Public Good Aspects of target: Natural Monopoly, Scale Economies, and Cost Allocation
Public Good Aspects of TARGET: Natural Monopoly, Scale Economies, and Cost Allocation

Wilko Bolt and David B. Humphrey *

* Views expressed are those of the individual authors and do not necessarily reflect official positions of De Nederlandsche Bank.
ABSTRACT

Public Good Aspects of TARGET: Natural Monopoly, Scale Economies, and Cost Allocation

Wilko Bolt * and David B. Humphrey **

This paper discusses various theoretic concepts which play a role in assessing the public benefits of Target, the large value RTGS payment network operated by the Eurosystem. These concepts touch upon natural monopoly, network externalities, competition and contestability, as well as economies of scale and scope. Based on empirical results for the Federal Reserve’s payment system (Fedwire), it is argued that if Target decided to standardize its operating platforms and consolidate its processing sites into one or a few centers, it too could realize strong scale economy benefits and lower unit costs.

The main thrust of the paper concerns natural monopoly and the possibility of lowering unit payment processing cost via economies of scale. The existence of a natural monopoly provides a rationale for a temporary partial or full subsidy in order for Target to achieve the ‘most efficient scale’ or apply the most efficient technology to lower unit costs. Such a subsidy could be implemented through temporary ‘penetration’ pricing (i.e., pricing at less than full current cost). When the lower costs are realized, the subsidy would be removed and full cost pricing implemented. Once users face the full costs of their payment decisions, they are better able to match benefits with actual costs and implement a more efficient allocation of payment resources than occurs today on Target.

Key words: public good, natural monopoly, most efficient scale, partial subsidy

JEL codes: G20, H41, L10

* Wilko Bolt: Research Department, De Nederlandsche Bank, Amsterdam, The Netherlands, e-mail corresponding author: w.bolt@dnb.nl
** Dave Humphrey, Department of Finance, Florida State University, Tallahassee, US.
Any views expressed in this paper are those of the authors and do not necessarily reflect the views of De Nederlandsche Bank, or the European System of Central Banks. Comments by Lex Hoogduin, Philipp Hartmann, and the members of the ECB Study Group on ‘The public good factor in Target2’ are acknowledged and appreciated.
1 INTRODUCTION

In a rapidly changing payment landscape, a lively policy debate was recently triggered about the cost effectiveness and the quantification of a possible ‘public good factor’ of the large-value payment system Target. This paper discusses the various public good issues related to the benefits that accrue to an interconnected Target payment network, the implications for market competition in large value payment services, how unit costs may change as payment volume expands on a network with standardized hardware and software, and the potential usefulness of knowing for pricing purposes how costs are associated with the different services provided by Target.

Broadly, the analysis will be carried out along two parallel lines. First, economic theory is used to identify relevant concepts that play a role in assessing the public good character of large-value payment systems. Theoretically, we take the position that achieving a ‘most efficient scale’ could justify a full or partial subsidy on Target for a limited time period, until the point at which potential scale economies are fully realized given the size of the market. Second, learning from the Fedwire consolidation and standardization experience in the eighties and nineties, in the empirical part it is argued that if Target decided to consolidate its processing sites into one or few centers, complemented by the implementation of new standardized processing technologies, it could realize strong scale benefits and markedly lower unit costs. Temporary ‘penetration’ pricing in the form of subsidizing Target payments could achieve these benefits more rapidly and result, once the lower unit costs are realized, in full cost pricing to users while saving resources now expended by central banks but are unrecovered in current pricing arrangements.

The argument for a subsidy is based on the fact that if full cost pricing were implemented on the current Target system, scale economy benefits either would not be realized (as a portion of current users would find alternative payment arrangements/networks) or would not be realized as rapidly (extending the period where currently high unit costs are incurred). Once lower unit costs are achieved on Target, the subsidy would be removed and full cost pricing implemented so that users would face the full costs of their payment decisions. This matches user benefits with user costs, a result which more properly allocates resources in a society.

The paper is organized as follows. In section 2, concepts that touch upon natural monopoly, network externalities, competition and contestability are discussed. Also the relevance of the
theory of two-sided markets for payment systems is briefly examined and possible lessons from
the telecommunications industry are outlined. Section 3 provides an empirical analysis suggesting
that large value payment operations, when properly structured, contain elements of a natural
monopoly and can experience strong scale benefits. The second part of the paper is concerned
with issues of cost allocation and the implications for pricing when scope economies exist on
Target. Although data are lacking, in section 4, we address the ways in which costs may be
allocated on a network such as Target which provides monetary control services for central banks
as well as funds transfers for the private banking sector. Finally, section 5 concludes.
2 PAYMENT SYSTEMS: NATURAL MONOPOLY, NETWORK EFFECTS, AND COMPETITION

Payment systems are essential components of well functioning economies and financial markets, facilitating the exchange of goods, services and assets. The speed, ease and reliability with which payments are processed and executed will in general affect economic activities, output and price levels. As such, one of the ECB’s key tasks is to promote and ensure efficient and safe payment systems. Below, some theoretic concepts that are useful for the assessment of public benefits of Target are presented.

2.1 Efficiency, natural monopoly, and scale

Before the public benefits of Target are analysed, it is convenient to first discuss briefly the concept of efficiency in payment systems. Payment systems impose resource costs on society so, at least from a theoretic viewpoint, it is important that the provision of payment services satisfies some basic principles of efficiency. But efficiency in payment systems is not easy to define, since many diverse aspects play a role here. Not only prices, social and private costs of making payments are important, but also such aspects as speed, convenience, safety, and reliability are contributing factors to the efficiency of payment systems. Loosely speaking, a payment system is said to be efficient if the net benefits it provides to society are being maximised.

More specifically, in economic theory three dimensions of efficiency are often identified:
- allocative efficiency: the extent to which total welfare of consumers and suppliers is being maximized,
- productive efficiency: the extent to which costs of production are being minimised given the current level of technology,
- dynamic efficiency: the extent to which suppliers are able to meet the changing needs and preferences of consumers over time (‘product innovation’) in the most efficient way (‘process innovation’).

Whereas allocative and productive efficiency (static efficiency) focus on ‘how to split up the pie’, dynamic efficiency refers to the ‘size of the pie’. These efficiency concepts will be useful when assessing the ‘public good factor’ of Target.
Payment systems are characterized by strong economies of scale and are often said to be natural monopolies. Several interpretations can be given to a natural monopoly. Theoretically, one may define an industry as a natural monopoly if, over the relevant range of demand, the cost function is subadditive. Subadditivity here means that it costs less to produce the various outputs together than to produce them separately. Formally, a cost function \( C(q) \) is said to be strictly subadditive if, for any \( n \) – tuple of outputs \( q_1, \ldots, q_n \) it holds that

\[
\sum_{i=1}^{n} C(q_i) > C(\sum_{i=1}^{n} q_i)
\]

If this is the case then a well-informed social planner would have no incentive to have several firms produce the output when the total aggregated volume could be produced more cheaply, thus more efficiently, by a single firm. Alternatively, in a more positive sense, one may look at actual behaviour in an unregulated industry. Assuming that a firm’s profits decrease with the number of firms in the industry (due to competition), an industry is said to be a natural monopoly if only one firm is making positive profits, but not two or more.

Cost subadditivity is particularly important here because it sheds light on the ‘most efficient scale’ of an industry. The most efficient scale is reached whenever the marginal cost curve intersects the average cost curve, which is at the point of minimum average costs. At this point scale economies are fully exhausted. In our view, the concepts of natural monopoly and cost additivity provide a rationale for a temporary partial or full subsidy on Target. We argue that as long as the most efficient scale has not yet been reached, a ‘penetration’ pricing policy in the form of a subsidy (i.e. setting prices lower than marginal costs) might be employed to boost demand for Target payments and achieve a higher, more efficient scale. However, after the potential scale benefits have been fully realized, this type of subsidization is removed, since it would then only distort allocative efficiency. Implementation of full cost pricing would better align user benefits with user costs, resulting in a more efficient allocation of resources.

In empirical studies, scale economies (SE) are often measured by the percent change in total costs divided by the percent change in transaction volume. That is,

1 The discussion here closely follows Tirole (1989), chapter 1.
2 It is easy to show that everywhere-decreasing marginal costs imply everywhere-decreasing average costs, which implies subadditivity.
Values of $SE < 1$ indicate the degree of scale economies, while $SE = 1$ points to constant returns to scale. In subsection 3.1, we will tentatively examine the degree of scale economies that exist across 21 large value payment systems in 2001.

Not only scale economies but also scope economies may exist among the various payments services supplied on a large value payment network. In particular, Target currently provides both government-related services and payment services to private banks of financial institutions. Below, in section 4.2, we will further motivate that the joint costs of providing government-related and banking industry services on Target are likely to be lower than the ‘stand-alone’ costs of providing these services separately. In theory, it is straightforward to generalize the concept of subadditivity for multiproduct firms, allowing a useful formulation of scope economies. Suppose that $q = (q_1, \cdots, q_k)$ is a production plan for the $k$ different outputs. Let $q^1, \cdots, q^m$ denote $m$ such production plans. Then (strict) subadditivity for multiproduct firms implies

$$
\sum_{i=1}^{m} C(q^i) > C(\sum_{i=1}^{m} q^i)
$$

for all $q$ such that $\sum_{i=1}^{m} q^i \neq 0$. To illustrate the concept of economies of scope, consider a two-product economy where $(q_1, q_2)$ denotes the two quantities of the two goods. Subadditivity in this two-type case means that

$$
C(q_1, 0) + C(0, q_2) > C(q_1, q_2)
$$

where $C(q_{1,0})$ and $C(0, q_2)$ denote the stand-alone costs of producing one good.

### 2.2 Network effects in payment systems

Payment systems are characterized by strong network externalities. Formally, network externalities arise when a good or service is more valuable to a user the more users adopt the
same good or service. Since these externalities are not priced, full consolidation of payment systems might be socially desirable. Below, in section 3, it is argued that this argument may also apply to Target, justifying a temporary public subsidization in terms of low prices for Target payments.

In theory, network externalities affect both the demand and supply side of the market. On the demand side, end-users need to coordinate expectations on system usage giving rise to multiple equilibria. If, for example, end-users hold pessimistic beliefs as to (future) usage of Target 2, this ‘new’ system may not be able to achieve enough critical mass making full cost recovery extra cumbersome. The size of a network is a key factor for its total value. There is, however, a danger that the ‘winner takes it all’. That is, by attracting more and more users, an existing network may increase its edge over competing networks, ultimately pushing smaller networks out of the market. This is called ‘tipping’, where end-users are effectively locked-in by the single existing network. Additionally, the process of attaining a large ‘installed base’ may be path-dependent, so that some specific chain of events in the start-up phase of the network turns out to be a decisive factor for the outcome. Because of tipping, lock-in and path dependence, one may be trapped in a situation of ‘excess inertia’ if an existing dominant network technology would prevent end-users to switch to a better technology.

More specifically, consider two different payment systems, a new one (‘n’) and an old one (‘o’), which induce an end-user surplus $v_n(q)$, respectively $v_o(q)$, where $q$ denotes the number of users (network size). Every end-user can either stick to the old network or join the new one. Network size induces a positive externality so that $v_i(q) > v_i(q')$, $i = o, n$, if $q > q'$. Let us assume that all $q$ end-users prefer to coordinate their decision on whatever type of network, i.e. $v_n(q) > v_o(q')$ and $v_o(q) > v_n(q')$ for $q > q'$. Now suppose that these $q$ end-users simultaneously choose whether to stick to the incumbent network or switch to the new network, then it is obvious that the need to coordinate gives rise to a multiplicity of equilibria (in pure strategies): all end-users can either stick or switch. Excess inertia arises when $v_n(q) > v_o(q)$ but all users stick to the old network. Coordinating on joining the new payment network would be

---

3 Also, network goods often feature interdependence of production technologies, in the sense that a firm’s choice on how to produce a good, depends on the techniques chosen by other firms.
welfare improving, but each user is afraid of moving alone, and prefers to wait for the others to move first. On the supply side, compatibility, interoperability and standardisation are necessary ingredients for extending the size of a network. Networks combine complementary components of a technology that enables the provision of network goods and services (see McAndrews, 1997). In payment systems, compatibility can be achieved by adherence to technical standards, infrastructural arrangements and interbank cooperation. Standardisation and compatibility make sure that economies of scale are fully exhausted. However, one should be aware of the potential dangers. First, in the early stages of setting up the network, firms might have been coordinating on the wrong technology, that later on, having attained a large installed base, is difficult to replace. Second, standardisation and cooperation between suppliers of network goods might limit effective competition. Third, a dominant network provider may have an incentive to make incompatible products in order to improve its market position. As a result, allocative and dynamic efficiency may be reduced.

Finally, as already noted, network externalities and economies of scale may induce a monopolistic market structure since duplication of fixed costs is often not socially desirable. As is well known from economic theory, a monopolist sets excessive prices, underprovides quality, and may have little incentive to innovate. As J.R. Hicks already stated: ‘the best of all monopoly profits is a quiet life.’ Although enhancing a more efficient scale on the one hand, subsidizing prices may on the other hand aggravate the problem of insufficient competition and lack of contestability. This trade-off between static and dynamic efficiency must be taken on board when designing an adequate pricing structure for Target payments.

2.3 Competition and contestability

The market for large-value payments tends to be highly concentrated, where central banks and big commercial banks dominate the whole landscape. The question rises whether this high concentration hampers adequate competition. In practice, no market is perfect: the theoretically

---

4 The opposite case, when \( v_i(q) > v_j(q) \) but all end-users switch to the new network, is called ‘excess momentum’. End-users join the new network out of fear of getting stranded with the old network technology.
ideal situation of perfect competition is never reached. What often matters in competition policy is the criterion of ‘effective competition’. Effective competition refers to a market environment which shows a wide choice of varieties of goods and services for end-users at attractive price-quality levels, and ‘normal’ profits for providers, with substantial innovative potential. Effective competition is one of the main drivers for increasing efficiency and market performance. It positively affects allocative efficiency through a better match of demand and supply and increased transparency in prices; it raises productive efficiency through the intensified search for cost-reducing production methods; and it increases dynamic efficiency through stronger incentives for producers to continuously innovate.

Modern competition theory often sketches two rough options. The first option points to ‘competition in the market’ and touches upon the concept of contestable markets. Contestability states that the mere threat of competition of potential entrants is already sufficient to constrain the behaviour of incumbents such that efficient prices and resource allocations will result. The disciplining effect of potential competition in a market with few players can be just as strong as actual competition in a market with many players. If markets are contestable, active price regulation is not necessary. However, a necessary condition for contestability to occur is the removal of barriers to entry to and exit from the market. In the context of competition between Target and the private large value payment network Euro1, allowing a price subsidy for Target makes the market less contestable and might increase economic barriers for Euro1 to enter the market.

The second option is dubbed ‘competition for the market’ and deals with infrastructural scarcity and competition between networks. Electronic payment systems typically have large set-up costs because of their relative capital intensity in terms of building high-technology electronic networks. Since duplication of these huge fixed costs does not seem desirable, actual network competition is often not a viable option. However, it is important to note that network competition may cause a trade-off between static and dynamic efficiency. In the short term, end-users would benefit from lower prices of services and goods, however, in the long run, low prices may lead to underinvestment in new network technologies. One runs the risk that newer, more advanced networks will not be able to gain critical mass.

The question comes to front how pure network activities can be separated from the provision of goods and services over the network, and whether competition in this latter segment is actually
possible. The ability for a potential entrant to access the existing network becomes an important issue here. This type of competition is usually called ‘system’ or ‘platform’ competition, which also triggers the complex issue of access pricing. The incumbent network may levy fees on the new entrants for accessing its network. If the incumbent network owner would charge a too high access fee (e.g. a fee higher than its own end-user price), then it is obvious that competition cannot be of any benefit to end-users. At the same time, access fees cannot be set too low, in order to prevent inefficient competitors to enter the market causing productive inefficiencies. The practice of setting high access fees is a form of ‘raising rival costs’. This issue may arise in the context of Target and Euro1, when ultimately Euro1 must access the Target network for final settlement.

2.4 Two-sided markets and payment systems

Theoretically, in assessing the public benefits and the social welfare aspects of payment systems, it is useful to note the market for large-value payments is in effect a two-sided market, where payees and payors (initiators and receivers) ‘consume’ the good (the payment). For a payment to be executed in Target you need both sides ‘on board’ using Target. Recently, there has been a surge of interest in two-sided markets. Rochet and Tirole (2004) propose a general definition of two-sided markets, identify the conditions under which ‘two-sidedness’ of a market occurs, and discuss its importance for business and public policies.

Two-sided markets involve two distinct types of end-users, each of whom derives value from interacting with users of the opposite type. In these markets, ‘platforms’ coordinate the demands of these distinct types accounting for the interactions between the two types on the demand side when devising pricing and investment policies. A key aspect of two-sided markets involves the optimal price structure, that is, the division of the revenues between the two sides of the markets that get on board. Optimal prices for the different types of end-users must balance the demand among these types - and the need for an optimal pricing structure as well as an optimal pricing level distinguishes a one-sided from a two-sided market. As Evans (2003) points out, in two-sided markets the demand for the product may completely vanish if the business does not get the pricing right.

Rochet and Tirole (2004) define a two-sided market as a market in which the total volume of transactions varies with the price for one side of the market while keeping the total price constant.
In a stylized theoretic framework, they show that the distribution of the total fee over the two types depends on the demand elasticities of the product. Although in their framework the optimal total fee obeys the well-known inverse-elasticity rule for monopolistic price setting, it may indeed be the case that the optimal price for one side of the market will be set below its marginal cost. These two-sided pricing principles may also be important for large-value payment markets. In setting the prices for a large value payment, the network switch (Target) needs to get both types on board. Currently, only one side of the market (initiators of payments) is charged for executing payments in Target. Although, on average, this pricing scheme will not result in major cost differentials between the two types of Target users (each user likely to be initiator and receiver half of the time), it does affect the total realized volume within Target.

Platform competition in two-sided markets is complex to analyse. However, within the same theoretic framework, it can be shown that the optimal pricing rule is qualitatively the same as in the monopoly case, except that the demand elasticities of the two types of end-users need to be modified to take into account the degree of ‘multihoming’ of end-users (i.e. connect to different platforms) and ‘steering’ by competitors (i.e. induce end-users to opt out of other competing platforms). Although the economic theory of multihoming and steering is not yet completely understood, it is clear that these aspects will play an important role for the competitive behaviour of Target and Euro1. Commercial banks may in principle do business with both platforms Target and Euro1, while this opens up the possibility for those platforms to steer customers away from their competitors by the prices they set. In general, system competition puts downward pressure on prices on both sides of the market, and the resulting impact on relative prices is not sure. In particular, system competition does not necessarily induce a greater allocative efficiency.

In the context of economic efficiency, Wright (2003) warns that straightforward application of ‘one-sided logic’ to two-sided industries may lead to false economic statements and policy recommendations. Issues about market definition, market power, efficient pricing and competition, predation, and regulation must be seen from a different economic angle. This naturally also holds for social welfare and public benefits of two-sided markets.

Surprisingly, at an interior solution, the most elastic side is charged the highest fee in the monopoly situation, see Rochet and Tirole (2003). Bolt and Tieman (2003) show that the reverse may be true for corner solutions.
2.5 Lessons from the telecommunications industry

The telecommunications business went through a significant change over the last two decades. Rapid technological progress and innovation in key inputs of telecommunications and computer-based services were the key driver for this change, which induced dramatic cost reductions and shifted demand conditions considerably. Moreover, the underlying telecommunications technology has become more and more digitalized moving towards multi-function programmable devices and interfaces. And as a result of competition, the telecommunications sector today consists of a myriad of different electronic networks. The complexity of the telecommunications industry in terms of pricing, access and interconnection, and essential facilities are a paramount policy concern for regulators and antitrust authorities.

Payment systems cannot exist without telecommunication networks. Hence, the study of telecommunication networks may provide useful insights about the functioning and design of payment system networks. Mason and Valletti (2001) describe four defining characteristics of communication networks that may also apply to payment systems. The first characteristic is the specific cost structure of networks, where one typically incurs large fixed cost attached to setting up the network but faces low marginal cost of operating the network. Strong complementarity between the different network components is the second factor, so that there are large gains to connecting two networks, but can also create a bottleneck problem. The third factor is demand externalities which, in short, states that networks are more valuable if there are more people using them. The fourth characteristic refers to social obligations, in the sense that these networks are often viewed as providing essential services. On top of these, we may add that both industries are two-sided where we distinguish payees and payors in payment systems, and senders and receivers in telecommunication networks. The combination of these five characteristics pose a real challenge to the analysis of pricing, social benefits and regulation of these networks.

The experience of the telecommunication firm AT&T in the US offers some parallels to determining the public good benefits/costs associated with Target. In particular, AT&T operated as a legal monopoly in providing both local and long-distance phone service. This allowed it to achieve lower unit costs due to the existence of marked scale economies in the provision of both local and long-distance phone service. AT&T was treated as if it were a natural monopoly. That is, in producing products or services where large scale economies exist, the public benefits of having a single producer results in the lowest cost of production. To ensure that these low costs
are passed on to consumers, prices were regulated to cover all cost plus a normal return on invested capital.

At the time, AT&T was charging prices lower than costs for local service but higher than costs on long-distance. When combined, all costs were covered but the pricing structure led to challenges by potential new entrants who wished to enter the long distance market but could only do so if AT&T was no longer deemed a natural monopoly. It was determined that the cost of the joint provision of local and long distance phone services by AT&T alone was not notably lower than if these two services would be provided separately by different firms. As scope economies did not apparently importantly exist between local and long distance, even though both services experience very large scale economies, AT&T was broken up and new firms entered the long distance market. The relevance of the AT&T experience and other telecom businesses to Target is that Target produces two services - government payment services as well transaction services used by banks. The same scope economy issue faced in the telecommunication industry also applies to Target. However, data is currently lacking to do this in practice although the procedure that one could follow is outlined in subsection 4.2 below.

In addition, it could be useful to explore the possibility that infrastructure and service provision can be workably separated on payment networks as they sometimes are in telecommunication arrangements, such as where a long-distance company purchases access to users through local service suppliers. If this is a reasonable possibility, then existing access to users of one large value payment network could be "merged" with processing these user's payments on a different network and costs may be saved. However, we expect that given the small number of users of large value payment networks and the relatively low expense of providing a direct connection between a user and a payment processor that the analogy of long-distance phone companies purchasing access to users through local phone networks would not involve significant savings compared to the complexity of actually implementing such an arrangement on a large value payment network. Moreover, this raises the issue of access and interconnection pricing, key to competition in telecommunications, which is generally very technical and requires an in-depth knowledge of the industry 6. Following Vogelsang (2003) these pricing issues have not yet been resolved.

6 For a formal theoretic approach to competitional issues in telecommunications, see Laffont and Tirole (2000).
3 SCALE ECONOMIES: THE EFFECTS OF CONSOLIDATION AND STANDARDIZATION ON FEDWIRE

This section deals with scale economies and whether public benefits can be expected from consolidating and standardizing the current processing sites of Target. In subsection 3.1 information is presented suggesting that scale economies are present on different large value networks. Moreover, the specific experience of Fedwire confirms this finding and is explored further in subsection 3.2. The existence of significant scale economies on large value networks can be used (via a natural monopoly argument) to justify a full or partial subsidy to Target for a limited time period - until the point at which potential scale economies are fully realized up to the limit imposed by the size of the market. Realizing such economies would permit Target to lower unit costs and, after the temporary subsidy is removed, implement full cost pricing so users of payment services can better match benefits with costs and generate a more proper allocation of resources than occurs currently on Target. The related concept of scope economies is examined in subsection 4.2 although a lack of appropriate data limit the meaningful application of this concept to Target at present.

3.1 Estimates of scale economies on large-value payment networks

Scale economies exist on payment networks as unit cost per transaction typically falls as volume is increased. This is seen in a recent analysis showing how unit cost falls as transaction volume expands across 11 Target processing sites in 2000. The figure shown there reflects strong scale effects when processed volume varies between 1 to 5 million transactions annually but unit costs appear to be flat for volumes greater than 5 million. While 22 million transactions were processed at the largest site on Target in 2000, two other analyses have found that unit cost continues to fall, bringing scale benefits, when annual volume exceeds the experience on Target.

An empirical study of payment costs on 21 public and private large value payment networks across 20 countries for 2001 found that the unit cost per transaction averaged $1.30 and ranged

between $0.29 to $3.91 (Khiaonarong, 2003). This also tended to vary by region, as shown in Table 1.

Table 1: Average Unit Cost Per Transaction on 21 Large Value Payment Networks (in U.S. dollars, 2001)

<table>
<thead>
<tr>
<th>Region</th>
<th>Number of Networks</th>
<th>Average Unit Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>European Union</td>
<td>10</td>
<td>$1.54</td>
</tr>
<tr>
<td>East Asia-Pacific</td>
<td>9</td>
<td>1.15</td>
</tr>
<tr>
<td>North America</td>
<td>2</td>
<td>0.84</td>
</tr>
<tr>
<td>Total</td>
<td>21</td>
<td>$1.30</td>
</tr>
</tbody>
</table>

Source: Adapted from Khiaonarong (2003) Table 6, page 29

Regressing unit or average cost (AC) on transaction volume (Vol) for these 21 large value networks gave the following results:

\[
\ln AC = 3.83 - 0.58 \ln Vol
\]  

where both parameters were significantly different from zero at the 99% level of confidence. Other variables added to this simple regression gave significant scale effect parameters ranging between -0.49 to -0.67. Focusing on the result in (1), this says that if transaction volume on a large value payment network were to double then average cost per transaction on average falls by 58%. Taking the average cost per transaction of $1.30 across the 21 networks, a doubling of transaction volume suggests that average cost could fall by perhaps $0.75 to a level of around $0.55 if no prices rose during this period (i.e., if inflation was zero). At present only the very largest Target processing sites seem to incur a unit cost of around $1.30 per transaction (others are higher). If transaction growth was 10% a year, as it was on Target over 1999-2002, it would take about 7.5 years to double transaction volume.

---

8 Unit costs in local currency values were translated into U.S. dollars at market exchange rates. Had purchasing power rates been used these average values would be somewhat higher. Also, not all unit costs were for 2001 as 3 were for 2000 and 2 were a forecast for 2002. In the range of unit costs shown in the text, one (clearly non-comparable) network in Khiaonarong (2003) was ignored.

9 In a cross-section study such as this, the common maintained hypothesis is that a payment network with relatively low volume and relatively high unit cost would ‘look like’ the average of payment networks with higher volumes and lower unit costs if its volume were to expand.
Greater accuracy in estimating scale effects could probably be obtained if (1) had been specified as a log quadratic with $(\ln\text{Vol})^2$ as an additional explanatory variable to capture better how scale benefits tend to be reduced as larger and larger volumes are attained. As well, the estimated scale parameter in (1) will incorporate differences in payment processing or telecommunication technology as well as different wage rates that likely exist among the 21 networks. Since higher volume networks are likely to employ more efficient processing and telecommunications technology, it is possible that the measured scale effect here - that average cost falls by 0.58% for each 1% rise in transaction volume - may be overstated. This is because the cost effects from differences in technology and input prices have not been held constant in (1). Even so, the estimate above gives a good first approximation of how strong scale effects may be on large value payment networks.

As already discussed in section 2.1, the usual way to determine scale economies (SE) is to relate the total cost (TC) of making a payment to the associated transaction volume on a network (Vol) so that $SE = \Delta\ln TC / \Delta\ln Vol$, which equals the percent change in total cost divided by the percent change in transaction volume. Values of $SE < 1.0$ indicate the degree of scale economies while an $SE = 1.0$ would reflect constant average costs as volume changes. Since (1) can be re-expressed as $\ln TC = 3.83 + 0.42 \ln Vol$, the cross-section relationship for 21 large value payment networks in 2001 yields a scale economy value of $SE = .42$. Thus if transaction volume were to double, total cost would be expected to only rise by 42% and average cost would fall. Equation (1) tells us the percent change in average cost as volume expands across the different networks, networks with a wider range of annual transaction volumes than currently exists on Target.

A more comprehensive analysis of large value payment network scale effects concerns an effort to estimate the relative importance of three major determinants of the reduction in unit cost on Fedwire over 1979 to 1996 (Hancock, Humphrey, and Wilcox, 1999). Unit payment costs combine processing costs and telecommunication expenses. Processing costs are composed of

---

10. Due to confidentiality issues, transaction volumes of the various networks were not reported in Khiaoanong (2003) and, while unit costs are listed, they were not identified by their network. Thus it has not been possible to re-specify and re-estimate (1).

11. The re-expression relies on the fact that $\ln AC = \ln TC - \ln Vol$ so that (1) can be rewritten as $\ln TC = 3.83 + (1 - 0.58) \ln Vol$. Had the original specification of (1) been $\ln TC = a + b \ln Vol$, the econometric result would have been $\ln TC = 3.83 + 0.42 \ln Vol$.

12. Scale economies reflect what happens to total or average cost when output rises but typically holds constant the effect that concurrent changes in input prices and technology may have on the output-associated change in cost.
labor, building, and computer expenses to electronically transfer funds from one party to another while telecommunication and other associated expenses are incurred in sending and receiving payment messages from the funds transfer initiating and receiving parties. Between 1979 and 1996, Fedwire average cost per transaction fell by about 24% in nominal terms (62% in real terms, which ‘removes’ the effect of concurrent price inflation).

The three major determinants of the reduction in Fedwire costs were: processing center consolidation, scale economies from expanded volume, and technological change which lowered processing and telecommunication costs directly. Data processing and telecommunication costs accounted for upwards of 86% of Fedwire expenses (the remaining costs were distributed among labor, building, and materials inputs). Data processing input prices (which reflected on-going technical improvements) fell by 51% over 1979-1996. Telecommunication input prices (which also reflected technical change and deregulation) rose by 38% over the same period while the prices of labor, buildings, and materials rose by 100% to 246%. The reduction in the price of data processing inputs relative to telecommunication inputs was associated with a rise in the quantity of data processing inputs relative to those for telecommunication, indicating that these two inputs are strong substitutes for one another. 13

Simply put, large value payments can be processed in a distributed manner where there are many processing sites and relatively little use is made of telecommunications or processing can be centralized and greater use is made of telecommunications. The change in relative prices over 1979-1996 clearly favored centralized over distributed processing. Responding to price incentives and reflecting a need to control Fedwire costs due to legislation that required the Federal Reserve to fully price its payment services, the 12 separate Fedwire processing sites in 1979 were consolidated into a single site in 1996 (with one ‘hot’ backup site that can immediately take over if problems arise and another ‘cold’ backup that could be operational after a delay of a number of hours). The cost reduction from consolidating processing operations was only partially offset by incurring greater telecommunication costs. In 1979, 36 customer service offices which handled wire transfer requests had telecommunication links with 12 processing sites. By 1996, there were 12 customer service offices (one in each Federal Reserve District) that communicated with a single processing center.

13 Quantities were derived by dividing nominal expenditures by the relevant technology or quality adjusted price indices.
Scale benefits on Fedwire were achieved in two ways. First, had Fedwire never consolidated, the 12 processing sites would have realized a lower average cost per transaction as payment volume grew by 136% over this period. Thus each site could have expanded by approximately 136%, although the site in New York, which processed close to half of the total, would likely have expanded somewhat more and the others less. Second, the result of consolidating 12 processing sites into just 1, say the largest site at the time - New York, would have meant a volume increase of 359%. While transaction volume rose by 136% nationwide, the consolidation of processing sites shifted this increase plus all of the volume processed at the other 11 sites into just 1 site. In combination, the total volume increase would have been 359%. Thus of the two ways scale benefits were realized on Fedwire, consolidation effectively more than doubled what would have occurred otherwise at the largest volume site in 1979.

The overall cost effect of consolidation, normal volume growth, and technical change on Fedwire unit cost is illustrated in Figure 1. The jagged line shows how actual Fedwire average cost per transaction (noted on the Y-axis) varied over 75 quarterly observations during 1979-1997 (numbered 1 to 75 on the X-axis). Over this period, annual Fedwire volume expanded from 35 to 83 million. The straight line is from a simple linear regression of average cost against time while the smooth curved line is a cubic spline fitted to the same data. Average costs on Fedwire fell from about $0.48 per transaction in 1979 to around $0.35 in 1996, a reduction of 26% in nominal terms. This cost reduction in nominal terms occurred even as the U.S. GDP deflator rose by 81% over this period and the cost-of-living index expanded by 116%.

Estimation of a cost function using Fedwire data on total costs, the number of payment transactions originated, the prices of data processing, telecommunications, labor, buildings, and materials inputs plus the number of processing sites and consumer service centers, along with three alternative representations of embodied technical change yielded scale economy values between .50 and .52. This suggests that a 1% rise in transaction volume is associated with only a .50% to .52% rise in total Fedwire costs. Re-expressing this result in terms of unit costs, a 1% rise in transaction volume is associated with a -.50% to -.48% change in average cost. This is

14 Most of the technical change was already accounted for in the technology or quality adjusted input prices for data processing and telecommunication inputs which either fell or rose more slowly than other prices in the U.S. economy. Panel data over time and across 12 Federal Reserve districts yielded the scale economy estimates reported above.
close to the -.58% change in average cost with a 1% rise in transaction volume reported above for a cross-section of 21 large value networks in 2001. Thus it is clear that strong scale economies exist for large value payment networks and that Target, Euro1, and other networks could expect to experience roughly similar scale effects with increases in their transaction volume over time.

Annual transaction volume on Target is around 65 million and its largest single processing site handled about 22 million transactions in 2000. This single site figure is considerably less than the single site volume of 115 million annual transactions on Fedwire or the 64 million on CHIPS. To obtain the benefits from scale economies, Target essentially has two choices: it can realize these economies over time consistent with its historical rate of volume growth of 10% a year or it can choose to consolidate its processing operations by folding some or most of its smaller volume sites into one of the currently larger sites. In the latter case, greater scale economies will be realized more rapidly. Alternatively, Target could adopt standardized and more efficient processing technology at each of its existing sites and this too could markedly reduce unit costs (Leinonen, 2002). It is unclear at this point which alternative - consolidation with new standardized technology or distributed processing at separate central bank sites with new standardized technology - may achieve the lowest cost. If the scale economies associated with consolidation are sufficiently large to more than offset the likely higher telecommunication costs of centralized processing, then unit cost is likely to be lower with centralized than with distributed processing. Even so, other considerations - such as operational ease and security - need to be considered along with unit cost. Of course, neither arrangement would disrupt the legal account relationship between banks and their individual central bank for settlement purposes.

The average cost per transaction on Target in 2000 was around 1.80 euros while the average cost at the two central banks with the greatest volume averaged around 1.15 euros. Although these figures are only approximate and preliminary, they contrast sharply with the current $0.34 cost

15 While the experience of Fedwire and CHIPS indicate that strong scale economies exist within a centralized processing framework, we have no similar experience to draw upon to assess the likely unit costs associated with implementing new and lower cost technology - along with related telecommunication and other expenses - within a distributed processing framework. Such a detailed and comparative study would be necessary before the ECB (and member central banks) could appropriately decide on the most cost effective configuration for Target.

16 ECB, "Overview of Target cost figures for 1999 and 2000", Unpublished, June 2001. At today's exchange rates, 1.15 euros translates into around $1.30, which was the average cost over 21 networks shown in Table 1.
per transaction on Fedwire and CHIPS which have centralized their processing operations 17. CHIPS processes about three times the volume as the largest volume site on Target (64 million compared to 22 million on Target) while Fedwire, which incurs greater telecommunication costs than does CHIPS, processes over five times as much (115 million compared to 22 million) 18. Since annual volume on Target is around 65 million, consolidating processing on Target would close to triple the volume that now exists at the largest processing site and yield considerably greater scale benefits than relying on the normal rate of volume growth of about 10% a year. Hence, by lowering overall unit costs and transaction prices, consolidation of processing centers would enhance allocative and productive efficiency.

3.2 Target and the public benefits of a natural monopoly

The existence of strong economies of scale, if realized, can confer large public benefits. Specifically, the cost of making a payment can be markedly reduced and resources would be released that find employment elsewhere in the economy. Financial markets, which intensively use large value payment networks, would experience lower transaction costs and output in the rest of the economy would be higher than otherwise. If scale economies remain strong as payment volume rises, the most efficient outcome would be to have only one supplier in the market -a natural monopoly. However, unless the single supplier is restricted in some manner to only charge a price that covers all its costs plus a normal return on invested capital, the elimination of competition that results tends to reduce incentives to actually produce at the very lowest cost and/or to continue to innovate and improve products or services for users. This would induce static and dynamic inefficiencies.

Given the strong scale economies that appear to exist on large value networks, an argument can be made that Target be considered a natural monopoly. To provide an incentive for potential users

17 CHIPS and Fedwire incur about the same unit costs ($0.34 per transaction) and both split this cost equally between the paying and receiving banks. Thus fees are about $0.17 for each party (Mohr, 2003). CHIPS is jointly owned by its commercial bank users and is operated as a "payment utility" that recovers all of its costs and funds transfers over Fedwire and CHIPS are in practice strong substitutes for one another. While this suggests that the type of ownership of a large value network per se has little effect on user choice, it does support the view that competition can be important for achieving low costs per transaction.

18 The low unit cost experienced on CHIPS is made possible by having a centralized processing facility with strong scale economies plus low telecommunication expenses (since all CHIPS transactions are made by entities located in New York City, rather than geographically distributed across the country as they are on Fedwire).
to expand transaction volume on Target and realize its full potential for scale economies, Target need not be required - at least initially - to cover all or most of its current costs. Instead, prices could be set to equal the lower unit cost expected to be incurred once most of its scale benefits (or benefits from improved processing technology - c.f., Leinonen, 2002) have been achieved. The social trade-off would then basically be a trade-off between static and dynamic efficiency. In particular, between:

1. The cost of a transition period which equals the discounted value of the expenses not recovered while Target expands its transaction volume at a consolidated processing site up to the limit established by the size of a market; and
2. The benefits of a mature system which equals the discounted value of the future cost savings from lowering expenses from their current level.

This comparison puts a premium on keeping the transition period as short as possible and illustrates the importance of determining when transaction volume growth has reached the limit associated with the size of the market. It is at this point that the justification for subsidizing unit cost only to achieve unrealized scale economies no longer applies or is very weak in economic terms.

At the present time, Target does not fully charge for the payment services it provides to the banking industry. Thus continuing this arrangement on Target, or only charging a low future expected price equivalent to unit cost once scale economies have been fully realized in the market (or new technology is in place), would be unlikely to result in any extra growth in payment volume over the "normal" growth it already experiences. Target, in its present configuration of some 16 separate processing sites, has already effectively reached the limit of the size of its market and future scale benefits will depend on the normal expansion of this market over time. Put differently, if Target were to continue with its current configuration of distributed processing and existing nonstandard technology, a scale economy argument for subsidizing unit cost really does not apply as the scale benefits associated with the size of the market have already been realized (although unit costs are higher than they would otherwise be with distributed processing.
However, Target could continue to be subsidized and not fully priced if - in addition to scale benefits achieved through "normal" growth - the provision of its payment services were determined to have strong public benefits in terms of ensuring financial stability resulting from its ability to: (a) create money and not financially fail to operate (as can occur on a private network); (b) reduce legal uncertainty in the event of participant failure on its network; and/or (c), provide full payment finality at lower social cost. Effectively, this has been the judgement the Bank of Japan made for its large value payment network. This consideration is developed more fully in Angelini and Maresca (2004) and Pagès and Humphrey (2004).

As Target has already reached the limit of its market share under its current configuration, a justification for continued subsidization on scale economy grounds alone can only reasonably be made if:

i) The scale benefit of having only one supplier for all large value payments is so strong that the provision of large value payments is deemed to be a regulated monopoly (similar to arguments that have in the past supported regulated monopolies for electricity and telephone services); or

ii) Target decides to consolidate its processing sites into one or a few centers to realize greater scale benefits (and lower costs). At least initially, these consolidated sites would not yet have realized the limit imposed by the size of the market since volume growth there would exceed normal, historical rates during a transition period.

There is no international agreement on the issue in (i). In the U.S., with two suppliers of large value payments (CHIPS and Fedwire), the regulated monopoly argument has been rejected in favor of competition versus greater scale benefits. In Japan this argument has been accepted and implemented with the full subsidization of their single large value network. Thus (i) is a judgement call on the part of policy makers who could point to arguments on both sides of this issue - but not rely on scale economies alone to make the case one way or another.

19 Put in terms of the infant industry argument often used to justify tariffs or quotas adopted to protect a domestic industry until a scale of operation and lower costs are achieved to allow it to compete with foreign imports, Target - in its current configuration - would not be considered to be an "infant industry" since it has already "grown up" and achieved the limit of scale benefits associated with the current size of the market under its current configuration.
The situation in (ii) could be used to justify full or partial subsidization on Target for a limited time period. Policy makers could decide to consolidate processing sites to achieve more rapidly scale economies compared to what the normal expansion of the payments market would bring. Here subsidization could be justified using the above criteria that compares the discounted cost of a subsidized transition period with the discounted cost savings obtained when a mature, consolidated system is achieved or when new standardized and lower cost technology is in place in a distributed processing system.

To summarize, strong scale economies exist on large value payment networks. In its current configuration with some 16 separate processing sites, Target has effectively realized all of the scale benefits it can expect given the size of the market and the existence of other payment suppliers (e.g., Euro1). At present, Target’s payment services are subsidized and additional scale benefits can only be realized as the size of the payment market continues to expand at historical rates of perhaps 10% a year. This arrangement can continue if it is decided that, like the Bank of Japan, the provision of Target’s services are essential for ensuring the stability of financial markets. Or, using the analogy to a regulated natural monopoly, policy makers could determine that there should only be one supplier of large value payment services in Europe. This would expand transaction volume at the current set of around 16 processing sites and pass on the lower costs from the expanded volume to users and society in general as other suppliers are forced (under regulatory authority or legislation) to exit the market. Alternatively, the subsidization of costs on Target can continue into the future during a transition period as payment volume at 16 processing sites are consolidated into 1 or 2 larger processing centers. Once the consolidation process is complete, the subsidization would cease and full cost pricing - at the lower level of costs being incurred - could be implemented.
4 COST RECOVERY, PRICING, AND SCOPE ECONOMIES ON TARGET

Large value payment systems are characterized by large fixed and sunk costs when setting up and maintaining the system. When the system is up and running, each additional payment generates very small costs. Such scale economies present a problem for efficient pricing and cost recovery. The possible existence of scope economies between central bank responsibilities for monetary control plus final settlement for other payment networks along with providing separate transaction services for the private sector complicates the pricing decision. Some issues that arise when analysing pricing policies, cost recovery and cost allocation of large-value payment systems are now discussed.

4.1 Pricing policy and cost recovery

Allocative efficiency is achieved when end-users of a payment system are charged a price equal to marginal costs since the extra benefit to the user is equated with the extra cost to the supplier. Marginal cost pricing corresponds to the ‘user’s pay principle’, avoids cross-subsidization schemes and induces efficient allocations. In effect, the user’s pay principle states that end-users of goods and services should themselves bear the costs they impose on society. Informative prices are the obvious channel which increase transparency and lead to a greater awareness by end-users to use more efficient modes of payment. Additionally, by removing (cross-)subsidies, other segments of the payment and banking sector get more efficient.

However, efficient pricing in payment systems is a complex matter. Despite its theoretic soundness, marginal cost pricing is often difficult to implement in practice. First, marginal costs are hard to accurately measure and, second, the problem of cost recovery comes to front. Especially in large-value payment systems, huge fixed cost to set up the necessary infrastructural networks play an important role. Due to economies of scale, marginal cost pricing will not recover full costs but other pricing rules can distort allocative efficiency. Moreover, cost recovery is an important factor for competitiveness: no potential competitor would be willing to enter the market if the full costs of production cannot be structurally recovered. In particular, average cost pricing implemented with a single fee would likely discourage users to utilize fully the scale

20 However, in two-sided markets the ‘user’s pay principle’ ceases to hold for the individual market sides, instead one should rather compare total price with total marginal costs (Wright, 2003).
benefits that are available. As these cost recovery and efficiency considerations may be important, also for the assessment of public benefits of Target, this raises the question of how marginal cost pricing should be adapted in order to at least break even on full production costs.

In principle, an obvious departure from marginal cost pricing is to relate the mark-up to the size of the own price elasticity. If the payment market as a whole becomes less elastic, higher total prices can be charged. These higher mark-ups can be used to recover a larger part of total costs, while the quantity demanded is not much reduced. However, empirical determination of elasticities is not easy, since data on prices and costs of payment services are scarce in many countries. Moreover, the departure from marginal cost pricing evidently leads to some cross-subsidisation. That is, elastic payment services are subsidised by inelastic payment services so that overall total costs can be matched by total revenues. However, the need to over-recover in some segments of the payment market opens up the opportunity for ‘cream skimming’. Competitors only operating in those inelastic segments may price at average costs, recouping all their expenses, and at the same time, have an competitive advantage over the suppliers that cross-subsidise. As demand will shift to these low-price competitors, the ability to cross-subsidise is reduced (Humphrey, Keppler and Montes-Negret, 1997).

A more elegant departure from marginal cost pricing is ‘two-part pricing’. A two-part tariff consists of a fixed part and variable part, e.g. a non-linear tariff \( t(q) = F + cq \). The fixed part is used to cover the average fixed cost of producing payment services and reflects the economies of scale in processing them, while the variable part - a transaction-based fee - is used to recoup the average variable costs. The benefits of high volume operations are passed on to those high volume users that generate the economies of scale. Hence, the average price of the payment service decreases with the number of units demanded. In this sense, two-part tariffs correspond to a quantity discount scheme. It also implies that the degree of cross-subsidisation between different types of end-users is limited. As such, a two-part pricing approach would resolve most of the ‘scale’ problems and is indeed a standard pricing approach where fixed costs are large but where all costs may need to be recovered.\(^\text{21}\)

\(^{21}\) Holthausen and Rochet (2002) consider efficient two-part pricing of large-value payments in the presence of unobservable heterogeneity about banks’ future payment volumes.
4.2 Cost allocation and scope economies

Target currently provides both government-related services (‘gov’) and payment services to private banks or financial institutions (‘bk’). Government-related services involve: (1) the maintenance of reserve accounts with the banking system for final settlement of net positions of retail or large value payment transactions made on other payment networks; (2) monetary operations of payment transactions in financial markets; and (3) other government payment functions. Bank-related services involve the processing of large value RTGS payment transactions. These two services are jointly produced as they likely share some of the costs of maintaining accounts, computer processing of payments, and telecommunication facilities between banks and the central bank. If expense data were available, the costs of providing these two services could be broken out into variable costs (VC) easily allocated to either service category (VC_{gov} or VC_{bk}) and those fixed and (some) variable expenses that are jointly incurred (‘joint’) and thus not easily allocated (FC_{joint}).

Even if Target did not process any large value RTGS payments for the banking system, certain site facility costs and payment processing and telecommunication expenses associated with government-related services would still occur. This suggests that some scope economies may exist on Target in providing jointly these government and banking industry services. That is, the sum of the cost of a stand-alone payment network that only provides the government services that Target provides plus the cost of a second stand-alone network that only provides payment services to the banking industry would likely be more expensive than a single network that provides both services jointly.

A standard (but uninformed) method for allocating these joint expenses would be to first determine the share that government-related transaction volume (Vol_{gov}) is of all types of payment transactions at a particular Target processing center: Vol_{gov}/(Vol_{gov} + Vol_{bk}). This share would then be multiplied by the value of the unallocated joint cost, giving the estimated cost (FC_{gov}) allocated to government-related services: FC_{gov} = (Vol_{gov}/(Vol_{gov} + Vol_{bk}))(FC_{joint}). Correspondingly, the portion of jointly incurred fixed expenses and (some) variable expenses allocated to the services provided to the private banking sector (FC_{bk}) would be FC_{bk} = FC_{joint} - FC_{gov}. While this simple transaction volume ratio approach to cost allocation is often used in practice as an accounting rule of thumb to distribute jointly incurred expenses, the private sector has devised and often uses other cost allocation approaches.
When scope economies exist, the allocation of joint production costs need not rely on a simple accounting rule-of-thumb to distribute the joint cost of the two service categories on Target. Indeed, the cost that is allocated to the provision of payment services for the banking industry can be the additional cost incurred for processing, telecommunication, and site operations after the stand-alone cost associated with providing government-related services is determined. All costs would be recovered as the government would cover the costs it incurs (the stand-alone government service cost) and the banking sector would cover the cost it incurs (the additional cost). While this approach needs to be refined with the assistance of technical experts who have a detailed knowledge of the internal cost structure and operations of Target, an indication can still be obtained.

The relative importance of the stand-alone costs of providing only government-related payment services on Target is roughly indicated by how large the share of government-related payment volume is or Vol_{gov}/(Vol_{gov} + Vol_{bk}). This share averages less than 1% across 16 Target processing sites with even lower values at sites with the largest total volumes. Even if the 1% government volume is associated with 10 times its share of processing equipment, telecommunication connections, labor input, and site facility cost - so that 10% of the cost on Target could be allocated to stand-alone government costs, this would still leave 90% of total costs to be recovered from private sector users. Although private sector users of Target would not have to recover around 99% of the total cost (as implied by the above standard accounting rule of thumb), a cost reduction of 10% reflected in the price charged to private users is unlikely to have much of an impact on stimulating demand (especially since these users are not now fully charged for their use of Target). In sum, the scope economy argument for cost allocation could be applied but its impact is likely to be relatively small.

4.3. **Empirical estimates of scope economies on large value payment networks**

We know of no study that has attempted to determine the possible scope economies among the various payment services offered on a large value network. However, there have been a few empirical studies that have attempted to determine scope economies among a broader range of payment and banking services. Importantly, when a proper functional form is used, these investigations have found little support for the existence of significant scope economies either within the commercial banking industry or between different payment services provided by the
Federal Reserve \textsuperscript{22}. This suggests that, aside from general managerial overhead and some likely sharing of infrastructure costs, scope effects are more likely to be small than very large for the central banks that comprise Target. A more definitive answer here will have to wait until the requisite cost accounting data on Target becomes available.

\textsuperscript{22} Little to no scope effects were found between Federal Reserve ACH and wire transfer operations (Adams, Bauer, and Sickles, 2002) and only weak scope effects were identified between U.S. commercial bank deposits and loans (Pulley and Humphrey, 1993) where most had expected to see a strong relationship.
5 CONCLUSION

This paper discusses various theoretic concepts which play a role in assessing the public benefits of Target. These concepts touch upon natural monopoly, network externalities, competition and contestability, as well as economies of scale and scope. The results of an earlier econometric analysis of the effects on unit cost from standardization and consolidation of the Federal Reserve's large value payment network (Fedwire) largely comprise the empirical part of this paper. Based on this analysis, it is argued that if Target decided to standardize its operating platforms and consolidate its processing sites into one or a few centers, it too could realize strong scale economy benefits and lower unit costs.

The concept of a natural monopoly provides a rationale for a temporary partial or full subsidy in order for Target to achieve the 'most efficient scale' and lower unit costs. Such a subsidy can be implemented through temporary 'penetration' pricing (i.e., pricing at less than full current cost). When the potential scale benefits are fully realized, given the size of the market, the subsidy would be removed and full cost pricing implemented. After unit costs have been lowered and reflected in the prices charged, users would face the full costs of their payment decisions. This would permit them to better match benefits with actual costs resulting in a more proper and efficient allocation of resources than occurs today on Target.

Possible scope economies on Target between central bank monetary operations responsibilities and providing payment services for the private sector could not be determined due to a lack of the necessary cost accounting data for these two broad categories of Target services. Information that does exist, however, suggests that scope benefits may be small, as they have been found to be in empirical studies of scope economies among payment services operated by the Federal Reserve.
BIBLIOGRAPHY


Figure 1: Fedwire Average Total Cost, 1979:1 to 1997:3
(75 quarterly observations)
Previous DNB Working Papers in 2005

No. 27 Jan Marc Berk and Beata K. Bierut, On the optimality of decisions made by hup-and-spokes monetary policy committees
No. 28 Ard H.J. den Reijer, Forecasting dutch GDP using large scale factor models
No. 29 Olivier Pierrard, Capital Market Frictions, Business Cycle and Monetary Transmission
No. 30 Jan Willem van den End and Mostafa Tabbae, Measuring financial stability: applying the MfRisk model to the Netherlands
No. 31 Carin van der Cruysen and Maria Demertzis, The Impact of Central Bank Transparency on Inflation Expectations
No. 32 Allard Bruinshoofd, Bertrand Candelon and Katharina Raabe, Banking Sector Strength and the Transmission of Currency Crises
No. 33 David-Jan Jansen and Jakob de Haan, Were verbal efforts to support the euro effective? A high-frequency analysis of ECB statements
No. 34 Riemer P. Faber and Ad C.J. Stokman, Price convergence in Europe from a macro perspective: Product categories and reliability
No. 35 Jan Kakes and Cees Ullersma, Financial acceleration of booms and busts
Financial acceleration of booms and busts