

ISSN 0871-2573

**Economias de Escala e Eficiência Produtiva na
Banca Portuguesa
Uma Revisão da Literatura**

**Paulo Soares de Pinho
Faculdade de Economia da Universidade Nova de
Lisboa
Working Paper n° 241
Janeiro, 1995**

**ECONOMIAS DE ESCALA E EFICIÊNCIA
PRODUTIVA NA BANCA PORTUGUESA:**

Uma Revisão da Literatura

Paulo Soares de Pinho
Faculdade de Economia
Universidade Nova de Lisboa

1. Introdução

A progressiva liberalização do sector bancário português iniciada em meados dos anos oitenta tem estado associada a estratégias de crescimento por parte da maioria das instituições de crédito do nosso país. De um modo geral, a triade de instrumentos competitivos, preço, publicidade e balcões tem sido utilizada com crescente intensidade com óbvios reflexos sobre a rendibilidade do conjunto do sector.

A maioria das instituições justifica o seu esforço de crescimento com a necessidade de adquirir "dimensão competitiva", i.e., uma dimensão que assegure a sua sobrevivência no longo prazo, dentro do contexto do mercado interno europeu. Tal justificação parece hoje um pouco esbatida, face ao reduzido esforço de entrada por parte de instituições estrangeiras e, até, às saídas que já se verificaram. Tal (ausência de) dinâmica de entrada vem, pelo contrário, sugerir um elevado grau de protecção do mercado doméstico, sendo portanto necessário encontrar outras explicações para esta procura de dimensão.

As motivações para crescer, quer através de uma dinâmica interna, quer através de fusões, centram-se, com frequência, no argumento das economias de escala. Se a tecnologia bancária tiver características tais que a expansão da escala de produção seja sustentada por um crescimento inferior dos recursos utilizados, então os bancos de maior dimensão diporão de uma clara vantagem competitiva: custos mais baixos.

Esta relação entre custos e dimensão, a verificar-se, aponta para uma situação de monopólio natural dentro da qual, no longo prazo, apenas sobreviverão os bancos de maior dimensão. Assim, o estágio actual do nosso sistema bancário deveria ser visto como puramente transitório, ao qual se seguiria um "shake-up" do qual só sobreviveriam as instituições mais capazes. Neste contexto, as fusões surgiriam naturalmente e dariam origem a bancos grandes e eficientes, dos quais resultaria um sistema de intermediação financeira mais barato que o actual.

Igualmente importante neste contexto é o apuramento dos eventuais ganhos de diversificação associados à partilha de recursos comuns na produção de diferentes produtos. Caso estas economias de *scope*¹ sejam significativas, então bancos com linhas de produtos alargadas terão vantagens de custos sobre os concorrentes mais

¹Não existe tradução universalmente consagrada em português, pelo que aqui se adopta a designação anglo-saxónica. Entre as várias traduções possíveis, registe-se "economias de fins", de "gama", de "âmbito", e de "diversificação".

especializados. Assim, a sua avaliação ocupa um papel central na literatura sobre a eficiência das instituições de crédito.

Este foco no lado dos custos peca por negligenciar os potenciais efeitos entre a dimensão e os proveitos das instituições de crédito. Com efeito, não sendo perfeitamente competitivos os mercados de depósitos e crédito, existe sempre a possibilidade de existir alguma relação entre poder de mercado e dimensão. Assim, e de um modo geral, é muito possível que empresas dominantes ou com grande quota de mercado disponham de vantagens significativas sobre os seus clientes, traduzidas em termos das margens praticadas sobre os clientes.

Se tal relação se verificar, os bancos serão levados a crescer como forma de obter poder de mercado, independentemente da existência de economias de escala. Neste contexto, as fusões resultarão do interesse em ganhar margem e reduzir o número de jogadores no mercado, condição para aumentar o grau de cooperação entre eles. No caso de tal dinâmica conduzir a uma elevada concentração do mercado, os bancos disporão de uma situação confortável, adquirida à custa dos seus clientes. No conjunto, o sistema de intermediação financeira ficaria mais caro e menos eficiente.

Desta discussão deve resultar claro que não é indiferente conhecer as causas subjacentes à actual dinâmica de algumas instituições de crédito. No caso de a situação actual poder ser descrita como de notórias economias de escala dentro de uma situação francamente competitiva, então os consumidores de serviços bancários acabarão por ser beneficiados com o processo actual. Se, pelo contrário, existirem rendimentos constantes ou deseconomias de escala, ao mesmo tempo que existe uma relação positiva entre dimensão e poder de mercado, os processos de expansão podem conduzir a importantes perdas de bem-estar na economia, devendo as autoridades de supervisão acompanhar cuidadosamente a situação e impedir as fusões entre empresas de grande dimensão.

Para além da determinação da escala óptima de produção nesta indústria, particular atenção tem sido dedicada ao problema da ineficiência produtiva na banca. Isto, porque uma determinada instituição pode exibir custos operacionais elevados por razões, em nada relacionadas com uma escolha menos eficiente de escala operativa (eficiência de escala), associadas a puro desperdício na afectação de recursos (eficiência produtiva). Assim, torna-se importante compreender até que ponto um dos efeitos domina sobre o outro, ou seja, se os bancos exibem custos elevados em virtude de não aproveitarem eventuais benefícios de escala ou por efectuarem uma má gestão de recursos produtivos.

A ineficiência produtiva pode, em Portugal, explicar uma componente significativa dos custos dos bancos, dada a situação de excesso de mão de obra resultante da descolonização. Por outro lado, as nacionalizações quase anularam a competição no sector, desencorajando os investimentos conducentes a ganhar produtividade, assim como poderão ter conduzido à existência de eventuais problemas de agência associados à separação entre propriedade e gestão.

A existência de dois estudos recentes sobre esta temática genérica, fundados em diferentes metodologias, sugere a necessidade de se proceder a uma análise comparativa dos mesmos, conducente à obtenção de conclusões mais sólidas que as resultantes de uma análise individual.

O presente artigo encontra-se organizado da seguinte forma: a secção dois aborda os conceitos teóricos associados à medição de economias de escala e eficiência produtiva, na terceira secção procede-se a uma análise comparativa dos resultados obtidos em diferentes estudos, sendo as conclusões gerais apresentadas na secção quatro.

2. Metodologia

2.1. O Modelo Base

Desde o início dos anos oitenta a literatura sobre a determinação dos rendimentos de escala verificados na indústria bancária tem vindo a basear-se numa abordagem *multi-output*, ou seja, reconhecendo o carácter multi-produto das empresas operando neste sector. No essencial, assume-se que as instituições bancárias utilizam um conjunto de factores de produção (vector X) para produzir um conjunto de produtos (vector Y), de acordo com uma função de transformação genérica:

$$F(Y,X) = 0 \tag{1}$$

Pela aplicação da dualidade entre funções de produção e de custos, em geral, e do lema de Shepherd (1953), em particular, as propriedades da tecnologia acima podem ser completamente caracterizadas por uma função de custos do tipo:

$$C = C(Y,W) \tag{2}$$

Onde W representa o vector de preços de *inputs* associado a X e C representa os custos da empresa bancária.

A necessidade de trabalhar com painéis de dados onde se misturam bancos de retalho com bancos grossistas², obriga a que se tenham algumas precauções quanto à eventual aplicabilidade da mesma tecnologia aos dois tipos de instituições. Como forma de corrigir este problema sem obrigar a partir a amostra em sub-amostras, é comum pressupor que a tecnologia pode sofrer variações em torno do número de balcões do banco³. Assim, é comum permitir a interacção entre vectores de *outputs* e preços com aquela variável (designada por B), numa função genérica do tipo:

$$C = C(Y, W, B) \quad (3)$$

Uma vez definidos os *inputs* e *outputs* da empresa bancária, o passo seguinte consiste na estimação de uma forma flexível para (2). A mais popular entre os investigadores é a translog flexível, que resulta de uma aproximação em série de Taylor de segunda ordem da função (2), em torno do vector unitário. A forma funcional correspondente é dada por⁴:

$$\begin{aligned} \ln(C) = & \alpha_0 + \sum_{i \in Y} \beta_i \ln(y_i) + \sum_{k \in W} \alpha_k \ln(w_k) + \frac{1}{2} \sum_{i \in Y} \sum_{j \in Y} \delta_{ij} \ln(y_i) \ln(y_j) + \\ & \frac{1}{2} \sum_{k \in W} \sum_{l \in W} \gamma_{kl} \ln(w_k) \ln(w_l) + \sum_{i \in Y} \sum_{k \in W} \rho_{ik} \ln(y_i) \ln(w_k) + \\ & \psi_b \ln(B) + \psi_{bb} \ln^2(B) + \sum_{i \in Y} \psi_i \ln(B) \ln(y_i) + \sum_{k \in W} \psi_k \ln(B) \ln(w_k) \end{aligned} \quad (4)$$

A expressão acima, tem como importante inconveniente o facto de apresentar custos nulos sempre que um dos *outputs* não seja produzido, o que se torna particularmente desagradável quando se lida com empresas pouco diversificadas. Em particular, esta situação conduz a problemas na avaliação das economias de *scope*. A fim de aliviar este problema, alguns autores optam por substituir os diferentes $\ln(y)$ da translog pela

²Ou, no caso dos EUA, bancos de estados "single-branch" com bancos de estados "multiple-branch".

³Sobre este assunto veja-se Benston *et al* (1982).

⁴Esta forma funcional foi consagrada por Berger *et al* (1987).

transformação Box-Cox (1964) sobre os *outputs* [f(y)], dando origem à denominada *translog híbrida*. A transformação é dada por:

$$f(y_i) = (y_i^{\phi_i} - 1) / \phi_i \quad \text{para } \phi_i \neq 0 \quad \wedge \quad f(y_i) = \ln(y_i) \quad \text{para } \phi_i = 0 \quad (5)$$

Esta abordagem implica a estimação de um parâmetro de transformação (ϕ) para cada um dos *outputs* considerados. Num recente survey desta literatura Berger, Hunter e Timme (1993, pg 225) mencionam o facto de num elevado numero de estudos serem estimados valores próximos de zero para aquele parâmetro, o que conduz à especificação translog. Num estudo igualmente recente, Pulley e Humphrey (1993) apresentam criticamente várias alternativas para ultrapassar aqueles inconvenientes.

Com vista a poder ser considerada como uma função custos, a translog deve ainda verificar as seguintes propriedades:

Simetria:

$$\delta_{i,j} = \delta_{j,i}, \quad \forall_{i \neq j} \quad (6)$$

$$\gamma_{k,l} = \gamma_{l,k}, \quad \forall_{k \neq l} \quad (7)$$

Homogeneidade nos Preços de Factores:

$$\sum_{k \in W} \alpha_k = 1 \quad (8)$$

$$\sum_{i \in W} \gamma_{ki} = 0, \quad \forall_{k \in W} \quad (9)$$

$$\sum_{k \in W} \rho_{ik} = 0, \quad \forall_{i \in Y} \quad (10)$$

$$\sum_{k \in W} \psi_k = 0 \quad (11)$$

Embora muitos autores prefiram testar *a posteriori* a sua verificação, por vezes, as restrições teóricas (6) a (11) são imediatamente incorporadas no modelo com vista a reduzir o numero de parâmetros a estimar. Habitualmente, a expressão (4) costuma ser estimada juntamente com as equações dos "shares", com vista a aumentar a eficiência dos estimadores. Normalmente, utiliza-se o método de Zellner para estimar o sistema de equações simultâneas aparentemente não correlacionadas constituído por

(4) e por n-1 equações de shares (sendo n igual a numero de *inputs*)⁵. Estas últimas são dadas por:

$$S_k = \alpha_k + \sum_{l \in W} \gamma_{kl} \ln(w_l) + \sum_{i \in Y} \rho_{ik} \ln(y_i) + \psi_k \ln(B), \quad (k \in W) \quad (12)$$

O processo de estimação acima descrito tem, no entanto, uma importante desvantagem: não permite uma separação directa entre eficiência de escala e eficiência produtiva. Com efeito, alguns autores vêm defendendo ser a primeira menos importante que a segunda, tornando crucial a possibilidade de obter estimativas directas do grau de ineficiência produtiva associado a cada observação⁶. Por outro, ao assumir simetria no resíduo de (4) está-se a permitir que problemas de ineficiência sejam confundidos com eventuais deseconomis de escala, visto que a dualidade entre produção e custos só é aplicável sobre a *fronteira eficiente de custos*, i.e., sobre a componente eficiente dos custos observados⁷.

Como forma de resolver os problemas acima, alguns autores optam por estimar fronteiras eficientes em alternativa à estimação econométrica tradicional. A metodologia foi introduzida por Aigner *et al* (1977) e consiste em assumir que os custos observados têm três componentes: uma corresponde ao valor *eficiente* (ou seja, sem desperdício de recursos) dos custos necessários para atingir o nível de produção da empresa, outro resulta da ineficiência-X, ou seja, consiste no montante de custos associado ao desperdício de recursos e, um terceiro, resulta de efeitos aleatórios não controlados pela empresa. O método consiste então na estimação de:

$$C_{i,t} = C(Y_{i,t}, w_{i,t}, B_{i,t}) + u_{i,t} + v_{i,t} \quad (13)$$

onde u é um resíduo estritamente positivo ($u \geq 0$) e associado à ineficiência-X, enquanto v é um resíduo com as propriedades econométricas habituais, associado à componente estocástica dos custos observados. A estimativa da primeira parcela é, normalmente, designada por *fronteira estocástica*. A palavra *fronteira* deriva do facto de consistir num envelope sobre o qual se encontram as empresas com eficiência total na utilização de recursos ($u=0$). Diz-se *estocástica* por permitir que haja desvios entre

⁵A discussão desta matéria encontra-se bastante clara em Mendes (1991, pp 457-8), pelo que é aqui omitida.

⁶Veja-se, a este propósito, Berger e Humphrey (1991) e Berger, Hunter e Timme (1993).

⁷Sobre este assunto, veja-se Berger, Hunter e Timme (1993, pp 223-7).

a fronteira e os custos observados resultantes de efeitos aleatórios, os quais são particularmente importantes em algumas indústrias⁸.

Escolhida uma forma funcional para (13), por exemplo (4), e assumindo uma determinada distribuição de probabilidades para u , o modelo é estimado pelo método de máxima verosimilhança⁹. No caso mais frequente, assume-se que u segue uma distribuição semi-normal e a correspondente função de log-verosimilhança é:

$$\ln(\mathcal{L}) = \frac{n}{2} \ln\left(\frac{2}{\pi}\right) - n \ln(\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^n \varepsilon_i^2 + \sum_{i=1}^n \ln\left(\Phi\left(\frac{\varepsilon_i \lambda}{\sigma}\right)\right) \quad (14)$$

onde:

$$\lambda = \frac{\sigma_v}{\sigma_u} \quad \sigma = \sqrt{\sigma_u^2 + \sigma_v^2} \quad \text{e} \quad \varepsilon_{i,t} = u_{i,t} + v_{i,t}$$

$\Phi(x)$ representa a função de distribuição normal

O modelo é estimado pelo método de máxima verosimilhança. Normalmente, usa-se como ponto de partida a estimativa OLS dos parâmetros da função custo e os valores de partida para λ e σ propostos em Waldman (1978) e reportados em Greene (1982). Os resíduos estimados (ε), podem ser decompostos nas suas componentes elementares, segundo o método proposto por Jondrow *et al* (1982)¹⁰:

$$E(u_{i,t} | \varepsilon_{i,t}) = \frac{\sigma_u \sigma_v}{\sigma} \left(\frac{\phi(\lambda \varepsilon_{i,t} / \sigma)}{\Phi(\lambda \varepsilon_{i,t} / \sigma)} + \frac{\lambda \varepsilon_{i,t}}{\sigma} \right) \quad (15)$$

onde:

$\phi(x)$ representa a função densidade de probabilidade normal

⁸Nos primeiros estudos sobre fronteiras, estas assumiam carácter determinístico; o modelo era estimado pelos métodos habituais e, posteriormente, subia-se o termo constante até que não existissem resíduos negativos. Os valores assim obtidos para estes últimos eram então interpretados como indicadores de ineficiência.

⁹Para um *survey* sobre estimação de fronteiras estocásticas e determinísticas veja-se Greene (1993).

¹⁰A fórmula aqui apresentada é, mais uma vez, relativa ao modelo semi-normal. Deve-se, ainda, notar que as estimativas assim obtidas não são consistentes.

Esta abordagem pelas fronteiras estocásticas tem, portanto, como principais vantagens o permitir uma fácil estimação conjunta das eficiências técnica e de escala. Nesta linha surgem os estudos de Bauer *et al* (1993), Cebenoyan *et al* (1993) e Mester (1993). Contudo, a necessidade de impor uma forma funcional (paramétrica) para a fronteira de custos constitui importante desvantagem, dado que os resultados obtidos dependem com frequência das características específicas da função adoptada. Assim, parte da investigação nesta área passou a ser devotada a métodos não paramétricos de avaliação da eficiência, os quais se baseiam na comparação directa entre as observações.

Os métodos não paramétricos partem das instituições com custos comparativamente mais baixos como termo de comparação para as restantes. Consequentemente, estes métodos são normalmente criticados por se basearem excessivamente em observações anormais ("outliers"). Dado que os estudos publicados sobre Portugal se baseiam em abordagens paramétricas optamos por não desenvolver este tópico. Contudo, o leitor interessado poderá obter mais informação sobre eles em Ali e Seiford (1993). Numa perspectiva da empresa bancária, as duas metodologias são comparadas em Evanoff e Israilevich (1991), enquanto Berger, Hunter e Timme (1993) apresentam uma interessante discussão sobre as vantagens relativas dos dois métodos.

2.2. Definição de Inputs e Outputs

Uma das questões controversas nesta literatura é a classificação dos *inputs* e *outputs* da empresa bancária. Tradicionalmente, a literatura divide-se em duas abordagens distintas: produção e intermediação¹¹.

A abordagem da produção vê a empresa bancária como produtora de serviços, pelo que o *output* consiste no volume dos mesmos, os quais são produzidos a partir dos *inputs* primários: capital e trabalho. Em termos de implementação empírica é usual usar como *proxies* dos *outputs* o *numero* de cada tipo de contas servidas pela instituição. O conceito de custos utilizado corresponde à soma dos custos do trabalho e do capital.

A abordagem da intermediação, desenvolvida por Sealey e Lindley (1977), vê a empresa bancária como simples intermediária, produzindo activos remunerados a

¹¹Uma excelente revisão desta literatura é proporcionada por Colwell e Davies (1992).

partir de factores de produção reais e monetários. Assim, em termos práticos, o problema é formulado com a empresa produzindo empréstimos e outras aplicações financeiras a partir dos factores de produção depósitos, outros fundos captados, capital e trabalho. O conceito de custos utilizado corresponde à soma dos custos operacionais (trabalho e capital) com os custos financeiros (juros).

Como é evidente a adopção da abordagem de produção pressupõe a disponibilidade do numero de contas (de dada tipo) de cada instituição, informação não disponível na maioria dos países. Assim, visto a abordagem de intermediação apenas exigir informação constante nas contas publicadas pelos bancos, esta tende a prevalecer nos estudos publicados. Os seus defensores argumentam ainda que dado o significativo peso que os juros assumem na estrutura de custos da empresa, a sua exclusão poderá conduzir a resultados enviados, sobretudo se existir algum *tradeoff* entre custos financeiros e custos operacionais.

Este último argumento, embora bastante comum, pode ser facilmente criticado por confundir dois conceitos fundamentais: custos operacionais e poder de mercado nos depósitos. Com efeito, o mencionado *tradeoff* está de algum modo relacionado com o facto de, muito frequentemente, instituições com pequena dimensão (e baixos custos operacionais) disporem de reduzido poder de mercado (elevados custos financeiros). Assim, os dois problemas devem ser objecto de tratamento individualizado, ainda que de forma integrada, o que não é possível quando se olha para a empresa exclusivamente sob a óptica dos custos e se ignoram os efeitos de interacção oligopolística. Um interessante estudo que integra estes dois aspectos foi recentemente publicado por Berg e Kim (1994). Outra alternativa de abordagem, recentemente proposta por Berger, Hancock e Humphrey (1993), consiste na estimação directa de funções lucro, a partir da qual são detectáveis os efeitos de escala sobre receitas e custos.

Apesar da indisponibilidade de dados relativos ao numero de contas apontar para uma adopção quase exclusiva da abordagem da intermediação, vários autores vêm divergindo da metodologia acima apresentada e optam por apresentar classificações alternativas de *inputs* e *outputs*. Assim, o valor monetário dos depósitos (sobretudo à ordem) é, por vezes, classificado como um produto em virtude da especial importância que estes assumem no consumo de recursos reais da empresa. Outros

autores justificam esta classificação com base no significativo contributo que os depósitos dão para a geração do valor acrescentado da empresa bancária¹².

Neste contexto surge, então, uma visão "revisionista" da abordagem de intermediação, na qual o processo produtivo da empresa bancária se processa em duas fases: Na primeira, originação, os bancos usam capital e trabalho para originar depósitos e crédito. Na segunda, decisão de tesouraria, os depósitos e capitais próprios são utilizados para financiar empréstimos, investimentos financeiros e o capital físico, sendo a tesouraria utilizada para ajustar a igualdade de balanço¹³.

Embora persista uma importante controvérsia sobre este tema, a análise comparativa dos estudos utilizando os mesmos dados e diferentes classificações de produtos e factores parece indicar que os resultados relativos a economias de escala e de *scope* são razoavelmente insensíveis à alternativa adoptada¹⁴. Contudo, no que respeita à medição da eficiência e produtividade tem sido detectada importante sensibilidade dos resultados à escolha de *inputs* e *outputs* adoptada.

2.3. Medição de Economias de Escala, de *Scope* e de Ineficiência Produtiva

A medição de economias de escala e de *scope* na indústria bancária vem no essencial seguindo as medidas multi-produto consagradas em Berger, Hanweck e Humphrey (1987). Recorrendo, mais uma vez, à dualidade entre produção e custos, os rendimentos de escala podem ser medidos através de:

$$RES = \sum_{i \in Y} \frac{\partial \ln(C)}{\partial \ln(y_i)} \quad (16)$$

Assim, se $RES=1$ os custos de produção aumentam na mesma proporção que a produção do banco, pelo que teremos rendimentos constantes à escala. Para $RES>1$, os primeiros aumentam mais que proporcionalmente que os segundos, indicando deseconomias de escala. Com $RES<1$ teremos economias de escala.

Um factor importante para a interpretação dos resultados acima é a especificação adoptada para a função custos. No caso de se ter adoptado a forma (2), em que o

¹²Veja-se, a este respeito, Berger e Humphrey (1991).

¹³Esta abordagem foi, no essencial, introduzida por Kolari e Zardkoohi (1987), retomada por Berger e Humphrey (1991) e formalizada algébricamente em Pinho (1994).

¹⁴Veja-se Clark (1988, pg 24).

numero de balcões não surge como variável autónoma, então os custos associados à variação dos balcões associada a mudanças de escala produtiva já estão incorporados naquela medida. Nesse caso, diz-se que a medida acima cai na família das *firm economies of scale*.

No caso de se ter optado pela especificação (3), então a medida acima é calculada no pressuposto de que o numero de balcões permanece constante. Os resultados acima obtidos são, portanto, apenas válidos para situações em que variações do nível de produção não impliquem alterações na dimensão da rede de balcões. Nesse caso, diz-se que a medida acima cai na família das *plant economies of scale*.

Contudo, os bancos, por razões de natureza competitiva, são normalmente forçados a aumentar a sua rede de balcões sempre que pretendem expandir o seu nível de actividade, pelo que as medidas do tipo *plant* são, na maioria das situações, praticamente inúteis. Por outro lado, a especificação da função custos (3) revela-se, na generalidade dos casos, superior a (2) em termos de implementação econométrica, pelo que se torna importante compatibilizar estas duas vertentes.

A solução de problema foi proposta, para a situação uni-produto, por Benston, Hanweck e Humphrey (1982), através do conceito de *augmented economies of scale*, cuja versão multi-produto é:

$$AES = \sum_{i \in Y} \frac{\partial \ln(C)}{\partial \ln(y_i)} + \sum_{i \in Y} \frac{\partial \ln(C)}{\partial \ln(B)} \frac{\partial \ln(B)}{\partial \ln(y_i)} \quad (17)$$

Nesta medida, o efeito da variação do numero de balcões é dado pela segunda parcela. O seu primeiro termo é directamente obtido a partir da função custos, enquanto o segundo deve resultar de estimação econométrica separada.

As medidas (16) e (17) acima caem na família das *ray economies of scale*, visto implicitamente assumirem que o mix de produtos não se altera com a expansão da actividade. No caso de aquele estar directamente ligado à dimensão, torna-se necessário recorrer a medidas que entrem em consideração com as alterações do *output-mix* decorrentes de um processo de expansão.

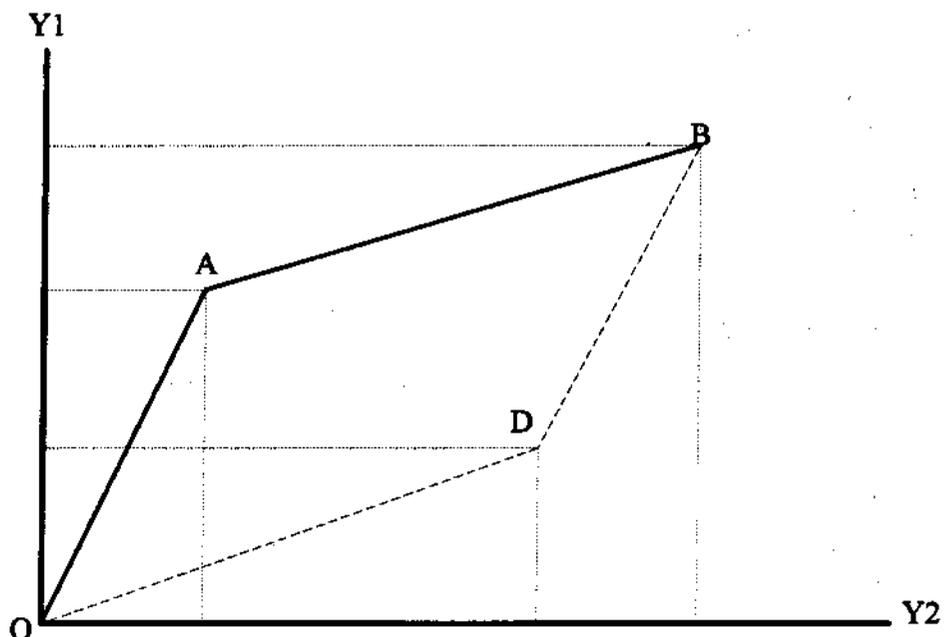
Recorrendo à figura 1, podemos observar um banco de pequena dimensão (A) e outro com maior volume de actividade (B). O cálculo da medida (16) para a instituição B assume que o banco expande a sua actividade ao longo do raio OB, quando na realidade essa expansão é feita ao longo da linha OAB. Assim, e podendo o tipo de rendimentos de escala depender do rácio de produtos, é preferível calcular

separadamente medidas para as expansões OA e AB. A medida proposta por Berger *et al* (1987), denominada *expansion path scale economies*, consiste na elasticidade do custo incremental relativamente à produção incremental, ao longo da linha de expansão de A para B e é calculada como:

$$EPSCE = \sum_{y \in Y} \left(\frac{(y^B - y^A) / y^B}{(C(y^B) - C(y^A)) / C(y^B)} \right) \times \frac{\partial \ln(C(y^B))}{\partial \ln(y^B)} \quad (18)$$

onde y^A e y^B representam, respectivamente, o volume de um determinado produto produzido pelos bancos A e B.

Figura 1
Output Mix e Expansão de Actividade



Objecto de particular atenção na literatura tem sido a medida de eventuais ganhos associados à produção simultânea de vários produtos. O conceito mais utilizado é o das já mencionadas economias de *scope*. Para o caso mais simples de dois produtos, existem economias deste tipo sempre que o custo de produzir cada um deles em separado seja superior ao custo de os produzir conjuntamente, ou seja:

$$C(y_1, 0) + C(0, y_2) > C(y_1, y_2) \quad (19)$$

A medida de economias de *scope* para empresas multi-produto foi proposta por Panzar e Willig (1981) e consiste em:

$$\text{SCOPE} = \sum_{y \in Y} (C(Y - y) - C(Y)) / C(Y) \quad (20)$$

Ou seja, para cada produto, determina-se o custo de produzir toda a gama de produtos da empresa com a exclusão do produto em causa e calcula-se a diferença entre aquele e o custo de produção de toda a gama. A soma destas diferenças é então dividida por aquele último.

No caso de aquela medida ser positiva, isso significa que excluir produtos da gama oferecida pela empresa aumenta os custos médios de produção. O valor assim obtido representa os ganhos de diversificação da gama em percentagem dos custos de produção. Um valor negativo significa que existem perdas associadas à produção conjunta daqueles produtos.

Dois problemas tornam particularmente pouco utilizável a expressão acima. Por um lado, a função mais utilizada neste tipo de estudos, a translog, não é definida para valores nulos de qualquer dos produtos. As tentativas de ultrapassar este problema passam, normalmente, pela definição de valores arbitrários para actuar como *proxy* de zero, a cujo valor aquela medida parece ser particularmente sensível. Por outro lado, na indústria bancária o mix de produtos varia consideravelmente com a dimensão das empresas, fazendo sentido estudar a possível subaditividade de custos associada a situações de expansão do produto (mix e quantidade).

Para responder a estas questões, Berger *et al* (1987) propuseram uma medida baseada na seguinte noção. Vamos supor que existe uma empresa pequena (A) e uma empresa grande (B), sendo a primeira mais especializada num tipo de produtos que a segunda. Podemos definir (veja-se a figura 1) uma terceira empresa (D), cujo vector de produtos é definido como $D=B-A$, com $B>A$ (para todos os produtos). As empresas A e D são, portanto, complementares, visto terem mix de *outputs* opostos. A empresa diversificada B, será mais eficiente que as empresas mais especializadas A e D no caso de:

$$C(Y^A) + C(Y^D) > C(Y^B) \quad (21)$$

A medida proposta por aqueles autores, designada por *expansion path subadditivity*, é então:

$$EPSUB = \frac{C(Y^A) + C(Y^D) - C(Y^B)}{C(Y^B)} \quad (22)$$

Que nos dá os ganhos (ou perdas, caso seja negativa) associados à diversificação, em percentagem dos custos da empresa grande e diversificada B.

A medição da eficiência produtiva encontra-se directamente relacionada com o resíduo assimétrico da expressão (13). A medida mais utilizada é a proposta por Farrell (1957) e consiste no rácio entre o valor eficiente de custos e a soma entre estes e a componente dos mesmos resultante da ineficiência-X. A partir de uma fronteira estocástica translog, este indicador é obtido como:

$$PE = \frac{\hat{C}}{\hat{C} + U} = \exp(-\hat{u}) \quad (23)$$

onde U representa os custos resultantes de ineficiência-X.

O indicador acima tem um valor máximo de um, para uma empresa onde não exista desperdício de recursos. Alguns autores tentam ir um pouco mais longe, dividindo a ineficiência-X em duas componentes: técnica (desperdício puro) e de afectação (uso dos factores na proporção incorrecta. Uma referência sobre esta matéria é Kopp e Diwert (1982).

3. Estudos Empíricos

3.1. Análise Comparativa

Existem dois estudos recentes sobre a eficiência das instituições de crédito portuguesas. O primeiro, Mendes (1990, 1991) cobre vários períodos distintos (65/73, 78/84 e 85/88), enquanto o segundo, Pinho (1994) cobre o período 1988/1992.

O estudo de Mendes (1990, 1991) segue a abordagem de intermediação, definindo os montantes monetários de crédito concedido, aplicações em títulos e contas extrapatrimoniais como os *outputs* da indústria. Os *inputs* considerados são depósitos totais, número de trabalhadores e a soma do imobilizado com os fornecimentos e

serviços externos¹⁵. As contas relativas a investimentos e financiamentos do mercado monetário não são incluídas em qualquer das classificações.

O conceito de custos utilizado neste estudo consiste na soma dos custos operacionais com a totalidade dos custos de operações passivas¹⁶. O preço dos depósitos é aproximado através da divisão entre os custos de operações passivas e os depósitos totais. Assim, ao excluir do denominador os outros recursos remunerados, aquele autor encontra preços relativamente elevados (Mendes, 1990, pg 143). O custo do capital é aproximado como a divisão entre a soma dos custos operacionais (excepto com o pessoal) e o activo total. O preço do trabalho é aproximado pelo rácio entre custos com pessoal e numero de trabalhadores.

Para o período 1985/88 o autor dispõe de 78 observações relativas a todos os bancos a operar em Portugal. O modelo, função custos e *shares*, é estimado, nas duas versões (2) e (3) pelo método de Zellner. O autor ensaia igualmente a substituição dos logaritmos de *outputs* pela transformação Box-Cox (1964).

O estudo de Pinho (1994) parte da já mencionada separação da tecnologia bancária em duas fases para formalizar o problema de optimização da empresa. Obtém como resultado que quer o crédito quer os depósitos devem, naquele contexto, ser tratados como *output*. A actividade extrapatrimonial, por dificuldade de compatibilização de planos de contas, não é incluída¹⁷. Os *inputs* considerados são o trabalho, o capital e os fornecimentos e serviços externos.

O conceito de custos empregue é o da soma dos custos do trabalho, do capital e os fornecimentos e serviços externos. Os primeiros são aproximados de forma semelhante à do estudo anterior. O custo do capital é aproximado pela soma entre a taxa de depreciação do imobilizado e o custo de oportunidade do respectivo financiamento (aproximado pela taxa média do MMI)¹⁸. Como este último elemento não figura na demonstração de resultados contabilística, a definição de custos totais empregue corresponde a uma noção económica, que excede a contabilística. A impossibilidade de obter um preço para os fornecimentos e serviços externos levou à

¹⁵O autor (Mendes, 1991, footnote 17) reconhece tratar-se da soma entre uma variável stock e um fluxo.

¹⁶Ou seja, os custos com os recursos captados no MMI são incorporados, apesar da exclusão destes na lista de inputs.

¹⁷O novo plano de contas para o sistema bancário entrou em vigor em 1990, portanto a meio do período considerado.

¹⁸A falta de melhor informação optou-se por aproximar o stock de capital e respectiva depreciação a partir dos respectivos valores contabilísticos.

obtenção de funções procura induzida para estes, que ao ser substituída na função custos, faz com que nesta última apenas figurem o capital e o trabalho.

Para o período 1988/92 o autor dispõe de 140 observações relativas a todos os bancos a operar em Portugal, com excepção da Caixa Geral de Depósitos, cuja presença na amostra piorava significativamente o ajustamento¹⁹. A função custos, nas suas versões (2) e (3), é estimada como fronteira estocástica, usando o modelo semi-normal.

3.2. Resultados

Os resultados obtidos no primeiro estudo (Mendes, 1990) para o período 1985-88 apontam para a relevância da solução Box-Cox quando se adopta a formulação (2), o mesmo não acontecendo quando a opção recai sobre o modelo mais geral (3). No primeiro daqueles casos, em que as medidas encontradas devem ser interpretadas num contexto "*firm*", o autor encontra significativas economias de escala para os bancos de mais pequena dimensão (activos abaixo de 3,6 milhões de contos a preços de 1997), sendo detectadas deseconomias de escala para quase todos os restantes (activos acima de 7 milhões de contos, preços de 1977). Curiosamente, quando se passa a usar uma medida do tipo *expansion path* (18) as conclusões mantêm-se, com excepção da descoberta de economias de escala para os bancos que se movem para dimensões acima dos 73 milhões de contos (preços de 1977).

Do mesmo modo, só são detectadas economias de *scope* entre os *outputs* considerados para os bancos de mais pequena dimensão, sendo encontradas deseconomias para todos os restantes, qualquer que seja a medida utilizada (20) ou (22).

Neste estudo é igualmente estimada uma versão onde os balcões interagem com as restantes variáveis do modelo (3). As medidas assim encontradas devem ser interpretadas num contexto "*plant*", visto que não foram publicadas medidas corrigidas de acordo com (17). Embora as medidas *ray* pareçam apontar para rendimentos constantes à escala (excepto para a classe de menor dimensão), as

¹⁹Os diferentes testes realizados levaram a considerar a CGD como um "outlier" naquela amostra. Aliás, deve-se ter em conta que parte da sua actividade naquele período se deve ao seu relacionamento especial com o Estado, tornando difícil, portanto, considerar que partilha da "tecnologia" utilizada por instituições que desenvolviam a sua actividade num regime mais competitivo.

medidas *expansion path* apontam para a existência de economias de escala em todas as classes. Contudo, é preciso ter presente que estes resultados são apenas válidos para expansões de actividade nas quais o numero de balcões permaneça constante. Aliás, no outro estudo a que vimos fazendo referência (Pinho, 1994) são encontradas, a este nível, economias de escala ainda mais fortes, as quais são posteriormente anuladas ao nível "*firm*" quando se introduz o efeito da expansão da rede de balcões dada em (17).

Os resultados obtidos no segundo estudo (Pinho, 1994) corroboram a existência de economias de escala para os bancos de menor dimensão (activos abaixo dos 20 milhões de contos, preços de 1986). Contudo, alguns problemas são encontrados para bancos situados numa dimensão intermédia, onde é possível a existência de algumas deseconomias de escala ao nível *firm*. Ao nível *plant* são detectadas economias de escala para todos os bancos, as quais são decrescentes com a dimensão. A correcção do impacto da expansão da rede de balcões (17), mostra um efeito bastante importante: a produtividade da rede de balcões é crescente com a dimensão do banco. Assim, passam a jogar-se dois efeitos contraditórios, os quais parecem anular-se, não sendo possível afirmar a existência de economias de escala para os bancos de maior dimensão.

Neste estudo, as medidas de economias de *scope* são medidas de acordo com a expressão (22). Com excepção dos bancos de maior dimensão (expansões para níveis acima de 500 milhões de contos, preços de 1986), são detectadas significativas economias associadas à produção conjunta de depósitos e crédito. Para os grandes bancos tais economias poderão desaparecer em virtude de uma maior especialização interna de funções que conduz à inexistência de uma significativa partilha de recursos entre as duas actividades.

As estimativas de eficiência apontam para um grau de eficiência produtiva de cerca de 82%, o que parece indicar que 18% dos recursos utilizados pelo sistema bancário resultam de desperdício puro. Embora os padrões de eficiência produtiva individuais sejam relativamente dispersos, alguns padrões de regularidade foram encontrados. Assim, os bancos públicos são normalmente mais ineficientes, o mesmo acontecendo com os bancos estrangeiros de retalho. Por outro lado, os bancos exibem indicadores de eficiência mais baixos nos períodos em que iniciam fases de expansão, nem sempre recuperando nos períodos posteriores. Com uma excepção, todos os bancos privatizados melhoraram a sua eficiência após abandonarem o sector público. Significativa é a detecção de um efeito temporal que parece indicar que, ao longo dos últimos anos, a eficiência produtiva do sector tem evoluído no sentido positivo.

4. Conclusões

Neste artigo são analisados dois estudos sobre a estrutura de custos do sector bancário português. O facto de cobrirem períodos apenas parcialmente coincidentes, usarem metodologias econométricas diferentes e divergirem na classificação de *inputs* e *outputs*, torna-os complementares e motiva a necessidade de uma comparação dos resultados obtidos.

Ambos os estudos apontam para a existência de economias de escala na banca portuguesa para os casos em que a expansão não seja conduzida pelo aumento da rede de balcões. Nos casos restantes, aliás mais frequentes, tais economias desaparecem para a maioria das instituições, apenas se mantendo para os bancos de menor dimensão.

As economias de *scope* dependem, naturalmente, da classificação de *inputs* e *outputs* adoptada em cada estudo. Assim, Mendes (1990, 1991) não encontra significativas economias associadas à produção conjunta de crédito, aplicações financeiras e actividade extrapatrimonial, detectando mesmo significativas deseconomias para os bancos de maior dimensão. Por seu turno, Pinho (1994) detecta economias resultantes da produção conjunta de crédito e depósitos, as quas só não se verificam para os bancos de maior dimensão.

O nível de eficiência produtiva da banca portuguesa é inferior ao que habitualmente é medido em estudos relativos a outros mercados. Entre os factores associados a ineficiência parecem estar a propriedade pública do banco e estratégias de entrada no retalho por parte de pequenos bancos estrangeiros. No conjunto, detecta-se uma progressão positiva do grau de eficiência ao longo do tempo.

Assim dos estudos realizados não podemos concluir que estratégias de crescimento no nosso sistema bancário venham a resultar em reduções de custos associadas a economias de escala. Tais poupanças de custos terão de passar, antes de tudo, por políticas de controle de custos e utilização racional de recursos que viabilizem a eliminação de desperdícios e o aumento da eficiência produtiva na banca. A explicação para a vontade de crescer manifestada por várias instituições terá, portanto, que ser encontrada ao nível do exercício de poder de mercado.

Referências

- AIGNER, Dennis, LOVELL, C. A. and SCHMIDT, Peter "Formulation and Estimation of Stochastic Frontier Production Function Models", *Journal of Econometrics*, 6 (1977), pp 21-37.
- ALL, Agha and SEIFORD, Lawrence "The Mathematical Programming Approach to Efficiency Analysis", in Fried, Lovell and Schmidt (eds) *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, Oxford (1993).
- BAUER, Paul BERGER, Allen and HUMPHREY, David "Efficiency and Productivity Growth in U.S. Banking" in Fried, Lovell and Schmidt (eds) *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, Oxford (1993).
- BENSTON, George, HANWECK, Gerald and HUMPHREY, David "Scale Economies in Banking: A Restructuring and Reassessment", *Journal of Money, Credit and Banking*, vol 14 (4), (November 1982, Part I), pp 435-456.
- BERG, S. Atle and KIM, Moshe "Oligopolist Interdependence and the Structure of Production in Banking: An Empirical Approach", *Journal of Money, Credit and Banking* (May, 1994).
- BERGER, Allen, HANCOCK, Diana and HUMPHREY, David "Bank Efficiency Derived from the Profit Function", *Journal of Banking and Finance*, 17 (2-3), (April 1993), pp 317-348.
- BERGER, Allen, HANWECK, Gerald, and HUMPHREY, David, "Competitive Viability in Banking: Scale, Scope and Product Mix Economies", *Journal of Monetary Economics*, 20, (1987), pp 501-520.
- BERGER, Allen and HUMPHREY, David "The Dominance of Inefficiencies over Scale and Product Mix Economies in Banking", *Journal of Monetary Economics*, 28 (1991), pp 117-148.

- BERGER, Allen, HUNTER, William and TIMME, Stephen "The Efficiency of Financial Institutions: A Review and Preview of Research Past, Present and Future", *Journal of Banking and Finance*, 17 (2-3), (April 1993), pp 221-250.
- BOX, G. E. and COX, D. R. "An Analysis of Transformations", *Journal of the Royal Statistical Society, Series B* (1964), pp 211-243.
- CEBONOYAN, A., COOPERMAN, E., REGISTER, C. and HUDGYNS, S. "The Relative Efficiency of Stock versus Mutual S&L's: A Stochastic Cost Frontier Approach", *Journal of Financial Services Research*, (1993), pp 151-170.
- CLARK, Jeffrey "Economies of Scale and Scope at Depository Financial Institutions: A Review of the Literature", *Economic Review*, Federal Reserve Bank of Kansas City, (September/October 1988).
- COLWELL, R. J. and DAVIS, E. P. "Output and Productivity in Banking", *Scandinavian Journal of Economics*, 94, Supplement (1992), pp S111-S129.
- EVANOFF, Douglas and ISRAILEVICH, Philip "Productive Efficiency in Banking", *Economic Review*, Federal Reserve Bank of Chicago, (July/August 1991).
- FARREL, M. "The Measurement of Productive Efficiency", *Journal of the Royal Statistical Society*, 120 (1957), pp 253-281.
- GREENE, William "Maximum Likelihood Estimation of Stochastic Frontier Production Models", *Journal of Econometrics*, 18 (1982), pp 285-289.
- GREENE, William "The Econometric Approach to Efficiency Analysis", in Fried, Lovell and Schmidt (eds) *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, Oxford (1993).
- JONDROW, James, LOVELL, C., MATEROV, Ivan and SCHMIDT, Peter "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model", *Journal of Econometrics*, 19 (1982), pp 233-238.

- KOLARI, James and ZARDKOOHI, Asghar *Bank Costs, Structure and Performance*, Lexington Books, Lexington MA (1987).
- KOPP, Raymond and DIEWERT, W. "The Decomposition of Frontier Cost Function Deviations into Measures of Technical and Allocative Efficiency", *Journal of Econometrics*, 19 (1982), pp 319-331.
- MENDES, Vitor *The Portuguese Banking Industry in 1965-88: Analysis of Scale and Scope Economies*, Doctoral Dissertation, University of South Carolina (1990).
- MENDES, Vitor "Scale and Scope Economies in Portuguese Commercial Banking: the years 1965-88", *Economia*, 15 (3), (Outubro 1991).
- MESTER, Loretta "Efficiency in the Savings and Loan Industry", *Journal of Banking and Finance* 17, vols 2-3 (April 1993), pp 267-286.
- PANZAR, J. and WILLIG, R. "Economies of Scope", *American Economic Review*, 71 (2), (May 1981), pp 268-272.
- PINHO, Paulo *Essays on Banking*, Doctoral Dissertation, City University Business School, London (1994).
- PULLEY, Lawrence and HUMPHREY, David "The Role of Fixed Costs and Cost Complementarities in Determining Scope Economies and the Cost of Narrow Bank Proposals", *Journal of Business*, vol 66 -3 (1993).
- SEALEY, C. W. and LINDLEY, James "*Inputs, Outputs and a Theory of Production and Cost at Depository Financial Institutions*", *Journal of Finance*, 34 (September, 1977), pp 1251-1265.
- SHEPHERD, Ronald *Cost and Production Functions*, Princeton University Press, Princeton (1953).

SHEPHERD, Ronald *Theory of Cost and Production Functions*, Princeton University Press, Princeton (1970).

WALDMAN, Donald "Estimation in Economic Frontier Functions", (mimeo), University of Wisconsin (1978).