econômica do bem-estar, a partir do final dos anos 60 e início dos anos 70, o quadro mudou. Passou-se a reconhecer a necessidade de o setor público — e, portanto, as empresas de utilidade pública, que geralmente estão ou sob o controle direto do governo ou sob seu controle indireto através de regulamentação — também seguir o critério de eficiência econômica, conforme demonstrado, por exemplo, por Diamond e Mirrlees (1971).

Por este raciocínio, parece claro que a questão deveria ser tratada como um problema do tipo segundo melhor.³ A abordagem do segundo melhor considera o caso em que um certo nível de perda de eficiência deve ser aceito de forma a atingir algum outro objetivo. No caso clássico das utilidades públicas, a imposição da cobertura dos custos totais via preços implica perdas de eficiência em decorrência do desvio dos preços em relação aos custos marginais ou devido à aplicação de impostos. É importante observar, no entanto, que o critério de eficiência econômica não é descartado dentro da abordagem do segundo melhor. O objetivo agora é assegurar a minimização das perdas de eficiência em face da restrição adicional imposta sobre a firma.⁴

Este artigo está centrado na abordagem do segundo melhor, reconhecida como a mais adequada para tratar da questão tarifária em empresas de utilidade pública, onde critérios outros que não o de eficiência — autofinanciamento dos custos totais e considerações distributivas, por exemplo — estão geralmente presentes. Nele são consideradas tanto estruturas tarifárias *lineares* quanto *não-lineares*. Em relação às primeiras, o caso clássico é constituído

³ Raciocínio enfatizado por Wiseman (1957), Baumol & Bradford (1970) e Blaug (1990).

⁴ Em outras palavras, o *trade-off* entre eficiência e autofinanciamento é controlado de modo a gerar um mínimo de distorção nos preços, garantindo-se, assim, a maximização do bem-estar líquido social (diferença entre o total de benefícios sociais gerados e o total de custos sociais incorridos).

pelos preços de Ramsey. As principais estruturas não-lineares tratadas são a tarifa de duas partes de Coase, a tarifa de duas partes de Feldstein (a qual incorpora a questão distributiva) e tarifas multipartes.

Como questões de capacidade são altamente relevantes nas utilidades públicas, considera-se também as tarifas do tipo *carga de pico* (*peak load tariffs*), as quais visam a atacar os problemas de potencial de geração dos serviços. Também será considerado qual o custo marginal relevante para o setor, se o de curto prazo ou o de longo prazo, uma vez expostos as principais características do mesmo. Por fim, analisa-se como a questão distributiva pode ser discutida e incorporada em estruturas tarifárias de segundo melhor.

3. Estruturas tarifárias do tipo segundo melhor

Nesta seção considera-se apenas o aspecto de eficiência das estruturas tarifárias quando uma restrição orçamentária é imposta sobre uma dada empresa de utilidade pública. Em termos gerais, é esse o caso que caracteriza a abordagem de segundo melhor em relação às utilidades públicas. Em outras palavras, considera-se o caso em que a política de determinação de preços é condicionada por alguma restrição de financiamento (por exemplo, que a empresa não opere com déficits). Tem-se, então, o seguinte problema de segundo melhor: determinar regras de preços ótimos para uma empresa de multiprodutos, produzindo em condições de retornos crescentes de escala, ofertando produtos passíveis de exclusão do consumo e tendo de gerar uma receita que lhe permita operar sem déficits.⁵

De modo a simplificar a busca de uma solução para esta questão, algumas premissas podem ser estabelecidas. Neste caso, todas as tarifas são assumidas como sendo reguladas pelo governo através de alguma autoridade ou agência regulatória. Tendo isso em mente, estabelece-se a hipótese de que a agência regulatória escolhe aquelas tarifas que levam à maximização da eficiência econômica ou bem-estar econômico em relação ao conjunto de consumidores servidos pela empresa em questão.⁶

A partir desta esquematização geral do problema de segundo melhor, apresenta-se a seguir várias alternativas de estruturas tarifárias que seguem esse esquema.

Tarifas lineares

Estruturas tarifárias lineares, também chamadas uniformes, são aquelas em que o gasto total de consumo para cada indivíduo é proporcional à quantidade comprada. O exemplo clássico de tarifas lineares é dado pelos preços de Ramsey, os quais são preços uniformes que maximizam o bem-estar total sujeito a uma restrição orçamentária que garante o equilíbrio financeiro da empresa.

⁵ Será ignorado o caso em que o governo pode recorrer ao sistema tributário para gerar receitas que cubram um possível déficit da empresa. A análise se concentrará no caso em que a empresa tem apenas seus preços como instrumento de geração de receitas, sem poder apelar para subsídios externos.

⁶ É importante notar que os demais preços praticados no resto da economia, e não controlados pelo governo, são tomados como exogenamente dados. Seguindo Bos (1985), entende-se que este procedimento é adequado quando se está preocupado com a análise do ajustamento de tarifas de serviços públicos em face de uma dada estrutura de preços não-controlados.

Preços de Ramsey

Os preços de Ramsey constituem preços ótimos do tipo segundo melhor, na medida em que garantem a maximização da eficiência econômica da empresa quando esta é obrigada a respeitar uma restrição orçamentária que serve como garantia de sua operação sem prejuízos, ainda que sem receber qualquer subsídio externo.

Dada uma série de hipóteses,⁷ pode-se representar a estrutura básica dos preços de Ramsey como o problema de determinação de preços regulados de utilidades públicas, exposto a seguir:

Maximize $W(\mathbf{P}, R)$, sujeito a $R = 0$, onde:

W = função de bem-estar a ser maximizada;

\mathbf{P} = vetor de preços dos produtos ofertados pela empresa;

R = restrição orçamentária.

A restrição orçamentária R pode ser mais bem-entendida definindo-a como:

$$R = \mathbf{P} \cdot \mathbf{X} - C(\mathbf{X}) = 0 \quad (1)$$

onde:

\mathbf{X} = vetor de produtos ofertados pela empresa;

$C(.)$ = custo total de produção da empresa.

A expressão em (1) diz que os preços praticados pela empresa devem ser tais que cubram os custos totais de produção associados ao vetor de quantidades produzidas.

A solução do problema de maximização acima apresentado gera como condições necessárias para a obtenção de um resultado ótimo a seguinte regra de preços, para um produto em particular:

$$(p_i - c_i) / p_i = \alpha / \epsilon_i \quad (2)$$

onde:

p_i = preço do bem ou serviço i ;

c_i = custo marginal associado ao bem ou serviço i ;

α = número de Ramsey;

ϵ_i = elasticidade-preço da demanda do bem ou serviço i .

As equações em (2) definem a regra fundamental de determinação de preços de Ramsey: o *markup* dos preços sobre o custo marginal deveria ser estabelecido de maneira inversamente proporcional às elasticidades-preço da demanda de cada serviço.

A constante de proporcionalidade α em (2), chamada de “número de Ramsey”, serve para ajustar uniformemente os *markups* em todos os mercados, de modo que a empresa obtenha

⁷ Preços iguais aos custos marginais no resto da economia; funções de demandas independentes; mercado de fatores em condições de concorrência perfeita; inexistência de externalidades no consumo, entre outras.

um equilíbrio entre receitas e despesas. Assim, α é determinado tendo-se em mente a restrição orçamentária R exogenamente imposta sobre a empresa,⁸ e, portanto, os *markups* também dependem do valor escolhido de R .⁹

A estrutura de preços de Ramsey sugere que em certas circunstâncias, como as já descritas, é desejável que se cobrem preços acima do custo marginal, com a diferença em cada caso sendo determinada pelas elasticidades-preço da demanda dos diversos serviços oferecidos pela empresa. Esta última deveria cobrar um preço mais alto, em relação ao custo marginal, para aqueles serviços com demandas inelásticas, do que para aqueles com demandas elásticas, e, assim, minimizar as perdas de eficiência associadas com tal política tarifária de cobertura dos custos totais de produção.¹⁰

Tarifas não-lineares

Tarifas não-lineares são aquelas que variam não somente entre, mas também intramercados (Brown & Sibley, 1986). Assim, consumidores dentro de um mesmo mercado se deparam com uma estrutura tarifária que tem como característica o fato de que os gastos totais de consumo não variam proporcionalmente com a quantidade consumida.¹¹ O preço a ser pago depende, portanto, da quantidade adquirida. Como resultado, abre-se a possibilidade de discriminação de preços entre mercados (isto é, entre diferentes bens e serviços) e entre consumidores em um mesmo mercado.

Uma estrutura tarifária não-linear seria, por exemplo, aquela em que a tarifa paga pelo consumidor fosse composta de uma parte fixa independente da quantidade consumida e de uma ou mais partes variáveis com o patamar de consumo. A tarifa de duas partes de Coase constitui a forma mais simples de tal abordagem tarifária.

Tarifa de duas partes de Coase

Coase (1946) propôs a forma padrão de tarifação de duas partes, com o preço do bem ou serviço ofertado sendo determinado segundo o custo marginal, e uma parte fixa sendo adicionada de modo a garantir a cobertura dos custos totais de produção da empresa.

⁸ α também reflete o fato de que a restrição orçamentária acarreta um custo social, por exemplo, ao desviar fundos que poderiam ser aplicados nos setores não-controlados da economia. Assim, α geralmente assume um valor positivo. Para uma discussão acerca do intervalo de possíveis valores que α poderia assumir, ver Laffont & Tirole (1993).

⁹ Como comparação, tome-se o caso em que o governo utiliza o sistema tributário, na forma de transferências do tipo *lump-sum*, para atacar um possível problema de déficit de uma empresa de utilidade pública. Então, α seria estabelecido igual a zero, e os preços seriam iguais aos custos marginais. Entretanto, empiricamente, é extremamente difícil se encontrar um mecanismo de utilização de transferências do tipo *lump-sum*. Em geral, todos os impostos envolvem perdas de eficiência.

¹⁰ É importante observar que os estudos empíricos sobre o comportamento do consumo de serviços de utilidade pública geralmente apontam para o fato de que a demanda apresenta baixas elasticidades-preço. Para o setor de saneamento, por exemplo, ver a resenha dos estudos internacionais sobre a questão feita por Cavalcanti (1991). Para uma referência sobre um estudo para o Brasil, ver Gabarrone & Reis (1994). Deve-se ressaltar que na maioria dos casos não há substitutos próximos para os serviços oferecidos pelas empresas de utilidade pública, o que explica em grande parte os resultados obtidos.

¹¹ Portanto, descontos e sobretaxas são possíveis de ocorrer dentro desta estrutura tarifária.

Formalmente, ter-se-ia:

$$G(q) = E + p \cdot q \quad (3)$$

onde:

G = gastos totais do consumidor com o consumo do bem ou serviço;

E = parte fixa;

p = preço unitário do bem ou serviço.

Na expressão em (3), E representa a parte fixa que o consumidor tem de desembolsar para poder ter o direito de acesso ao mercado em questão (por exemplo, o direito de ter sua residência ligada à rede de energia elétrica ou ao sistema de telefonia), enquanto $p \cdot q$ representa a parte variável com a quantidade consumida (por exemplo, quilowatts/hora).¹²

O encargo fixo E geralmente é estabelecido de modo a assegurar a cobertura dos custos fixos incorridos pela empresa (barragens, adutoras, redes de distribuição etc.). Para o caso de n consumidores¹³ presentes no mercado, E seria igual ao custo fixo dividido por n .

A eficiência máxima seria obtida com a aplicação dessa estrutura tarifária, dado que cada unidade adicional consumida pode ser vendida ao custo marginal de se produzir aquela unidade adicional, ou seja, o preço p seria estabelecido igual ao custo marginal c . Por outro lado, a firma cobriria suas necessidades de financiamento, uma vez que a receita total seria igual ao custo total.¹⁴

Pode-se mostrar, entretanto, que este resultado não é geral, desde que tarifas não-uniformes nem sempre geram soluções de primeiro melhor (ótimo de Pareto). Mas como Braeutigan (1989) observou, tarifas não-uniformes podem ser usadas para aumentar o grau de eficiência econômica em relação à solução de segundo melhor do tipo uniforme (preços de Ramsey, por exemplo), quando não é possível atingir uma solução de primeiro melhor. Isto se daria através de uma estrutura tarifária não-uniforme que estipulasse preços mais próximos do custo marginal do que os preços de Ramsey.

A importância de tal estrutura de preços para a política fiscal do governo em relação às empresas de utilidade pública reside no fato de que um maior grau de eficiência pode ser alcançado sem a necessidade de se subsidiar a empresa, garantindo a autonomia financeira desta última. Além do mais, essa estrutura faz com que o custo de oferta do produto seja pago pelos que se beneficiam do consumo do mesmo.¹⁵

¹² A fim de se evitar confusão, é preciso ter bem claro que o preço unitário p tanto pode ser tomado como constante (não variando com a quantidade consumida) quanto como variável (com o preço por unidade adicional do bem ou serviço variando com a quantidade consumida). Neste último caso, p é comumente chamado de preço marginal do bem ou serviço. Observe-se, também, que a tarifa de duas partes diz respeito à determinação de preço de dois bens complementares: acesso a um dado mercado e aquisição de um certo montante de um bem ou serviço, uma vez dentro daquele mercado (Varian, 1989).

¹³ Para simplificar a questão, já que não se está interessado aqui na questão distributiva, consideram-se N consumidores idênticos.

¹⁴ A receita total da empresa seria $n(E + p \cdot q)$, enquanto seu custo total seria $F + n \cdot c$, onde F é o custo fixo incorrido pela empresa. Como, por definição, $n \cdot E = F$ e $p = c$, segue-se a igualdade entre a receita total e o custo total.

¹⁵ A quantidade ideal do produto seria, então, alcançada, sendo aquela para a qual o benefício marginal derivado do consumo do bem ou serviço se iguala ao custo marginal de produção do mesmo. Note-se, também, que na tarifa de duas partes de Coase o encargo fixo E equivaleria a um efeito puramente distributivo entre a firma e os consumidores, não afetando o excedente total gerado pelo método de determinação de preços segundo o custo marginal. Em outras palavras, Coase separou o problema de eficiência do problema da cobertura dos custos totais.

Um problema com a tarifa simples de duas partes de Coase se refere ao fato de que, ao se sair de um mundo onde todos os consumidores são idênticos para outro onde os mesmos se diferenciam entre si (em termos de gostos, por exemplo), o encargo fixo E pode vir a ser fixado num patamar tal que resulte na expulsão do mercado de alguns consumidores que estariam dispostos a pagar um preço igual ao custo marginal, mas para os quais o encargo fixo se apresenta como excessivo. Ter-se-ia, então, um *trade-off* entre a cobrança de uma tarifa igual ao custo marginal e a necessidade de se impor o encargo fixo para a cobertura dos custos fixos.¹⁶

Desse modo, a regulação das tarifas deve levar em conta dois aspectos igualmente importantes que compõem aquele *trade-off*:

- a) como determinar um preço que induza um consumo eficiente do serviço ofertado;
- b) como determinar o encargo fixo de forma a fazer com que o número correto de consumidores participe do mercado.

Na ausência de *assimetria de informação*, o órgão regulador poderia facilmente identificar e “catalogar” os indivíduos consumidores e determinar tarifas separadamente para cada consumidor. Conseqüentemente, para cada consumidor, teríamos uma tarifa de duas partes ótima. O conjunto de tarifas de duas partes ótimas assim determinado resultaria em maximização do bem-estar e na cobertura dos custos totais de produção.

Entretanto, o mundo real é permeado pelo fenômeno da *assimetria de informação*. Como resultado, o órgão regulador não dispõe da informação completa necessária para desenhar o conjunto de tarifas individuais de duas partes ótimas.

O resultado ideal sugere, porém, que seria mais vantajoso, em termos de eficiência, diferenciar os consumidores do que cobrar uma tarifa uniforme para todos, mesmo em situações onde não se dispõe de informação completa sobre as características de cada consumidor.¹⁷

A conciliação do objetivo de se cobrar uma tarifa economicamente eficiente com aquele de se garantir a participação no mercado do número correto (ou desejável) de consumidores implica, no entanto, que geralmente o preço deve divergir do custo marginal, resultando em perdas de eficiência.

Na determinação da tarifa de duas partes ótima, o problema é escolher E e p em (3) de modo a maximizar o bem-estar econômico dos indivíduos, condicionado pela restrição orçamentária imposta sobre a empresa.

Seja β um parâmetro que define as características dos consumidores. Pode-se, então, estabelecer as seguintes funções:

$p(q, \beta)$ = função de demanda inversa;

$d(\beta)$ = função de densidade de β ;

$D(\beta)$ = função de distribuição acumulada;

$p(q)$ = tarifa por unidade de produto adquirida por um consumidor.

¹⁶ Em uma linguagem mais formal, tem-se que o aumento potencial de bem-estar (aumento do excedente do consumidor) devido ao movimento da tarifa na direção do custo marginal poderia ser mais do que compensado pelas perdas associadas à cobrança do encargo fixo.

¹⁷ Willig (1978) demonstrou que estruturas de preço não-uniformes são superiores, no sentido de Pareto, às estruturas de preço uniformes. Note-se, entretanto, como observou Varian (1989), que o que se tem é uma melhoria potencial no sentido de Pareto, pois pode-se ter o caso em que o órgão regulador, devido a problemas de informação, não necessariamente escolha a estrutura correta de preços.

Para uma dada quantidade q e um dado preço p existe um valor crítico de β , denominado β' , tal que os consumidores do tipo β' constituem o limite inferior do número total de consumidores participantes do mercado. O limite inferior β' representa aqueles indivíduos que estão dispostos a pagar apenas o preço marginal pela q -ésima unidade do produto.¹⁸

Com estas novas definições, o processo de maximização do bem-estar sujeito à restrição orçamentária resulta na seguinte solução modificada de (2):¹⁹

$$(p_i - c_i)/p_i = m/\epsilon_i$$

onde:

$$m = \alpha / (1 + \alpha) \quad (4)$$

Portanto, a tarifa de duas partes ótima seria formada pelo encargo fixo E mais um preço dado pela expressão (4) acima. Este resultado engloba a regra de Ramsey dada por (2), refletindo o fato de que o *markup* do preço sobre o custo marginal continua a depender do inverso da elasticidade-preço da demanda.²⁰

O modelo acima descrito enfatiza a importância da informação sobre a distribuição de frequência de β . Não é necessário, nem é factível, que o órgão regulador classifique os consumidores individuais segundo seus tipos, mas tão-somente que conheça a distribuição dos consumidores de acordo com β .²¹

Em relação ao encargo fixo E , é de se supor que, no que diz respeito a serviços tão essenciais quanto saneamento ou energia elétrica, mesmo com E sendo fixado em um patamar alto, dificilmente um indivíduo consumidor optaria por deixar o mercado.

Pode-se agora generalizar a abordagem da tarifa de duas partes de modo que, para uma dada distribuição de β , os consumidores sejam alocados ao longo de diferentes blocos tarifários. Tal método caracteriza as tarifas multipartes.

Tarifas multipartes

As tarifas multipartes se referem àquelas estruturas tarifárias não-uniformes constituídas por um número n finito de blocos tarifários, com $n \geq 2$. Cada bloco tarifário é composto de um encargo fixo E , comum a todos os blocos,²² e pelo menos um preço p , que varia com a quantidade adquirida q .

¹⁸ Para os consumidores abaixo do limite dado por β' , a melhor opção é sair do mercado, dado que o custo de permanecer no mesmo supera o benefício gerado pelo consumo do produto em questão.

¹⁹ Para uma demonstração formal deste resultado, ver Brown & Sibley (1986) e Navajas & Porto (1990). Note-se que os preços em (4) são ótimos em termos de segundo melhor.

²⁰ A expressão em (4) mostra que, apesar de os preços (uniformes) de Ramsey serem inferiores, no sentido de Pareto, aos preços não-uniformes, a regra do inverso da elasticidade nela embutida ainda permanece como um elemento crucial na determinação de preços em estruturas não-lineares.

²¹ Em certos casos é possível ter-se acesso à informação sobre β sem grandes dificuldades. Por exemplo, em relação à demanda por energia elétrica residencial, a renda familiar é um elemento crucial na determinação do consumo (Brown & Sibley, 1986).

²² Isto se deve ao fato de que talvez seja ilegal cobrar encargos fixos diferenciados para diferentes consumidores. A taxa de acesso ao mercado deveria refletir a condição de que todos têm o mesmo direito de participar do mesmo.

Uma estrutura geral para as tarifas multipartes, T , seria:

$$T(q) = E + \sum_{i=1}^{n-1} p_i q_i + p_n (q - q_{n-1}), \quad q \geq q_{n-1} \quad (5)$$

Assim, cada consumidor pagaria um total dado pela soma do encargo fixo mais o valor correspondente à quantidade efetivamente consumida. Este último valor, por sua vez, seria determinado de acordo com a faixa de consumo em que o consumidor se situa. Como ilustração, para o caso de $n = 2$ (ou seja, $i = 1, 2$) o consumidor pagaria um total composto da parte fixa E , do preço correspondente à faixa de consumo $i = 1$ e do preço correspondente à faixa de consumo $i = 2$.

A idéia por trás das tarifas multipartes é permitir que a empresa (ou o agente regulador) tenha mais graus de liberdade, em relação à tarifa de duas partes de Coase, para cobrar dos consumidores uma tarifa segundo seus perfis de consumo. Dada a informação disponível sobre a distribuição de consumidores por tipo β , a tarifa multiparte seria estabelecida através do cálculo dos preços correspondentes aos diversos blocos de consumo segundo o custo marginal associado a cada um desses blocos, com o encargo fixo sendo estabelecido de forma a que os custos totais sejam efetivamente cobertos.

Alternativamente, poder-se-ia conceber a tarifa multiparte em termos semelhantes ao caso dos preços de Ramsey, só que incorporando a diferenciação por blocos de consumo.²³ Neste caso, a tarifa seria cobrada de acordo com a seguinte formulação:²⁴

$$[p(q) - c(q)]/p(q) = m/\epsilon(q, p(q)) \quad (6)$$

A fórmula em (6) sugere uma estrutura ótima de preços onde a cada bloco de consumo é associado um determinado preço, sendo que o conjunto de preços é tal que o bem-estar geral é maximizado e os custos totais inteiramente cobertos. A regra de Ramsey permite, então, descontos e sobretaxas, se o preço para um dado bloco for menor ou maior do que o custo marginal correspondente.²⁵ Seguindo-se a fórmula de preços dada por (6), a estratégia seria fazer com que aqueles indivíduos compondo os grupos com demandas inelásticas contribuam em um grau maior para a cobertura dos custos totais da empresa.

É possível também combinar a estrutura da tarifa de duas partes com aquela dos preços de Ramsey, e este é o caso desejável quando se está interessado na provável interação entre a parte fixa E e o vetor de preços \mathbf{p} . Primeiramente, observe-se que E diz respeito à decisão de entrar ou não no mercado em questão (de água ou telefone, por exemplo), enquanto os diferentes preços dizem respeito à decisão de quantas unidades se quer adquirir. Tem-se, assim, dois mercados a considerar: o de participação e o de consumo.

Até aqui supôs-se que não haveria qualquer interação entre os dois mercados. Entretanto, numa visão mais realista, é óbvio que os dois mercados não estão completamente dissociados. Por exemplo, se E é fixado num patamar tal que algum grupo de indivíduos opte por não participar do mercado de água ou telefone, seu consumo daqueles serviços cai a zero, influenci-

²³ Esses preços são conhecidos também como preços de Ramsey não-uniformes (Brown & Sibley, 1986).

²⁴ Assumindo-se que as demandas por unidades adicionais de consumo sejam independentes.

²⁵ Isto depende do formato da curva de demanda de cada indivíduo (ou grupo de indivíduos) e da frequência de distribuição de β .

ando o mercado de consumo de água ou telefone. Por outro lado, se o preço cobrado pelo consumo do serviço aumenta, a quantidade consumida é reduzida e o grupo de consumo marginal se retira do mercado, influenciando, portanto, o mercado de participação.

Conseqüentemente, é necessário que se considerem os efeitos cruzados entre os dois mercados. Neste caso, a expressão em (6) deve ser modificada da seguinte maneira:²⁶

$$[p(q) - c(q)]/p(q) = \{\alpha/\varepsilon(q, p(q))\} \cdot \{[1 - q'(p, \beta_0)]/Q\} \quad (7)$$

onde:

Q = consumo médio;

q' = consumo do consumidor marginal;

β_0 = consumidor marginal.²⁷

O primeiro termo entre colchetes no lado direito de (7) reflete os efeitos cruzados entre o mercado de participação e o de consumo. Como observado por Brown & Sibley (1986), este termo obriga a empresa ou o agente regulador a considerar o fato de que quando o preço de consumo sobe, não apenas cai o consumo, como cai também a participação. Isto contribui para que o *markup* do preço sobre o custo marginal permaneça num patamar mais baixo em relação ao caso em que os efeitos cruzados entre os mercados não sejam levados em conta.

Uma abordagem alternativa à até aqui focalizada de maximização de bem-estar é aquela que considera apenas os ganhos de eficiência que podem ser alcançados a partir de uma dada situação inicial. Esta abordagem se refere àquelas estruturas tarifárias que são ditas superiores no sentido de Pareto, isto é, aquelas que fazem com que o bem-estar geral aumente sem prejudicar ninguém.

Por essa visão, é possível tomar o sistema tarifário prevalecente em um certo período como base para um processo de reforma que leve a um aumento da eficiência econômica, sem se objetivar, entretanto, a maximização de bem-estar, ou seja, ganhos graduais de eficiência é que são perseguidos.

O *rationale* básico desta visão está em que uma estrutura tarifária de várias partes pode ser vista como envolvendo um conjunto de tarifas de duas partes, conjunto este gerado a partir de uma dada tarifa uniforme inicial.²⁸

Suponha-se que haja n consumidores de serviços de água ou telefone e que, inicialmente, um preço uniforme seja cobrado dos consumidores pelo consumo do serviço em questão. É possível mostrar que, ao se estabelecer um vetor \mathbf{p} de preços e um vetor de encargos fixos \mathbf{E} , tal que²⁹

²⁶ Para uma derivação formal deste resultado, ver Brown & Sibley (1986, ap. do cap. 4).

²⁷ O consumidor, ou grupo de consumidores, marginal representa o "ponto de corte", entre aqueles que participam e os que não participam do mercado.

²⁸ Duas importantes contribuições teóricas sustentam esta visão. Faulhaber & Panzar (1978) mostraram que uma tarifa ótima de $n + 1$ partes é mais eficiente do que uma tarifa ótima de n partes, enquanto Willig (1978) demonstrou que uma tarifa de n partes pode ser construída a partir de um dado preço inicial, tal que o total de bem-estar é aumentado em relação ao caso da cobrança da tarifa uniforme inicial.

²⁹ Note-se que o exemplo a seguir se refere ao caso de um sistema de tarifas de blocos crescentes, ou seja, com preços crescentes para sucessivas unidades adicionais de consumo. Seria perfeitamente possível lidar com o caso oposto, isto é, com um sistema tarifário de blocos decrescentes envolvendo descontos de preço para sucessivas unidades adicionais de consumo.

$$\begin{aligned}
 p_1 &> \dots > p_n \geq c \\
 0 &= E_1 < \dots < E_n \\
 E_i &= q_i(p_1)(p_1 - p_i), \quad i = 1, \dots, n
 \end{aligned}$$

Os pares (E_i, p_i) formariam, então, um conjunto de tarifas de duas partes que pode ser visto como uma estrutura tarifária multiparte.³⁰ Esta última é superior à tarifa uniforme ou a uma tarifa multiparte com um número menor de blocos de consumo.³¹

Como resultado, pode-se argumentar que é possível aumentar o bem-estar geral quando se passa de uma dada tarifa uniforme inicial para uma estrutura tarifária não-linear.

Este resultado depende crucialmente, no entanto, de o sistema de tarifas ser compatível em termos de incentivos (*incentive compatible*) individuais. Com isto, quer-se dizer que o *menu* de tarifas de duas partes oferecido aos consumidores deve ser tal que estes últimos sejam induzidos a selecionar a tarifa que seja mais adequada às suas preferências individuais, isto é, que os consumidores tenham incentivo para manifestar suas verdadeiras preferências (em termos dos reais benefícios que adquirem do consumo do serviço em questão) no mercado através da escolha do par (E_i, p_i) .³²

Neste caso, certos possíveis conflitos entre a empresa e o agente regulador podem ser minimizados uma vez que não há perdedores, em relação à situação anterior, com a nova estrutura tarifária: esta última representa, como já observado, uma melhoria no sentido de Pareto.³³

A seguir, considera-se o problema de capacidade associado ao desenho de sistemas de tarifas ótimas e a questão de qual conceito de custo marginal, de curto prazo ou de longo prazo, é o mais adequado.

O problema da capacidade

As empresas de utilidade pública são caracterizadas pelos grandes investimentos em capacidade produtiva necessários para assegurar uma oferta confiável³⁴ dos serviços. Segue-se, então, que a decisão inicial de quanto investir em capacidade é crucial para aquelas empresas. Normalmente, um estudo de previsão do comportamento do mercado (por exemplo, número de consumidores ou unidades de consumo a ser atendido, quantidade média de consumo por consumidor ou unidade de consumo etc.) serve para estabelecer as necessidades iniciais de investimento em capacidade.³⁵

³⁰ Formalmente, uma tarifa de multiparte seria o contorno inferior do conjunto de tarifas duas partes.

³¹ Este resultado depende de certas hipóteses, como, por exemplo, a de que as funções de demanda obedecem ao requisito de monotonicidade forte, ou seja, a quantidade demandada decresce à medida em que o preço sobe.

³² Neste caso, o consumidor revela o quanto ele está realmente disposto a pagar (*willingness to pay*) por determinada quantidade do serviço.

³³ Por exemplo, se o agente regulador propõe uma dada tarifa uniforme ou uma dada tarifa de n partes, a empresa pode tomar aquela sugestão como base para a construção de uma tarifa Pareto-dominante que represente ganhos de eficiência para a empresa, sem prejudicar nenhum consumidor. Politicamente, esta solução seria mais adequada aos olhos do agente regulador do que uma outra estrutura que implicasse ganhadores e perdedores.

³⁴ Confiável no sentido de atender a todos os demandantes do serviço em qualquer período.

³⁵ Está-se supondo aqui que a decisão é de quanto investir, e não de se o investimento deve ser realizado ou não. A razão de se adotar essa hipótese é que o serviço a ser ofertado é considerado essencial e indispensável à população, de modo que o investimento tem de ser realizado.

Como ilustração, o serviço de abastecimento de água é geralmente concebido de forma a se ter uma capacidade de atendimento das necessidades de consumo da população por um determinado período, digamos 30 anos, a partir da entrada em operação do sistema de abastecimento.

Mas o problema da capacidade não se resume à determinação do investimento inicial. A demanda por diversos serviços (como energia elétrica e água) é caracteristicamente cíclica, variando, por exemplo, de acordo com a hora do dia ou a época do ano. Certas horas do dia e certos meses do ano apresentam um crescimento acentuado no consumo do serviço, impondo uma pressão adicional sobre o sistema, que deve estar preparado para tal situação.

Estar preparado para enfrentar os chamados períodos de pico da demanda, quando a capacidade instalada é utilizada plenamente ou próxima de tal nível, implica custos de capacidade ou de confiabilidade,³⁶ e surge então a questão de como cobrar por estes custos: as tarifas deveriam ser as mesmas para qualquer período ou deveriam variar com o mesmo? É esta a questão abordada pela literatura sobre tarifas de carga de pico (*peak-load tariffs*).³⁷

Tarifas de carga de pico³⁸

Uma estrutura tarifária de carga de pico é aquela que tenta estabelecer preços para demandas que variam de acordo com o período. A demanda total é dividida em classes de demanda segundo o período a que se referem, e a questão tarifária envolve o uso eficiente da capacidade produtiva existente de modo a atender àquelas classes de demanda, tomadas separadamente (Williamson, 1966).

Para assegurar o uso eficiente da capacidade produtiva instalada, a empresa pode recorrer à discriminação de preços, sendo a tarifa de carga de pico mais um exemplo de uma estrutura de preço não-uniforme. Em um setor como o de energia ou de telefonia, dado que o produto é não-armazenável, a discriminação tarifária segundo um período de consumo é aconselhável para se evitar “congestionamento” no sistema em certos períodos.

O problema de capacidade se apresenta como importante quando a demanda total supera as possibilidades planejadas de atendimento, ou seja, quando a capacidade do sistema se torna insuficiente para atender ao aumento no consumo total. Aumento populacional e ampliação da parcela de uma dada população que é atendida com serviços de utilidade pública são dois fatores, entre outros, que podem causar essa situação. Em tais casos, tem-se de optar entre racionar a oferta do serviço (via preço ou quantidade) ou investir em capacidade adicional.

Suponha-se um sistema de abastecimento de água³⁹ que apresenta custos operacionais b e custos de capacidade d , ambos definidos por unidade de produto e por período.⁴⁰ Seguindo Williamson (1966), pode-se definir o custo total como:

³⁶ Ver nota 2.

³⁷ Tais tarifas são também conhecidas como horossazonais.

³⁸ Para uma resenha desta literatura, ver Crew & Kleindorf (1979), Williamson (1966) e Bos (1985).

³⁹ Assume-se, por enquanto, para efeito de simplificação, que o sistema de abastecimento é completamente divisível.

⁴⁰ Os custos operacionais b podem ser vistos como os custos marginais de curto prazo. Quando a empresa incorre em custos de capacidade d , a soma $b + d$ pode ser encarada como os custos marginais de longo prazo.

$$CT = (b + d)Q \quad (8)$$

onde Q é o produto total do serviço.

As condições de primeira ordem para a maximização de uma função de bem-estar social definida como a diferença entre benefícios sociais e custos sociais resultam em que o preço deve ser igual ao custo marginal. Portanto, $p = b + d$, onde p é o preço por unidade do produto.

Suponha-se, agora, que o sistema de abastecimento esteja operando em um nível de produção tal que $Q \leq Q'$, onde Q' representa a capacidade física máxima de produção do serviço em questão. Neste caso, a empresa não está incorrendo em custos de capacidade, ou seja, $d = 0$.

Maximizando a função de bem-estar W levando em conta a restrição $Q \leq Q'$, resulta em que $p = b$. Em outras palavras, o preço se iguala ao custo marginal de curto prazo.

Entretanto, a partir do momento em que a capacidade total instalada é atingida, qualquer pressão de demanda sobre o sistema de abastecimento faz com que a empresa passe a incorrer em custos de capacidade, de modo que o preço passa a ser definido como $p = b + d$, como resultado do fato de que o sistema de abastecimento necessita ser ampliado.

Uma hipótese mais realista seria a de indivisibilidade do sistema de abastecimento, ou seja, que o mesmo é construído com uma capacidade física em termos de múltiplos inteiros das unidades do produto ofertado.

A principal diferença para o caso anterior de um sistema perfeitamente divisível reside em que, agora, o equilíbrio estático de longo prazo é caracterizado por lucros líquidos diferentes de zero.⁴¹ Ter-se-ia, então, o problema de como financiar a empresa caso seus lucros líquidos sejam negativos, ou de se saber se no caso de lucros líquidos positivos haveria algum tipo de tributação sobre os mesmos.

No tocante ao primeiro caso, que é o que mais interessa tendo em vista as discussões anteriores, a empresa, em vez de ser subsidiada, poderia recorrer a uma tarifação de duas ou várias partes, com a parte fixa da tarifa sendo usada para cobrir o déficit financeiro.

Considerou-se apenas o caso em que o problema de capacidade é resolvido através da ampliação do sistema. A literatura recente considera, entretanto, uma outra alternativa, que é o racionamento da capacidade⁴² existente entre os consumidores. O resultado principal dessa literatura sugere que pode ser melhor, em termos de bem-estar, aceitar a existência de excesso de demanda do que gastar o “dinheiro público” na ampliação da capacidade física das utilidades públicas ou aumentar em demasia as tarifas do serviço. Consequentemente, racionamento de quantidade pode ser mais eficiente do que racionamento via preço ou ampliação de capacidade (Bos, 1985).

Por esta abordagem, pode-se mostrar que é eficiente ter-se excesso de demanda, isto é, racionamento, para o caso em que haja flutuações de demanda dentro de um determinado período.

A seguir, abordar-se-á a questão da escolha entre os conceitos de custo marginal de curto prazo ou custo marginal de longo prazo como base para a determinação de tarifas de serviços de utilidade pública.

⁴¹ Mesmo assumindo-se retornos constantes de escala, a empresa opera com lucros líquidos positivos ou negativos. Como Williamson (1966) observou, apenas acidentalmente a empresa obterá lucro zero no equilíbrio.

⁴² Pode-se entender a demanda pelo serviço em questão como uma demanda por capacidade de produção.

A questão do conceito apropriado de custo marginal

Paralelamente à discussão sobre tarifação baseada no custo marginal há um outro debate sobre o conceito mais apropriado de custo marginal, se o de *curto prazo* ou o de *longo prazo*, em relação à determinação de tarifas de serviços ofertados por empresas de utilidade pública.

Por exemplo, a abordagem seguida pelo Banco Mundial para preços no setor público compreende dois passos: primeiro, o cálculo do custo marginal; em segundo lugar, o ajustamento daquele cálculo inicial de modo a levar em consideração outros fatores tais como financiamento, objetivos distributivos etc. Mas, então, surge a questão: qual conceito de custo marginal deve ser usado no cálculo envolvido no primeiro passo?

A prática seguida pelo Banco Mundial não é bem clara quanto ao conceito apropriado de custo marginal. Segundo Heady (1989), aquela prática não tem sido uniforme nos trabalhos oriundos do Banco Mundial. Por exemplo, Walters (1968), considerando o caso de tarifas de utilização de rodovias, recomenda o uso de custo marginal de curto prazo, enquanto que Munasingh (1981), lidando com tarifas de energia elétrica, privilegia o uso de custo marginal de longo prazo.

A discussão centra-se fundamentalmente sobre se os custos de capital devem ou não ser incluídos na definição de custo marginal. Heady (1989) apresenta uma lúcida discussão sobre essa questão, e o restante desta subseção baseia-se grandemente na exposição deste autor.

Uma regra seria a que diz que se a empresa opera com excesso de capacidade (capacidade ociosa), de forma que um aumento na demanda não requer investimentos adicionais de capital, o uso do custo marginal de curto prazo seria mais adequado.

Considere-se o caso em que o estoque de capital (a usina de produção e sua capacidade) é fixo. Variações na oferta do serviço só podem ser obtidas, então, através de variações na quantidade utilizada de insumos variáveis. Assim, o custo social marginal de se produzir uma unidade adicional do serviço é dado pelo custo marginal das unidades adicionais dos insumos variáveis necessárias para ofertar aquela unidade extra do serviço.

No longo prazo, no entanto, unidades adicionais do serviço podem ser obtidas por meio de variações no estoque de capital (capacidade instalada). Acontece que variações no estoque de capital constituem uma decisão de investimento, não de preço. Decisões “puras” de preço seriam aquelas que considerassem um aumento da tarifa como forma de enfrentar uma demanda superior à capacidade instalada. É preciso, portanto, ter em mente que variações de capacidade envolvem decisões de investimento, não sendo meramente decisões a respeito de preço.

Considere-se agora o caso em que as necessidades de atendimento da demanda superam a capacidade planejada da usina de produção. Se a tarifa fosse determinada segundo o custo marginal de longo prazo, é facilmente percebido que não haveria equilíbrio no mercado (entre oferta e demanda) na ausência de racionamento. Isto porque haveria um hiato temporal entre o momento da decisão de variar o estoque de capital e o momento em que a capacidade adicional de produção entrasse em operação. Por outro lado, determinando-se a tarifa segundo o custo marginal de curto prazo, que é o preço que equilibra oferta e demanda no mercado, ter-se-ia uma eficiente alocação do produto e uma eficiente utilização da capacidade instalada.

Poder-se-ia também ter o caso em que uma insuficiente capacidade instalada refletisse o fato de que o custo marginal de curto prazo fosse maior do que o custo marginal de longo prazo. Determinar a tarifa de acordo com este último significaria que a tarifa cobrada seria

menor do que o custo social associado à oferta de unidades adicionais do serviço, o que é claramente ineficiente.⁴³ A regra, portanto, é: uma vez que os investimentos tenham sido realizados, tarifas baseadas no custo marginal de curto prazo constituem o meio mais eficaz de se fazer com que a capacidade instalada seja usada de modo mais eficiente.

Mas a utilização do custo marginal de curto prazo como base para a determinação tarifária também é objeto de críticas. A objeção mais importante é que esse método produz grandes variações de preços.⁴⁴ Observe-se que são estas variações que asseguram que se tenha uma utilização eficiente da capacidade existente; no entanto, elas resultam em dois problemas importantes.

O primeiro são os custos administrativos e políticos resultantes das variações das tarifas. Mudanças frequentes dos níveis tarifários implicam constantes cálculos, revisões de cálculos, preenchimentos de formulários, divulgação das novas tarifas etc. Além do mais, talvez seja politicamente desinteressante para o governo submeter a população a contínuas mudanças de tarifas.

O segundo problema está afeto diretamente ao bem-estar dos consumidores. Mudanças frequentes de preços criam um clima de incerteza entre os consumidores em relação aos possíveis gastos com o serviço, levando, por exemplo, a um aumento do custo de aquisição de informação sobre as tarifas. Tal situação de incerteza pode afetar negativamente as decisões de investimento dos agentes econômicos.⁴⁵

Quando uma restrição é imposta sobre a empresa, no sentido de que as tarifas devam permanecer constantes por um determinado período — por motivos políticos, por exemplo —, surgem implicações negativas para a utilização do conceito de custo marginal de curto prazo.⁴⁶

Seguindo-se o critério do custo marginal de curto prazo, o nível das tarifas varia de acordo com o nível de utilização da capacidade instalada. À medida que o ponto de plena capacidade se aproxima, as tarifas tendem a aumentar de forma a reduzir a demanda e controlar os custos de se prover unidades adicionais do serviço. Quando novos investimentos em capacidade são realizados, gerando excesso de capacidade, as tarifas tendem a diminuir de maneira a induzir um aumento da demanda.

Por outro lado, adotando-se o critério do custo marginal de longo prazo, as tarifas não se relacionam diretamente com o nível de utilização da capacidade instalada. O ponto crucial aqui é determinar qual o período adequado para a realização de novos investimentos em expansão da capacidade a um custo mínimo.

⁴³ Como Heady (1989:4) observou, o fato de que o nível do serviço ofertado poder ser aumentado a um menor custo por meio da expansão da capacidade não constitui um argumento para baixar a tarifa antes do aumento de a capacidade ter ocorrido.

⁴⁴ Os custos marginais de longo prazo também variam no decorrer do tempo, mas não tanto quanto os custos marginais de curto prazo.

⁴⁵ Principalmente em relação àquelas decisões que são complementares ao consumo do serviço. Como ilustração, pode-se citar o caso em que uma empresa, ao escolher o local de instalação de uma fábrica que necessita do insumo água para a sua produção, leva em conta as opções em termos da confiabilidade da oferta de água e o custo de tal serviço entre várias localidades.

⁴⁶ Uma maneira de se lidar com o problema da variabilidade das tarifas poderia ser a realização e divulgação de estudos de previsão de mudanças dos níveis de preço. Entretanto, além do custo inerente a tais estudos, têm-se a dificuldade de se fazer a previsão de mudanças na demanda, a parca informação disponível sobre a relação entre o comportamento dos custos marginais e o nível do produto, e a dificuldade em se saber ao certo quando a expansão em capacidade estará pronta para ser usada.

Portanto, a regra de preço a ser usada depende da avaliação de uma série de fatores, sendo necessário manter-se alerta para as circunstâncias específicas de cada caso a ser analisado.

A seguir será discutido como a introdução de considerações distributivas afeta a análise da determinação de tarifas.

3. Aspectos distributivos da determinação de tarifas

Após o período que ficou denominado como a “controvérsia da determinação do preço pelo custo marginal”, nos anos 40, a idéia predominante ficou sendo a de que a regra do preço igual ao custo marginal seria a ideal, pois através dela atinge-se a maximização da eficiência econômica. Aquela mesma regra ainda serviria como a referência básica para a avaliação de esquemas de preço alternativos.

Não houve consenso, entretanto, quanto à forma de financiar os déficits das empresas de utilidade pública e outras que operassem com tecnologia exibindo retornos crescentes de escala, para as quais os custos médios são maiores que os marginais. A restrição orçamentária imposta sobre os modelos de determinação de preços em empresas de utilidade pública, como forma de assegurar seu equilíbrio financeiro, passou a dominar as contribuições teóricas sobre a questão tarifária, resultando em várias contribuições que visavam a buscar a melhor forma de os preços se desviarem do custo marginal, por causa daquela restrição adicional, ao mínimo custo possível para o sistema econômico.

Uma característica comum à maioria das contribuições teóricas a respeito de empresas de utilidade pública é considerar que o objetivo de tais empresas é a maximização do bem-estar social. É amplamente reconhecido, entretanto, que a maximização do bem-estar, definido usualmente em termos de eficiência econômica (ótimo de Pareto), não garante que, do ponto de vista de justiça social, atinja-se a posição ideal⁴⁷ de distribuição do total de bem-estar entre os indivíduos ou grupos de indivíduos que formam a sociedade.

A teoria econômica que embasa a análise da determinação de tarifas tem se concentrado fortemente sobre a questão da fixação de preços eficientes, isto é, aqueles preços que garantem a maximização do bem-estar total (definido pela soma dos excedentes dos consumidores e dos produtores), não importando como aquele total está distribuído entre os indivíduos e, conseqüentemente, não se atribuindo relevância à posição relativa dos indivíduos dentro da sociedade e não se preocupando com possíveis conseqüências indesejadas do ponto de vista distributivo que os diferentes sistemas de preço possam trazer.⁴⁸

As razões para a desconsideração da questão-distributiva quando da análise de esquemas tarifários podem ser as mais diversas. Uma delas pode ser o desejo de se concentrar em modelos simples, que possam gerar regras claras e diretas para a formulação de políticas de pre-

⁴⁷ Qual o significado do termo “posição ideal” é uma questão de difícil resolução, pois entra-se no território das chamadas escolhas sociais (formalmente, a especificação de uma função de bem-estar social), uma área pontuada de controvérsias e quase nenhum consenso. Não se deterá aqui sobre essa questão, assumindo-se apenas que a sociedade, por algum mecanismo não especificado de escolha social, elegeu uma dada função de bem-estar social representativa dos objetivos que aquela sociedade considera como sua meta principal.

⁴⁸ A teoria econômica mostra que posições de maximização de eficiência são perfeitamente compatíveis com extremas desigualdades de renda entre os indivíduos. Pode-se ter, por exemplo, uma situação em que uma determinada estrutura tarifária promova a alocação eficiente de recursos, mas ao mesmo tempo impeça o acesso ao serviço oferecido por parte de algumas camadas da população.

ços. Uma outra razão pode estar na pressuposição de que o perfil da distribuição de renda já é justa ou aceitável (segundo algum critério social).

Outra razão pode ainda ser encontrada na visão de que a questão distributiva seria mais bem resolvida fora do sistema de preços (por exemplo, através de políticas compensatórias, subsídios diretos às camadas mais pobres etc.).⁴⁹

A importância da questão distributiva em relação a determinadas utilidades públicas, como o setor saneamento, reside basicamente nos fatos de que os serviços de água e esgoto são essenciais para a população, não podendo deixar de ser ofertados, e por princípio, todos devem ter acesso aos mesmos.

Como as tarifas são essencialmente instrumentos equilibradores dos níveis de oferta e demanda, elas podem, se estabelecidas em um patamar muito alto, excluir do consumo um certo número de consumidores. Já que exclusão do consumo dos serviços de utilidade pública não é desejável, a política tarifária praticada pelo setor é de fundamental importância para possibilitar o acesso universal àqueles serviços sem comprometer a restrição de equilíbrio financeiro das empresas.

Portanto, tem-se que a questão distributiva deve ser levada em consideração simultaneamente com a questão de eficiência econômica e de financiamento. Nesta seção, procurar-se-á mostrar uma forma de se abordar o problema distributivo nas tarifas,⁵⁰ a chamada *tarifa de duas partes de Feldstein*.

A tarifa de duas partes de Feldstein

É fácil perceber que a aplicação da regra de Ramsey (ver equação (2)) à determinação de tarifas apresenta um efeito distributivo altamente regressivo na medida em que aqueles grupos de consumidores com demandas menos elásticas, e que, portanto, deveriam, segundo a regra, contribuir mais para o financiamento dos custos totais da empresa, são geralmente mais pobres. A descon sideração deste impacto distributivo negativo, implícito na regra de Ramsey, fez surgir críticas à sua utilização como único critério de determinação de tarifas ótimas.

Em relação à tarifa de duas partes de Coase, tem-se que o encargo fixo cobrado independente do nível de renda do consumidor. Isto implica que o pagamento desse encargo pode ser bastante oneroso para os consumidores de baixa renda, tornando-o regressivo.

Feldstein (1972a; 1972b) propôs uma estrutura de determinação de preços ótimos que incorpora, além dos critérios de eficiência e equilíbrio financeiro, considerações sobre justiça distributiva. Em outras palavras, Feldstein incorporou o aspecto distributivo explicitamente na determinação de tarifas ótimas.⁵¹

⁴⁹ Na implausibilidade de se poder utilizar transferências do tipo *lump-sum* (que não afetam as condições de eficiência), o que é amplamente reconhecido na literatura, esta opção pela resolução das questões distributivas exogenamente ao sistema de preços necessitaria de uma regra ou conjunto de regras que especificasse um mecanismo distributivo que não implicasse perdas de bem-estar maiores do que aquelas resultantes, por exemplo, da aplicação de impostos para realizar-se políticas redistributivas.

⁵⁰ Assim, não lidaremos com a opção de se utilizar do sistema tributário para atacar os objetivos de justiça social.

⁵¹ O impacto distributivo embutido na regra de preço igual ao custo marginal já havia sido notado, entre outros, por Hotelling (1938), Coase (1946) e Henderson (1947). Entretanto, antes de Feldstein ninguém havia incorporado explicitamente a questão distributiva na determinação de preços ótimos.

Como requerimento básico para a formulação de uma estrutura tarifária nos moldes propostos por Feldstein, há a necessidade de se especificar parâmetros de demanda e de distribuição de renda, com as considerações distributivas estando ligadas à relação entre a utilidade social marginal da renda e o nível de renda entre os indivíduos ou grupos de indivíduos.⁵²

A regra especificada para representar aquela relação é de que a utilidade social marginal da renda diminui à medida em que a renda aumenta, implicando um peso maior a ser dado ao bem-estar daqueles grupos de indivíduos com rendas mais baixas (Feldstein, 1972a; 1972b).

Seja $S(p, Y)$ o excedente do consumidor, que é a medida de seu bem-estar individual, com p sendo o preço por unidade do serviço e Y a renda do indivíduo consumidor. Subtraindo-se de S o montante E que o consumidor paga como encargo fixo, obtém-se o excedente líquido do consumidor. Assim, em termos agregados, tem-se a seguinte função de bem-estar a ser maximizada:

$$W = N \int_0^{\infty} \{S(p, Y) - E\} u'(Y) f(Y) dY \quad (9)$$

onde:

N = número de consumidores;

$u'(Y)$ = utilidade social marginal da renda do indivíduo de renda Y ;

$f(Y)$ = função de densidade relativa representando a distribuição de rendas individuais.

A expressão em (9) representa a soma dos excedentes líquidos dos consumidores, sendo o fator de ponderação a utilidade social marginal da renda dos indivíduos.⁵³

O problema de maximização é, então:

$$\text{Max } W$$

$$s. a \ E \geq 0$$

O método de Feldstein para lidar com a questão distributiva consiste em explicitar as características distributivas do(s) serviço(s) dentro da fórmula que determina a tarifa. A característica distributiva R_i é definida como a soma ponderada da participação do consumidor de renda Y no consumo total do serviço em questão. Tem-se, então, que:

$$R_i = (N / Q_i) \int_0^{\infty} q_i(Y) u'(Y) f(Y) dY \quad (10)$$

onde:

Q_i = quantidade total ofertada do bem ou serviço i ;

$q_i(Y)$ = quantidade do bem ou serviço i consumido pelo consumidor de renda Y .

⁵² Assume-se a hipótese de que a utilidade social marginal da renda para um indivíduo é independente do preço por unidade do serviço e do encargo fixo. A utilidade social marginal da renda indica a variação no bem-estar social em relação à variação na utilidade de um indivíduo, quando a renda deste último varia.

⁵³ Dada a hipótese especificada na nota anterior, o uso do excedente do consumidor para mensurar o bem-estar dos indivíduos não apresenta maiores problemas, pois não há a necessidade de se diferenciar entre demanda não-compensada e demanda compensada (Feldstein, 1972b).

Quanto mais alto o valor da característica distributiva, mais o consumo do serviço estará concentrado nas famílias de baixa renda, e, portanto, mais baixo deveria ser o preço relativo do serviço. É razoável, portanto, se supor que o valor da característica distributiva será maior para um bem necessário ou essencial do que para um bem de luxo (Wilson, 1977).⁵⁴

Suponha-se, por exemplo, o caso em que há dois serviços, água e esgoto, oferecidos aos consumidores. A função de bem-estar assume a seguinte forma:

$$W = N \int_0^{\infty} S(p_1, p_2, Y) u'(Y) f(Y) dY \quad (11)$$

onde: p_1, p_2 são as tarifas de água e esgoto, respectivamente.

A tarifa de duas partes de Feldstein para um dado serviço, então, seria definida da seguinte forma:

$$T = [\varepsilon / \varepsilon(R_i - 1)] c + E \quad (12)$$

onde:

ε = elasticidade-preço da demanda pelo serviço;

R_i = característica distributiva do serviço;

c = custo marginal.

Conseqüentemente, a percentagem em que a tarifa deve divergir do custo marginal será tanto maior quanto menor for a característica distributiva do serviço e a elasticidade-preço da demanda pelo mesmo.

A tarifa de duas partes de Feldstein compreende, assim, uma regra de preços envolvendo dois fatores básicos: um fator de eficiência, representado pela regra do inverso da elasticidade (regra de Ramsey), e um fator de justiça social, dado pela característica distributiva (Wilson, 1977).

Com o preço divergindo do custo marginal, perdas de eficiência ocorrem, as quais são contrabalançadas pelos ganhos em termos de justiça distributiva. A regra de Ramsey é complementada pela característica distributiva do serviço, de modo a corrigir o efeito regressivo, em termos de distribuição, da formulação original daquela regra.

O resultado final garante a minimização das perdas de eficiência e o atendimento do objetivo de justiça distributiva. Note-se, também, que o *markup* do preço sobre o custo marginal não depende do encargo fixo, ou seja, há uma completa separação entre os mercados de participação e de consumo em relação à fixação da tarifa.⁵⁵

⁵⁴ Como Feldstein (1972a:33) afirmou: "A hipótese de bem-estar convencional de que $u'(Y)$ declina à medida que Y aumenta, implica que o valor de R_i será maior para um bem essencial do que para um bem de luxo. Quanto maior a elasticidade-renda da demanda por um bem, menor o valor de R_i " (tradução nossa). Para uma discussão detalhada da relação entre preços e a característica distributiva, ver Andrade (1993), que também discute como, para uma dada estrutura tarifária, pode-se calcular os julgamentos de bem-estar embutidos.

⁵⁵ Este aspecto do modelo de Feldstein, que implicitamente implica que a elasticidade-preço no mercado de participação é igual a zero, tem sido criticado por desconsiderar os efeitos, em termos de eficiência, associados à determinação do encargo fixo (Navajas & Porto, 1990).

5. Considerações finais

A tônica das discussões de esquemas alternativos de tarificação realizadas neste artigo consistiu em, essencialmente, estabelecer uma restrição orçamentária para a empresa e desenhar uma tarifa que respeitasse aquela restrição e maximizasse o bem-estar dos indivíduos. A importância da imposição de tal restrição é realçada quando se leva em conta um quadro de dificuldades financeiras que obriga as empresas a buscarem um maior nível de eficiência e a geração de recursos próprios para financiar suas necessidades de investimento. Nesse sentido, pode-se dizer que a restrição orçamentária atua como um incentivo para se obter ganhos de eficiência (Heady, 1989).

Seja através de tarifas lineares do tipo de Ramsey, seja por meio de tarifas não-lineares (tarifa de duas partes de Coase, tarifas multipartes etc.), é possível induzir as empresas a buscarem um aumento na eficiência da oferta de seus serviços.

Mas os critérios de eficiência econômica e de equilíbrio financeiro não são os únicos a balizarem a questão tarifária. É legítimo supor que a sociedade também tem um interesse todo especial na questão de justiça social. Esta revisão da literatura de tarifas ótimas mostrou que é possível lidar explicitamente com a questão distributiva dentro das estruturas tarifárias, um exemplo sendo a tarifa de duas partes de Feldstein.⁵⁶

A abordagem do segundo melhor chama a atenção para o inevitável *trade-off* entre os objetivos de ganhos de eficiência e de ganhos distributivos, e, dado o princípio da universalidade de acesso aos serviços, esse *trade-off* tende a ocorrer contra o objetivo de eficiência. Uma política regulatória apropriada procuraria equilibrar criteriosamente os dois objetivos e evitar uma exacerbação daquela tendência, de modo a não resultar em perdas de eficiência maiores do que os ganhos distributivos.

Se o critério de eficiência econômica fosse soberano e o problema da assimetria de informação não fosse tão pronunciado, a tarifa de Ramsey, na presença de uma restrição orçamentária, poderia ser vista como a mais indicada, pois seria capaz de gerar um resultado bem próximo do primeiro melhor.⁵⁷ Tarifas de Ramsey são, entretanto, notoriamente de difícil implementação devido aos requisitos de informação exigidos (informação sobre custos, demanda, gostos etc.), os quais são improváveis de estarem à disposição do agente regulador.

As tarifas de duas ou várias partes também incorrem em problemas de informação, mas dada a possibilidade de se recorrer à discriminação de preços, esses problemas podem ser minimizados. Além do mais, essas estruturas tarifárias são mais adequadas para lidarem com a questão distributiva, o que é uma grande vantagem em se tratando do caso onde os serviços prestados são considerados essenciais e de grande importância social.

Como as empresas de utilidade pública tendem a operar com um componente alto de custos fixos e com problemas periódicos de capacidade, um cuidado todo especial deve ser tomado em relação ao conceito de custo marginal a ser usado na determinação de tarifas. Há bons argumentos para se usar tanto o custo marginal de curto prazo quanto o de longo prazo, dependendo da intensidade do problema de capacidade. Não há necessidade de, e não seria aconselhável, estabelecer-se um único conceito de custo marginal para todos os casos.

Há que se levar em conta também que obstáculos políticos e administrativos podem restringir a utilização de tarifas ótimas, caso em que se poderia apelar para uma política tarifária

⁵⁶ É possível argumentar que a necessidade de se levar em consideração a questão distributiva na determinação de preços também é justificável do ponto de vista estritamente teórico (Brown & Heal, 1979).

⁵⁷ O mesmo argumento se aplica às tarifas do tipo carga de pico.

que objetivasse a busca de melhorias de eficiência a partir do *status quo*. Uma dificuldade aqui é que se o *status quo* representa uma situação bem distante da posição ótima, pode haver uma variedade enorme de opções que podem conduzir a melhorias de eficiência, impedindo a obtenção de regras claras para se definir a escolha da melhor opção (Bos, 1985). No entanto, considerações outras que não de eficiência (por exemplo, objetivos de justiça social) podem ajudar nessa escolha.

Para a elaboração de uma adequada política tarifária, é imprescindível que se façam estudos sobre a estrutura de custos (especialmente dos custos marginais), de demanda (elasticidades preço e renda), de padrões de preferências etc. Todos os esquemas alternativos de tarifas discutidos neste artigo dependem fundamentalmente dessa massa de dados para poderem ser testados. Além do mais, é a especificidade de cada serviço de utilidade pública, tornada explícita através do estudo empírico e da coleta de dados, que deverá sugerir em última instância sua estrutura tarifária mais adequada.

O estudo criterioso do que constitui uma estrutura tarifária ótima é, na ótica dos autores, condição *sine qua non* para se analisar o *status quo* e para propor-lhe alternativas. Como bem alertou Wiseman (1957), um importante papel do economista consiste em examinar criticamente as políticas correntes adotadas pelos governos com o intuito de tornar claros a natureza e as consequências dessas políticas.

Referências bibliográficas

- Andrade, T. Objetivos distributivos e preços discriminatórios para empresas prestadoras de serviços públicos. *Pesquisa e Planejamento Econômico*, 23, 1993.
- Baumol, W. & Bradford, D. Optimal departures from marginal cost pricing. *American Economic Review*, 60, 1970.
- Blaug, M. Marginal cost pricing: no empty box. In: Blaug, M. *Economic theories, true or false?* London, Edward Elgar, 1990.
- Boiteux, M. On the management of public monopolies subject to budgetary constraints. *Journal of Economic Theory*, 3, 1971.
- Bos, D. Public sector pricing. In: Auerbach, A. & Feldstein, M. (eds.). *Handbook of public economics*. Amsterdam, North-Holland, 1985. v.1.
- Braeutigan, R. Optimal policies for natural monopolies. In: Schmalensee, R. & Willig, R. *Handbook of industrial organization*. Amsterdam, North-Holland, 1989.
- Brown, D. & Heal, G. Two-part tariffs, marginal cost pricing and increasing returns in a general equilibrium model. *Journal of Public Economics*, 13, 1979.
- Brown, S. & Sibley, D. *The theory of public utility pricing*. Cambridge, Cambridge University Press, 1986.
- Cavalcanti, J. C. *Economic aspects of the provision and development of water supply in 19th century Britain*. University of Manchester, Great Britain, 1991. (Tese de Doutorado.)
- Coase, R. The marginal cost controversy. *Economica*, 13, 1946.
- Crew, M. & Kleindorf, P. *Public utility economics*. London, MacMillan, 1979.
- Diamond, P. & Mirrlees, J. Optimal taxation and public production (I): production efficiency. *American Economic Review*, 61, 1971.

- Faulhaber, G. & Panzar, J. *Optimal two part tariffs with self-selection*. 1978. (Bell Laboratories Economic Discussion Paper, 74.)
- Feldstein, M. Distributional equity and the optimal structure of public prices. *American Economic Review*, 62, 1972a.
- . Equity and efficiency in public sector pricing: the optimal two-part tariff. *Quarterly Journal of Economics*, 86, 1972b.
- Gabarrone, C. & Reis, L. *Diagnóstico sobre a questão tarifária em saneamento*. Brasília, Ipea, 1994.
- Heady, C. Public sector pricing in a fiscal context. World Bank, 1989. (PPR Working Paper Series, WPS179.)
- Henderson, A. The pricing of public utility undertakings. *The Manchester School*, 1947.
- Hotelling, H. The general welfare in relation to problems of taxation and of railway and utility rates. *Econometrica*, 6, 1938.
- Laffont, J-J. & Tirole, J. *A theory of incentives in procurement and regulation*. Boston, MIT Press, 1993.
- Munasingh, M. Principles of modern electricity pricing. World Bank, 1981. (World Bank Staff Occasional Papers, 5.)
- Navajas, F. & Porto, A. La tarifa en dos partes cuasi óptima: eficiencia, equidad y financiamiento. *El Trimestre Económico*, 1990.
- Ruggles, N. Recent developments in the theory of marginal cost pricing. *Review of Economic Studies*, 1949/50.
- Sheshinski, E. Positive second-best theory: a brief survey of the theory of Ramsey pricing. In: Arrow, K. & Intriligator, M. (eds.). *Handbook of mathematical economics*. Amsterdam, North-Holland, 1986. v. 3.
- Stiglitz, J. *Economics of the public sector*. 2 ed. New York, W. W. Norton, 1988.
- . Another century of economic science. *Economic Journal*, 101, 1991.
- Varian, H. Price discrimination. In: Schmalensee, R. & Willig, R. (eds.). *Handbook of industrial organization*. Amsterdam, North-Holland, 1989. v. 1.
- Walters, A. *The economics of road user charges*. World Bank, 1986. (World Bank Staff Occasional Papers, 5.)
- Williamson, O. Peak load pricing and optimal capacity under indivisibility constraints. *American Economic Review*, 56, 1966.
- Willig, R. D. Pareto-superior nonlinear outlay schedule. *Bell Journal of Economics*, 9:56-69, 1978.
- Wilson, L. The interaction of equity and efficiency factors in optimal pricing rules. *Journal of Public Economics*, 7, 1977.
- Wiseman, J. The theory of public utility price: an empty box. In: Rowley, C. (ed.). *Readings in industrial economics*. London, MacMillan, 1957. v. 2.
- Zajac, E. Basic Ramsey prices in the regulated sector. In: Barke, S. & Elliott, C. (eds.). *Readings in public sector economics*. Lexington, D.C., Heath and Company, 1990.