CHAPTER 19

The economics of financial market infrastructures

Updated on 17 December 2018

In the economic sense, "infrastructures" generally refers to institutions used to exchange goods, information or rights between agents. As the previous chapters have shown, financial market infrastructures (hereinafter FMI) specifically serving the functioning of financial markets are the subject of particular attention by legislators and regulators because they generally have characteristics that influence how smoothly those markets function. FMIs' economic characteristics make them an archetype for market failure situations, in which market price formation mechanisms alone are unable to fully capture the costs and benefits of the activities under consideration and to lead to an optimal allocation of resources.

Economic analysis of FMIs' functioning can largely be done using the tools developed in the field of network economics (particularly communication). According to Nicholas Economides,[1] networks can be formally defined as a set of "links" connecting "nodes", whose different components function in a complementary manner.[2] This definition can be applied to FMIs, all of which are designed to establish links between participants (nodes), via their connection to the same technical system, and thereby provide the desired service. A payment system, for example, connects participants, and the complementarity between the different components comes from the fact that participants need to be connected to the system for the payment service to be provided. Generally speaking, networks can be understood through two different prisms:[3] a technical prism, where they consist in an interconnection of components that 'cooperate' in order to transport flows such as assets or information, and an economic prism, where essentially they act as a physical medium for economic intermediation (transactional view).

This chapter describes the impact of FMIs' economic characteristics (I) on market organisation and dynamics (II) and investigates pricing-related issues (III).

1. The economic nature of FMIs

The economic nature of FMIs is notable for its significant externalities – in particular network externalities, as well as a cost structure that enables economies of scale and scope.

1.1. The presence of major, particularly network, externalities

The externality concept refers to situations where one agent's production or consumption affects another's usefulness, externally to the market. If the effect is positive,[4] the externality is said to be positive; otherwise, it is said to be negative.[5] This phenomenon implies that the price mechanisms at play do not make it possible to satisfactorily capture all the costs and/or benefits associated with the production or consumption in question. This discordance between social costs and private costs leads in most cases to a divergence between social and individual optimal levels of consumption or production and complicates measurement of social benefits generated by the activities concerned.

Financial market infrastructures have two main categories of externalities: either negative or positive externalities linked to their potential impact on financial stability, or so-called network (or 'club') externalities.

Positive financial stability-related externalities pertain to the contribution that FMIs can make to financial stability due to their systemic nature,[6] which justifies the intervention of public authorities to impose their use in certain markets. Take EMIR, for example.[7] This regulation imposes, among other things, an obligation to report derivative transactions to trade repositories. This is with good reason, because the reporting of transactions in a given market to a trade repository enhances that market's transparency, which makes it possible to more appropriately assess the risks that its activity may pose to financial

1 Economides N., "The economics of networks", *International journal of industrial organization*, 1996.

2 This implies that all the components are necessary for the provision of a service.

3 Curien N., *Economie des réseaux*, La Découverte, 2005.

4 A classic example is when a beekeeper and a farmer operate adjacently and the beekeeper's bees pollinate the farmer's fields at no cost to the farmer.

5 An example of this is when pollution by industrial activities generates disease and economic costs for nearby inhabitants.

6 This systemic nature, which is described in more detail in Chapter 18, results in the efforts that FMIs undertake individually benefiting the financial system as a whole.

7 For further details, see Chapter 16 (trade repositories) and Chapter 11 (CCPs).

stability and enables public authorities to promptly contain those risks for the benefit of all participants. Also in order to ensure financial stability, EMIR imposes, via CCPs, a multilateral clearing obligation on counterparties to certain categories of OTC derivative transaction[8] (the main role of CCPs being to provide their participants with a risk management mechanism).

Negative financial stability-related externalities, by contrast, pertain to the potential systemic risks that financial market infrastructures can pose to financial stability due to their systemic nature.[9] This explains why FMIs are subject to strict regulation aimed at ensuring that their operators take into account the risks they can generate for the ecosystem as a whole. This is reflected, for example, in the PFMIs' provisions on governance, which stipulate that *"An FMI should have governance arrangements that are clear and transparent, promote the safety and efficiency of the FMI, and support the stability of the broader financial system, other relevant public interest considerations, and the objectives of relevant stakeholders."*

With regard to FMIs, on the one hand all ecosystem players benefit equally from a stable and resilient financial system, without the fact of one of them benefiting from it preventing the others from doing the same, and on the other hand it is impossible to exclude any player whatsoever from the system. Consequently, FMI-generated financial stability-related externalities fall under the specific economic category of public goods, which introduces the risk of possible free-rider behaviour[10] by market players that can, at the level of the industry as a whole, lead to a sub-optimal level of financial risk management. This problem is accentuated by the risks of moral hazard[11] induced by participants' confidence that the public authorities will, whatever the circumstances, rescue entities of systemic importance (i.e., because they are too big, or too interconnected, to fail).

Network externalities arise from the fact that an individual participant's utility is positively correlated to the (current and future) number of a network's participants:[12] a user's decision to join a network thus corresponds to another type of positive externality. This effect can be both direct and indirect;[13] direct in that the connection of new users directly affects the utility of others by allowing them to be connected to a larger number of counterparties, and indirect in that an increase in the number of network users leads to an improvement in the characteristics of the network's offering (i.e. in the quality of its services) or the supply of complementary goods and services (greater variety), which makes the network more attractive to service providers.[14]

As a result of these network effects, adoption decisions by users are linked to the FMI's current and expected number of users. However, users naturally do not decide whether or not to join a network based on that decision's potential impact on other users' usefulness, and that is where the externality lies. For FMIs, therefore, there is a risk that the network's size at equilibrium may be smaller than its socially optimal size.

In the case of stock exchanges, for example, from the investors' point of view network externalities originate in the search for the trading platform offering the greatest liquidity for a given type of security. This leads directly to even greater liquidity for the type of security in question and therefore an increase in the chosen platform's attractiveness for other investors.[15] From an issuer's point of view, they stem from the fact that the bigger the investor network, the more easily it will absorb the issuer's capital requirements.[16]

At the clearing and settlement stage, network externalities come mainly from transaction processing time savings, while the establishment of netting mechanisms reduces the opportunity costs associated with regulatory capital requirements.

8 See Chapter 11 (CCPs) for further details.

9 For further details, see Chapter 17.

10 While it may be in everyone's interest to have a more stable financial system, certain participants may make others bear the cost of making it more secure.

11 The possibility that a party insured against a risk behaves differently than if he or she were fully exposed to it.

12 In economic terms, therefore, the usefulness of the various parties is interdependent and the parties' consumption of the network's goods or services complementary.

13 Katz M., Shapiro C., "Network Externalities, Competition and Compatibility", *American Economic Review*, 1985; Katz M., Shapiro C., "Systems Competition and Network Effects", *The Journal of Economic Perspectives*, 1994. See also footnote 509.

14 Familiar phenomenon in the field of programming with the hardware/software paradigm, for example, where an increase in demand for hardware can lead to an increase in the development or quality of compatible software (and vice versa). As regards credit cards, meanwhile, it reflects the fact that the more credit card users there are, the more merchants with a terminal enabling the use of credit cards there will be (complementarity), which in turn increases the credit cards' usefulness for their owners (see Chapter 4).

15 For a review of literature on liquidity as an externality, see: Serval T., "Lorsque les réseaux d'information deviendront des bourses", *Presses de Sciences Po*, 2001.

16 See footnote 3.

1.2. A cost structure favouring economies of scale and scope

1.2.1 FMIs generate economies of scale

One of FMIs' chief roles is to enable financial players to outsource to a single point processing that was previously done in different locations. By doing this, they generate economies of scale.

This is because the fixed costs to set up an FMI are considerable, insofar as they are largely independent of transaction volumes. The costs in question are those associated with physical infrastructures (servers and server buildings), IT developments (program-writing, tests, error correction, etc.) and the implementation of a governance system (legal documentation, etc.). To some extent, they may also include costs to maintain and monitor the system and train staff – although these items are not entirely independent of the number of transactions. Lastly, users also incur fixed costs when they first connect to the infrastructure. However, the variable costs related to the operation of the networks are generally quite low.

The high fixed costs and low variable costs mean that average costs per user generally decrease as the number of users increases, resulting in economies of scale, or increasing returns, which are also linked to learning effects[17] (increased process mastery).

Empirical studies on the subject have found the economies of scale in payment systems (retail or wholesale) to be significant, with fixed costs as a share of total costs ranging from 50% to 80%.[18] The possibility of economies of scale was also an important argument put forward when TARGET2 was set up.[19]

Stock exchanges also offer economies of scale at the trading stage linked to the combination of high trading venue set-up fixed costs and low incremental costs per transaction thereafter. Thus, at comparable fixed costs, the platform with the largest volume of transactions will be able to benefit from the lowest average transaction costs.

Lastly, these economies of scale can also be seen at the clearing and settlement and delivery stages. The marginal cost of clearing, for example, is close to zero, while through its activity a clearing house makes savings in terms of collecting and analysing information about its members. Indeed, monitoring credit and liquidity risks involves the implementation of sophisticated risk monitoring techniques, which requires costly investments (IT, modelling, organisational, etc.). It is more rational to pool such investments in a single entity – the CCP. In addition, a CCP's central position gives it a greater capacity to monitor and manage these risks than that deployable by individual players acting alone. Insofar as margin calls are made on the basis of participants' net positions, this system enables them to save liquidity. It also reduces settlement and delivery volumes as a proportion of cleared transaction volumes, especially as the larger the volume of transactions on a given instrument, the greater the likelihood that some of these transactions will be in the opposite direction.

1.2.2. FMIs generate economies of scope

Moreover, the FMIs can generate economies of scope, i.e. situations where it is more cost-effective to jointly produce several types of goods or services through a single firm than it is to use separate operators. It should be noted, however, that the realisation of such savings may have an impact on other forms of efficiency – in particular systemic; see Chapter 17 on FMI-related risks.

With regard to central banks' operation of large-value settlement systems, it has been suggested that this offers economies of scope thanks to i) their management, on behalf of credit institutions, of a system of accounts for reserves and interbank settlements and ii) their operation of the system that makes settlements between these accounts.[20] In addition, these systems enable central banks to provide services to governments[21] and private

17 These effects correspond to the increasing return from the labour factor through the repetition of certain tasks over time.

18 Khiaonarong T., "Payment systems efficiency, policy approaches, and the role of the central bank", *Bank of Finland Discussion Papers*, 2003.

19 Bolt W. and Beijnen C., "Size matters: economies of scale in European processing", *DNB Working Paper*, 2007; Bolt W. and Humphrey D., "Payment Network Scale Economies, SEPA, and Cash Replacement", *Review of Network Economics*, 2007; Bolt W. and Humphrey D., "Public good issues in TARGET", *ECB Working Paper Series*, 2005.

20 Millard S., Haldane A. and Saporta V., *The Future of Payment Systems*, Routledge, 2008; or Green E.J. and Todd R.M., "Thoughts on the Fed's Role in the Payments System", *Federal Reserve Bank of Minneapolis Quarterly Review*, 2001; or Green E. J. "The Role of the Central Bank in Payment Systems", 2005.

21 Reflecting the historical role of central banks as an intermediary between governments and their lenders.

banking institutions, thus enabling them to benefit from synergies linked to the joint provision of these services.[22]

With regard to securities trading venues, economies of scope can result from the possibility of trading in several types of securities on the same platform. Once the infrastructure has been set up, the extension of the trading services to an additional type of security can be offered at a modest incremental cost, especially if a network of active buyers and sellers is already in place. It is also economically optimal for users to group their activities on different types of securities onto as few trading platforms as possible – notably to achieve back-office savings.

22 Bolt W. and Humphrey D., "Public good issues in TARGET", *ECB Working Paper Series*, 2005.

Box 1: Economies of scale and scope expected from T2S' implementation

Implementation of the T2S project provides a good example of the cost savings that can be achieved by consolidating activities in the area of settlement and delivery (for further details on how T2S works, see Chapter 14).

The gains expected from T2S in terms of exploiting economies of scale and scope and network externalities were detailed in a study carried out in 2007 by the ECB.[1] The provision of this common platform has generated economies of scale through the consolidation of both investment and operating costs linked to the maintenance of platforms (specific, prior to migration, to each of the euro area countries' 24 CSDs).

With regard to the CSDs' users, T2S' implementation enables them to centralise their securities accounts with a single central depository of their choice (provided that they have established the necessary relationships) or their cash account at the level of a single central bank, thereby reducing account dispersion-related costs. This should enable them to benefit in particular from opportunities to streamline their cross-border activities and make back-office savings, notably by having a guaranteed single entry point to several markets, harmonising their internal procedures, using their guarantees jointly for their cash and securities activities (creation of a single collateral pool that reduces cross-border guarantee mobilisation costs) and having to rely on fewer intermediaries (CSDs, but also asset management-related). The use of a single platform and standardised communication protocols allows issuers to reach more investors and so increase investor demand, while enabling investors to reduce the cost of managing an international securities portfolio – and thus increase the return on that portfolio.

By opening up domestic markets, the project should increase competition between CSDs in the single market and lead to systems consolidation in Europe, which could in turn generate additional economies of scale and scope. Still in the medium term, the benefits linked to the use of a single technical platform will also materialise when it is upgraded or, if necessary, overhauled, when there will then be only one project to manage instead of 23.

Lastly, the current work to consolidate TARGET2 and T2S aims to exploit synergies between the two platforms[2] (examples of economies of scope) in the areas of the use of IT resources and architecture, possibilities for reuse of existing communications technologies, and organisation of support and operational functions. The exploitation of these synergies could even result in the future merger of the two platforms (see Chapter 7, Section 6.3).

1 ECB, 2007, TARGET2-Securities – Economic Feasibility.

2 As detailed in this document, for example: *https://www.ecb.europa.eu/pub/pdf/other/t2seconomicfeasibility0703en.pdf?8e36385d37d399eaf9a3615292b80c08*

At the post-market stage, clearing several categories of financial instruments makes it possible, subject to the framework for managing risks – which remain specific to each category regardless – to pool certain resources (risk or legal teams for example, or technological infrastructures, which represent a significant cost centre). Clearing a wide range of financial instruments can also, thanks to a lower correlation between the various instruments' risk factors, enable the CCP to reduce exposure variance and, consequently, the amount of the collateral it sets aside against these risks. This argument also applies for participants who choose to have their transactions cleared through a single rather than multiple CCPs: since the risk associated with each participant's portfolio is smaller than the sum of the risks per instrument, a single CCP can accept a margin calculated on the basis of a lower net exposure (portfolio management models). In Europe, for example, LCH Ltd recently launched the Spider offer, which makes it possible to jointly clear listed interest rate futures and OTC interest rate swaps by calculating a net margin on all positions. This allows LCH Ltd to grow in the listed market by taking advantage of its strong position in interest rate swaps; conversely, Eurex has launched an initiative to enter the swap market based on its position in futures.

Finally, economies of scope can also come from a single group's provision of securities trading, clearing and settlement services, which are highly complementary.[23] This enables straight-through transaction processing within the same group, leading to reduced communication costs between the various activities, can promote the implementation of common standards for data transmission between the various stages, and improves the process of disseminating innovations along the length of the chain by reducing coordination needs. The establishment of a vertical 'silo' of FMIs covering the whole securities processing chain, from trading and clearing to settlement and delivery, shows that this type of saving is possible. However, it poses challenges in terms of managing risks (particularly systemic), as described in Chapter 17 (Section 2.2.1), and competition.

2. The impact of FMIs' economic characteristics on market organisation and dynamics

The existence of economies of scale on the supply side and network externalities on the demand side can hamper competition in network industries and encourage operator consolidation, on the one hand due to a tendency towards horizontal and vertical concentration, and on the other hand due to the fact that these two characteristics reinforce each other and create feedback effects. Such a concentration trend raises key financial stability issues. For example, concentration of a clearing or settlement and delivery activity on a very small number of players, or even on a single player by category of activity, results in the creation of systemic players whose failure would have extremely destabilising consequences; such players would be considered 'too big to fail'. The creation of such monopolistic players leads to a significant moral hazard, insofar as it requires the authorities to intervene to prevent their failure (see Chapter 17 on FMI-related risks). Public authorities therefore face an industrial and competition policy challenge relating to FMIs, to encourage the most efficient market organisation possible by avoiding the creation of excessively systemic, monopolistic players.

2.1. A tendency towards horizontal and vertical concentration

As a result of the network externalities and economies of scale and scope that their activities generate, FMIs are natural monopolies in the sense that it can be optimal for a single entity to be responsible for meeting all market demand.

23 In the sense that demand for each service is not independent of the other services' price.

Box 2: T2S and the delimitation of 'pure' infrastructure provision activities

As described in Chapter 14, T2S is not considered as a securities settlement system or CSD, but DvP rather a technical platform providing harmonised IT infrastructures that enables CSDs to develop their services on an identical basis. The establishment of T2S can be likened to choices made in other network industries (telecommunications, rail, electricity, etc.) historically organised as monopolies to break down these monopolies into the supply of 'pure' network infrastructure provision activities, in which service by a single player is the most efficient form of market organisation. By contrast, the part of the network corresponding to commercial service provision activities based on this infrastructure would be open to competition, due in particular to its lower fixed costs.

In the case of T2S, the economic rationale for separating the infrastructure layer from the service layer is similar, but the historical direction is the opposite, with the public sector taking over a pure infrastructure provision activity (the technical platform) for efficiency reasons, while the CSDs remain in control of the commercial services provided to their clients (the 'service' layer).

FMIs' fundamental economic characteristics therefore constitute a barrier to entry for newcomers, insofar as they must be able to replicate the significant fixed costs of existing operators, most of whose fixed costs are generally very specific and thus sunk in the event of an exit from the market, and who as pre-established players will always benefit from lower average costs per user. This situation increases the importance for a new operator of being able to rapidly reach a critical mass of clients that allows it to exploit economies of scale and network effects; uncertainty about a potential entrant's ability to reach such a critical mass can also hinder its entry into the market. These characteristics show the productive efficiency of the FMI concentration trend, both horizontally and vertically.

In practice, the large-value payments segment therefore often operates as a duopoly at the domestic level, as we have seen in Chapters 6 to 8 (TARGET2 and EURO1 in Europe, FEDwire and CHIPS in the United States, etc.). Similarly, the retail payments market (see Chapter 10) is often dominated at the domestic level by a single player (for example CORE(FR) in France, operated by STET). European domestic financial markets are also generally organised around a single national CSD and a single CCP, although T2S' recent commissioning could lead to changes in this set-up.

Securities' post-market landscape is logically more integrated in the United States than in Europe, the United States being a single domestic market in this respect. In the US, settlement and delivery and clearing are organised mainly around the Depository Trust and Clearing Corporation (DTCC),[24] which also acts as a central depository, while in Europe the compartmentalisation of domestic systems (for historical, technical, institutional, legal or tax reasons) leads to relatively high costs for cross-border transactions[25] (due to the complexity and number of intermediaries potentially involved in these transactions).

This fragmentation prevents networks' positive externalities and possible economies of scale from being fully exploited. However, in recent years there has been a consolidation movement in

24 Created in 1999 to consolidate and integrate the operations of the Depository Trust Company (DTC) and the National Securities Clearing Corporation (NSCC).

Europe, both horizontally (concentration of entities providing similar services at the same point in the post-market value chain) and vertically (entities providing different services that integrate the whole post-market value chain offering).

Horizontal concentration movements have notably been observed in mergers between trading venues: creation of the Euronext group in 2000 through the merger of the Amsterdam, Brussels and Paris stock exchanges, followed by its merger with the Lisbon and Porto stock exchanges in 2002, and with the Dublin stock exchange in 2017. They have also been observed at the clearing level (consolidation of LCH and Clearnet into a single group in 2003), and at the custody and settlement and delivery level, with in particular Euroclear's strategy of merging with several national CSDs and the establishment of the Euroclear Settlement of Euronext-zone Securities (ESES) platform in 2009 for all transactions carried out on the Euronext markets.[26] The trio formed between Euronext for trading, the LCH SA central counterparty (controlled since 2013 by the London Stock Exchange) for clearing and the Euroclear central depository illustrates, at the level of each part of the securities processing value chain, the horizontal integration model. In these examples, FMIs' with a specific business focus (trading, clearing, and settlement and delivery) seek to serve several geographic markets. The horizontal model can benefit participants and other users in that it makes it possible to generate strong commercial synergies: openness and positioning in different segments of the infrastructure's core business; diversification possibilities; innovation; and, because the infrastructure's activity is not reliant exclusively on a single market, greater independence.

Vertical silo integration is also seen in other markets. Under this organisation, the entire securities transaction chain (trading venue, clearing house and central securities depository managing a settlement and delivery settlement system) is handled on a straight-through basis by a single-capital group's infrastructures and proprietary systems. Such is the case with the German model, for example (where the Deutsche Börse Group controls the Eurex platform, the Eurex Clearing CCP and the Clearstream Banking Frankfurt CSD) and the Italian model (with the Borsa Italiana Group, in which the company Borsa Italiana – itself controlled by the LSE – controls the MTS trading venue, the Cassa di Compensazione e Garanzia clearing house and the Monte Titoli national CSD). Gains are made possible through an alignment of strategic interests that can result in increased innovation (reduction in coordination costs or insourcing of R&D-related externalities) or integration of tools and processes enabling better coordination and dissemination of learning benefits. However, this model could lead to quasi-exclusive relations between these players, which would raise competitiveness risks, currently an area of keen regulatory focus (see below). Furthermore, such vertical integration of different infrastructures with different risk profiles and regulatory constraints can pose a problem when it comes to preserving each infrastructure's necessary independence.

2.2. Consequences weighing on innovation dynamics

Because of the network externalities and economies of scale and scope described above, supply and demand generally interact in network industries on a so-called feedback basis:[27] the more a network service is distributed, the more increasing adoption yields will fuel growth in demand and thus increase its dissemination at the expense of competing services (an increase in supply leads to an increase in demand, which leads to an increase in supply, and so on), and vice versa.

These effects firstly strengthen the trend towards a natural monopoly, leading to market polarisation around a limited number of network operators and service providers, in which firms with a significant base strengthen their position. This makes

25 The cost differential between a domestic transaction and a cross-border transaction was estimated in 2011 as ranging from 1 to 10 – "Settling Without Borders", European Central Bank, 2011.

26 With the exception of Portugal.

27 See footnote 511. Also: Foray D. "Innovation et concurrence dans les industries de réseau", *Revue française de gestion*; Katz M. and Shapiro C., "Technology adoption in the presence of network externalities", *Journal of Political Economy*, 1986; Shapiro C. and Varian R., "Information Rules", *Harvard Business School Press*, 1999.

them a 'growth amplifier' and creates a 'winner takes all' logic where even a small advantage can shift the market in favour of its beneficiary, while it will be very difficult for dominated firms to survive (also known as a 'market tipping' phenomenon).

A classic consequence of such effects is that the dominant technology will not necessarily be the most efficient: efficient technology can be excluded because it takes too long for it to reach the market, and a standard may come to dominance due to early selection (the 'first mover advantage') rather than because it offers the best quality.

Secondly, feedback effects result in innovation dynamics being beset by excessive inertia linked to coordination failures.[28] This results from both the potential lack of incentives for users to adopt new technologies, and a lack of incentives for suppliers to invest in new technologies.

Adoption inertia is primarily attributable to the fact that network technologies are generally characterised by significant migration costs. This can make consumers captive and be an additional barrier to entry for alternative suppliers or technologies. Migration costs can be defined[29] as the sum of the cost borne by the consumer (learning and specific investments) to change technology and the cost borne by the new supplier to replicate the consumer's position with its previous supplier. Thus, in the case of FMIs' underlying information technologies, the participant has generally had to make long-term investments in additional resources[30] specific to the technology in question. A change in the main technology therefore necessitates replacing or at least adapting these additional resources, which may also have a different lifespan, meaning that it will never make sense to start over entirely, which reinforces the inertia effect. An example of migration costs as regards FMIs are the investments that CSDs had to make in order to be able to use T2S, including in particular architectural changes in the systems they operate (for example, modification of the IT architecture of the ESES platform for Euroclear). These migration costs support the argument that technological inertia can sometimes be economically efficient, even if the replacement technology is better. This is particularly the case if the migration cost is higher than the benefits generated by the new technology.

Adoption inertia can also arise from the fact that, even though it may be in all participants' interests to adopt a new technology, the risk of finding themselves isolated in a new network may result in each of them deciding to postpone their decision to switch until the new network's user base is big enough. If all users follow the same logic, the change stalls. This impasse can also be seen as a form of prisoner's dilemma, the result of which would be more favourable if the actors cooperated by agreeing on the decision for a technological change, but where in the absence of cooperation, and in a situation of uncertainty about the choices that others will make, the best course is for each participant not to change technology (sub-optimal Nash equilibrium scenario).[31]

When the market is left unfettered, therefore, its dynamics do not always lead to economically efficient results – hence government intervention in the markets and, particularly as regards FMIs, the catalyst role that central banks can play.

2.3. Public authorities' responses to competitiveness issues

2.3.1. FMIs and competition policy: the issue of access to FMIs

In addition to the link between FMIs' size and the systemic risks they can pose to the financial system, specifically dealt with in Chapter 17 on FMI-related risks, the sector's natural concentration tendency also poses a problem due to the pricing practices that operators can implement – at the expense of consumers and economic exchange – thanks to their market power.[32] As such, European competition policy[33] strives to

28 Katz M. and Shapiro C., "Systems Competition and Network Effects", *Journal of Economic Perspectives*, 1994; Farrell J. and Saloner G, "Standardization, compatibility and innovation", *Rand Journal of Economics*, 1985; Milne A., "What is in it for us? Network effects and bank payment innovation", *Journal of Banking and Finance,* 2005.

29 Shapiro and Varian (1999).

30 In the sense that they are necessary for access to the service provided by the infrastructure.

31 In game theory, this refers to a situation in which each player adopts the best strategy possible given the strategy adopted by the other player.

32 See theoretical economic literature on the deadweight loss of the monopoly.

33 This refers to all mechanisms, in particular legal, aimed at organising and controlling markets in such a way as to encourage the maintenance of sufficiently competitive functioning for maximum economic efficiency.

combat the possibility that operators use their market power to carry out restrictive competition practices aimed at creating or strengthening barriers to entry.

In their role as catalyst, the European authorities have thus chosen to build the single market on a model based on competition between the various providers of financial services, be that between the institutions themselves or between FMIs, particularly at the cross-border level. This commitment has been reflected in particular in efforts to remove the barriers identified in the Giovannini reports, MiFID,[34] of which this approach was the mainstay, and the development of the post-trading code of conduct carried out under the aegis of the European Commission. It should be stressed that this approach does not contradict the recognition of concentration's potential for productive efficiency, insofar as the objective pursued is to put in place a single European financial market: the de-fragmentation of the various geographical markets makes it possible to introduce competition between players (by moving markets' geographical borders), and concentration movements can, in the longer term, be a consequence of this competition (see Box 2 on T2S).

Legislators and regulators pay particular attention to the conditions of access to FMIs, as such access may be a prerequisite for participating in certain markets. This makes FMIs an essential infrastructure[35] that operators can use, for markets with an essential need of infrastructure access, as a bottleneck[36] facility. Certain FMIs' essential infrastructure status can also be based on regulatory requirements: the Dodd-Frank Act in the United States (Title VII) and Europe's EMIR, for example, impose a clearing obligation on counterparties to certain derivative transactions. It is therefore important for market participants subject to this obligation to have access to the infrastructures that offer these services. Essential infrastructure theory advocates the access obligations imposed under sectoral regulations, which for FMIs are contained in PFMI principle 18, which states that *"An FMI should have objective, risk-based, and publicly disclosed criteria for participation, which permit fair and open access"* in order to promote competition between market participants.

Furthermore, operators' market power can allow them to apply a leverage strategy[37] – or even direct crowding-out strategies[38] – in upstream, downstream or ancillary markets to the ones that they serve – these risks being particularly significant in vertically integrated infrastructures. For example, CCPs are generally owned by operators that also have post-trading activities (e.g. the Deutsche Börse Group), and the choice of CCP is often made by the trading venue's operator, which can decide to use a single CCP and thereby exclude the others.

The principle of fair and open access as regards relations between market infrastructures is tackled in various regulations (EMIR,[39] CSDR, MiFID II/MiFIR) that have been introduced since the financial crisis.

For example, the open access principle allows a trading venue to have its transactions cleared by the clearing house of its choice, and a clearing house to have access to the transaction flows of any trading venue. If an operator on the downstream market is denied access to a clearing house or has less favourable access conditions imposed than other infrastructure members, its costs may rise or it may even be squeezed out of the market in question. At the same time, in the absence of regulations there is a risk that a trading venue may refuse a CCP the right to clear transactions traded on its platform, which would cut off transaction flows to the CCP. Sectoral regulations advocate the open access principle to mitigate this type of risk and encourage the opening of vertical silos. In the other direction, this principle allows a CSD to have access to transactions cleared by any clearing house and a clearing house to have access to the transaction flows of any trading venue.

34 Directive 2004/39/EC of the European Parliament and of the Council of 21 April 2004 on markets in financial instruments.

35 Originally the creation of the US courts in their application of the Sherman Act's provisions prohibiting monopolies, this theory was incorporated into Community and French law as part of the fight against the abuse of a dominant position and the opening up to competition of markets previously run by public monopolies, before being expanded. For more information, see Supreme Court of the United States, United States v. Terminal Railroad Association, 224 US 383, 1912; 2005 annual report of the *Cour de Cassation* highlighting a ruling of 12 July 2005 of the French Chamber of Commerce (no. 04-12388); Chang F.B., "Financial Market Bottlenecks and the "Openness" Mandate", *University of Cincinnati*, 2015; Sealink affair, European Commission, 94/19/EC, 1993, where the Commission explicitly uses the expression "essential facility"; and, lastly, CJEU ruling nos. C-241/91, C-07/97 and C-418-01.

36 See footnote 509.

37 A company's ability to increase its sales in a market – the 'linked market' – by exploiting its dominant position in an adjacent market, making it more difficult to access these markets.

38 The effectiveness of such strategies has been a subject of debate in economic literature, and in particular disputed by the University of Chicago; for a presentation on how the control of an essential infrastructure can make it possible to increase competitors' costs, see for example Patrick Rey and Jean Tirole's *Handbook of Industrial Organization*.

39 Articles 7 and 8.

Box 3: Case law examples of FMI access issues

In certain cases FMIs have had to revise their access policy to comply with legal requirements. For example, it was only after such access rules were enacted that LCH Clearnet's SwapClear platform removed a clause requiring its members to maintain a $1,000bn interest rate swap portfolio balance, seen as a potential means of reserving the market for major institutional sellers.[1]

In Europe, in a decision of 2 June 2004[2] (confirmed on 9 September 2009 by a decision of the CJEU[3]) the Commission ruled that by refusing to provide cross-border clearing services (within the meaning of establishing reciprocal contractual obligations between buyer and seller) and settlement services for registered shares issued under German law, and by adopting discriminatory pricing practices, at the expense of Euroclear Bank, Clearstream Banking AG and its parent company Clearstream International SA had abused their dominant position in the markets concerned. In this case, the Commission emphasised the fact that, including for the settlement and delivery market, while competition law recognised the "freedom of companies to choose their trading partners", companies in a dominant position had a "special responsibility". It considered that, in this case, Clearstream was the only depository of German securities held in collective custody and that a new market entry was not a realistic assumption, which made it an unavoidable trading partner, that Euroclear Bank could not duplicate the services that it was requesting and that Clearstream's behaviour had the effect of impairing Euroclear Bank's ability to provide cross-border clearing and settlement services to clients in the single market between 1997 and 2002.

1 Felix B. Chang, "Financial Market Bottlenecks and the "Openness" Mandate", *University of Cincinnati*, 2015. Core Principle C (v) of the Dodd Frank Act explicitly prohibits derivatives clearing institutions from requiring their members to maintain a particular value or volume of outstanding swap positions.

2 Press releases: *http://europa.eu/rapid/press-release_IP-04-705_fr.htm*.

3 *http://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:62004TJ0301&from=EN*

2.3.2. FMI interconnection and compatibility strategies

Interconnection, or interoperability, means the connectivity between the various networks that enables the users of one network to communicate with those of another or to access services provided by different operators.[40] This compatibility makes it possible to realise the positive externalities resulting from the complementarities between the network components.[41] This possibility of 'dialogue' between systems may be imposed by the regulations or reflect a choice on the part of the service providers.

As regards FMIs, an example of a lack of interoperability in Europe was the situation prior to the migration of T2S, characterised by fragmentation of the settlement and delivery market between the EU's various Member States. Concerning work to implement the SEPA for means of payment, according to the Eurosystem's interpretation of the SEPA regulation the interoperability concept implied that once payment service providers participated in a given retail payment system they would be able to reach all counterparties in the SEPA, whether or not they belonged to a different system. This required putting in place interoperability links.

Interoperability is based in particular on the implementation of compatible technical, operational and legal standards (e.g. messaging system flows or compatibility between hardware, software and operating systems). The use of open, transparent and non-proprietary standards facilitates interoperability and thereby stimulates competition in the market by opening up the possibility of transactions between users of the various infrastructures without users

40 One definition of this concept, for example, appears in Article L.32 of the French Postal and Telecommunications Code.

41 See footnote 509: "Links on a network are potentially complementary, but it is compatibility that makes complementarity actual."

having to belong to all of them. It facilitates the simultaneous activity of several infrastructures, making it possible to avoid the excessive concentration phenomenon described above and thereby potentially leading to an improvement in the quality of the services offered to consumers and reducing the risks that a player in a dominant position abuses that position. Common standards can make it possible to avoid coordination problems in firms' technological choices and anchor users' expectations about the chances of a given technology's adoption. This reduction in consumer uncertainty itself makes it possible to mitigate their risk of being locked in to technologies that end up not being accepted by the market as a whole, and therefore to limit wait-and-see, non-adoptive behaviour. When an industry agrees on a standard it generally leads to an increase in the pace of adoption and evens out competition.

Standardisation[42] does however pose the risk of being in thrall to bad technologies, in particular due to the feedback mechanisms described above, and may increase the costs of transition to new technologies. Furthermore, its costs can vary depending on the market players concerned – i.e. how big they are or whether they are already in place and have incurred sunk development and compliance costs relating to a set of standards that may be more or less different from that targeted by the process. Finally, the requirement of full compatibility poses a risk of encouraging free-rider behaviour at the R&D and innovation stages, ultimately leading to reduced incentives to invest in the network's operational improvement.

From an individual operator's viewpoint,[43] the choice to use the same standards as its competitors may be driven by the prospect of being able to access their clients, in which case, thanks to the network effect described above, it could increase the usefulness for these clients of access to their network and/or services (and therefore, notably, their willingness to pay) and accelerate their ability to reach a critical mass of consumers. In principle, the larger an operator's existing client base, the less useful having an interconnection strategy will be, because its existing participants and potential new entrants will be in divergent situations.[44] Conversely, an operator may refuse to interconnect on economic grounds, in that it wishes to offer differentiated services from its competitors (the competition effect, reflecting the fact that compatibility between services offered by two suppliers makes those services more substitutable) or sometimes even squeeze its rivals out of the market. Such a strategy could, for example, be adopted by a new entrant with better technology than the existing players that seeks to impose it based on a rapid uptake effect (positive feedback spiral). In this sense, the establishment of standards can constitute a renewed form of monopoly enabling players that win the 'standards wars' to enjoy an economic premium. For this reason, FMI-related legislation, notably MiFID, MiFIR and EMIR, requires FMIs to interconnect if another FMI so requests. Risk considerations are the only acceptable reason for an FMI to refuse interconnection under these regulations – if it can prove that such interconnection would pose a risk to its security and operational efficiency, for example.

This leads to the question of whether standards' compatibility is a natural market equilibrium – i.e. whether or not the participants have a vested interest in promoting it to maximise the externalities that can be produced by expanding the network. Without sufficient incentives, this is not necessarily the case.

The remedy for this market failure and the impetus for common standards for the industry's full interoperability can come from action by the public authorities, international bodies or committees from the industry itself, or even from individual market participants' initiatives. As regards securities settlement and delivery, for example, the 2003 Giovannini report conferred upon SWIFT and the Securities Market Practice Group[45] a facilitating role for a project to harmonise messaging and interoperability

42 See Milne A., "Standards setting and competition in securities settlement", *Bank of Finland Research Discussion Papers*, 2005. The author stresses the fact that proprietary standards could encourage more innovation and be promoted more aggressively by the companies that develop them, and therefore stand a greater chance of wide acceptance in the industry, and that, moreover, companies using incompatible standards could compete more intensely on prices.

43 See footnote 3, as well as Kemppainen K. and Salo S., "Promoting Integration of European Retail Payment Systems: Role of competition, Cooperation and Regulation", Bank of Finland, 2006.

44 See footnote 3.

45 Working group of practitioners aimed at promoting the harmonisation of securities industry market practices. It relies on the National Market Practice Group, which operates in more than 30 countries.

standards at industry level, in order to remove one of the barriers to completion of the single market that it had identified. In the area of payments, meanwhile, the European Automated Clearing Houses Association (EACHA) has worked to develop an interoperability framework between the various retail payment systems handling instant payment transactions.

With regard to the public authorities, part of the Eurosystem's mandate is to encourage dialogue with and between the various stakeholders as well as the pursuit of solutions promoting interoperability. This catalyst role was illustrated by the SEPA project and, more recently, the Target Instant Payments Settlement (TIPS) project, which is part of the Eurosystem's Vision 2020 strategy (see Chapter 7, Section 6). T2S' implementation was also an opportunity for the Eurosystem to play the role of catalyst in harmonising the functioning of the various national markets, by providing the impetus for an alignment of legislative, operational and technical frameworks (harmonisation of settlement cycle timings, message and data format etc.), which was necessary for T2S' smooth functioning. This work was carried out based on close collaboration with and strong involvement from the industry, with the consultation forums validating all standards choices and exerting peer pressure to move these harmonisation efforts forward.

3. The issue of pricing for FMI operators

The consequences of the market's functioning on possible pricing constraints and strategies call for public intervention, whether in terms of regulating the players themselves or acting directly as operator. In reality, the main pricing obligation imposed by FMI overseers is a requirement for transparency.

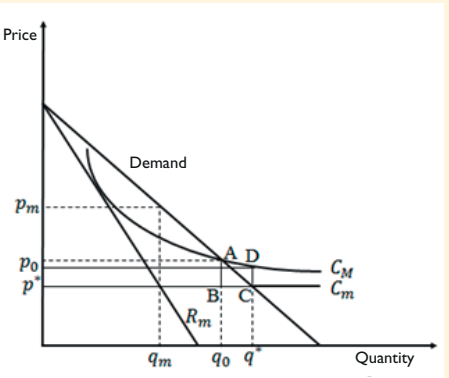
The Chart 1[46] illustrates the pricing issues associated with a single product monopoly. A number of price levels warrant observations. Pricing at marginal cost (price set at p^*) in theory represents a 'first-rate' solution, showing efficient allocation in the Pareto sense, as it maximises the collective surplus, thereby enabling all gains from the transaction to be consumed.[47] In practice, such pricing is unlikely to occur in the case of FMIs, not only because of the operators' market power described above, but also because in an industry characterised by high fixed costs and low variable costs (natural monopoly characteristics), it does not allow the operator that implements it to recover its costs (given that if it operates in the average cost zone, its average costs are by definition higher than its marginal costs). In the graph, income is equal to $p^* q^*$ and lower than costs $C_M q^*$, and losses are equal to the rectangle $p^* p_0$DC, which corresponds to the fixed costs.

Pricing below the average cost can be a temporary way of encouraging consumers to join the network in the short term to enable it to fully realise the associated positive network externalities. More specifically, a possible pricing policy could be to set prices at the lowest expected transaction processing unit cost once economies of

46 Taken from *Economie de la règlementation*, Lévêque F., La Découverte, 2004, which deals very lucidly with this subject.

47 This is the definition of a Pareto optimality.

C1: Pricing of a single product monopoly



Note: R_m is the monopoly's marginal income, and the cost function is determined on the basis of $C(q)=F+C_m$, where F corresponds to fixed costs and C_m the marginal cost, which are assumed to be constant and always lower than the average cost, C_M, which decreases as fixed costs are amortised (this corresponds to the cost structure described in Section 1). p^*, p_0 and p_m are respectively the prices set as the marginal cost, the average cost and the monopoly cost (the price that leads to a maximisation of profits), with the corresponding quantities q^*, q_0 and q_m.

scale have been fully achieved.[48] Where a central bank operates the infrastructure, this pricing equates to a public subsidy, which, in the event of contention about the service provided, is likely to discourage potential private operators from entering the market as they cannot recover their costs. Where a private player operates the infrastructure, such a policy can, as in the case of a public operator, also be designed to achieve the critical mass above which network externalities become positive. However, it may also be the result of a predatory strategy aimed at preventing an incoming or potential competitor from being profitable, in order to subsequently exploit market power and generate competition-free profits.[49]

A complex issue for FMIs operators is therefore, in this context, determining the optimal price structure to enable recovery of their costs.

First, assuming network effects, uniform pricing at the average cost (which on the graph corresponds to the fixed price p_0), and which in principle allows the operator to recover its costs but not to turn a profit, could discourage some users from joining the network (those whose willingness to pay is situated between the average cost and the marginal cost), with the result that the positive network externalities are not fully realised – which is also a problem from the point of view of allocative efficiency. Allocative inefficiency is represented in the graph by triangle ABC.

Second, in theory a natural monopoly will automatically choose to sell quantity q_m at the monopoly price p_m that maximises its profit and corresponds to where its marginal income equals its marginal cost. This price level is not Pareto-efficient insofar as the monopoly could profitably serve all consumers whose willingness to pay was situated between p_m and p_0.

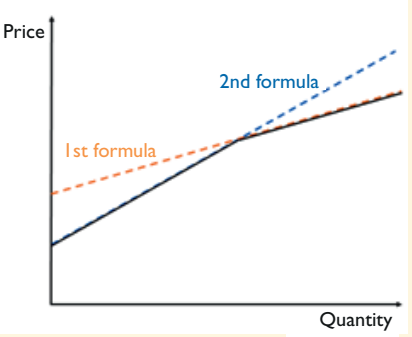
Finally, the pricing of the infrastructure's use can also be set in a non-uniform manner, with different prices for the same service[50] or depending on the volumes processed. The dominant pricing choice as far as FMIs are concerned is a non-linear pricing policy, in which the price per transaction processed by the infrastructure depends on the number of transactions that a participant initiates.

One variation on this pricing consists in determining a non-linear, two-part pricing structure $t(q)=F+cq$, including a fixed lump-sum part F (the objective of which is to recover fixed costs and reflects the resulting economies of scale), and a variable part proportional to the number of transactions (the objective of which is to recover the average variable costs). This is the type of pricing applied by the Bank of England in its CHAPS private large-value payment systems service,[51] with the aim of recovering long-term costs, at neither profit nor loss, and without cross-subsidisation between the various service lines.

A second alternative consists in two-part double pricing, based on the same principle, where individual participants choose the best-suited pricing method for them.

These two formulae are shown in the Chart 2. The graph shows that for small quantities it is more profitable to apply formula 1 (first part of the solid line) and for larger quantities formula 2 (second part of the solid line): an operator applying

C2: Double pricing of a single product monopoly



Note: The intercepts correspond to each formula's fixed cost and the lines' slopes to the constant marginal costs.

48 Bolt W. and Humphrey D., "Public Good Aspects of TARGET: Natural Monopoly, Scale Economies, and Cost Allocation", DNB Working Paper, 2005.

49 When pursued by a dominant operator, such a practice is likely to infringe competition rules. However, the competition authorities consider the proof of such an infringement to be conclusive only if particularly stringent criteria have been met.

50 The simplest scenario is to charge an amount on top of the marginal production cost that increases as the demand price elasticity decreases. However, in practice an adjustable pricing like this requires measuring demand price elasticity. It can lead, in a multi-product scenario, to the implementation of a cross-subsidy policy in which the services for which demand is most elastic are subsidised by those for which it is less elastic. Such a policy contravenes the objectives of certain central banks (e.g. the Fed) of recovering costs by service line, and (correspondingly) can enable certain players to pursue a 'cream skimming' policy by only serving market segments with low demand price elasticity and covering their related costs while benefiting from a comparative advantage over an operator implementing a cross-subsidy policy, which ultimately makes such a policy ineffective.

51 Annual membership price of £15,000 for both CHAPS and the DvP service; per item fee of £0.155 for CHAPS and £1.90 for DvP.

adjustable pricing will take into account the fact that each type of participant will choose the most profitable formula for them, and will set the pricing scales accordingly. In this case, the distinction is between participants carrying out a large number of transactions via the systems, and others.

This is the pricing method for the core services chosen for TARGET2, where participants can choose between two options. The first, in which there is a single price consisting of a monthly fee of EUR 150 and a transaction price of EUR 0.80, is intended for institutions with low transaction volumes. The second, consisting of a monthly fee of EUR 1,875 and sliding tariffs by transaction volume threshold, is intended for large institutions.

In both cases, the average price decreases as volumes processed increase, which has the same effect as a volume discount policy and limits cross-subsidisation between the different types of participants. Such a structure is generally considered to be efficient,[52] and encourages large-volume participants' use of the infrastructure.

52 For its application to payment systems, see: Holthausen C. and Rochet J.-C., "Efficient Pricing of Large Value Interbank Payment Systems", *Ohio State University Press*, 2006.

Box 4: Recovery of costs by central banks for the provision of RTGS services

Most central banks operating RTGS have a partial or total cost recovery policy. Some (the Bank of Japan, for example) suggest that RTGS should be subsidised given the benefits it can offer the community as a whole in terms of economic stability. The more contestable the nature of the services that they provide, in that they could to some extent be provided by the private sector, the more sensitive central banks are to the need to recover their costs. By contrast, if they seek only to partially recover their costs it can be assumed that there is nothing contestable about the service provided (for example, because payments in central bank money can only be made in RTGS operated by the central bank), or that it is feared that a total recovery policy will lead to pricing that discourages the use of RTGS for riskier payment systems.

In some cases this objective is a legal one. In the United States, the Monetary Control Act of 1980[1] imposes on the Federal Reserve a general long-term objective of recovering its direct and indirect costs. The objective pursued by Congress was both to stimulate competition (fair competition concept) and so provide services at the lowest possible cost for society, and to ensure that those services were adequate (role of prices as a signal with the aim of allocative efficiency). This principle is interpreted in a restrictive manner by the Fed, which sets itself an objective of total recovery of costs (production, investment and operational (including maintenance and operation), adjusted for those that would have been incurred by the private sector)[2] by service line.

In the Eurosystem, Article 2 of the Statute of the ESCB and of the ECB requires the ESCB to *"act in accordance with the principle of an open market economy with free competition, favouring an efficient allocation of resources"*, which suggests that the pricing of the services it provides should seek as often as possible to meet a cost recovery objective if a subsidy would be likely to hinder private-sector competition. The stated objective ofTARGET2 is to recover "significant" costs[3] in order to avoid unfair competition with private payment systems.

.../...

1 *https://www.federalreserve.gov/paymentsystems/pfs_pricingpol.htm.*

2 Defined as "an allowance for the taxes that would have been paid and the return on capital that would have been provided had the Federal Reserve's priced services been furnished by a private-sector firm."

3 *https://www.ecb.europa.eu/paym/t2/shared/pdf/professionals/TARGET2_pricing_guide_v4_updated.pdf?67e41c1f1858a8e9af59c9667e3660cf*, p.4.

In contrast, TARGET2 is not intended to fully recover its costs due to the contribution of a public good factor corresponding to the positive externalities generated by its use, in particular in terms of reducing systemic risk. This public good factor can be viewed and measured as the subsidy that would be needed for the private sector to 'insource' the costs related to taking into account systemic risks in the way it operates the payment system.[4] A number of economic studies conducted prior to the launch of TARGET2 estimated this public good factor at 20% of total costs. It reflects the fact that, in addition to its large-value payment settlement services for directly executed transactions, TARGET2 also provides:

- reserve account-keeping services for the final settlement in central bank money of net positions in transactions carried out in other FMIs (a core activity for central banks that only they can provide);
- intraday credit facilities, an extension of overnight credit systems, (by its nature, unquestionably a central bank activity); and
- fund transfer services for commercial banks' reserve accounts.

Given all these functions performed in addition to simple settlement, imposing a total cost recovery objective could lead to dissuasive pricing compared to private systems that do not fulfil the same roles and do not have the same constraints, which could prove to be less than socially optimal. In practice, when measuring the public good factor, the cost of producing services equivalent to commercial services is nevertheless more easily determined than that relating to externalities concerned with strengthening the economy's overall security.

As regards T2S, however, whereas the objective is one of full cost recovery, the platform does not seek to operate for profit. This objective was notably decided on for competitive reasons, insofar as it was a question of 'insourcing' activities (operation of the technical platform providing the settlement and delivery service) that were previously CSDs' responsibility, and which therefore was by definition partially questionable.[5]

4 For further details, see: Holthausen C. and Rochet J.-C., "Incorporating a "public good factor" into pricing of large-value payment systems", European Central Bank Working Paper, July 2005.

5 "Partially" because, whatever the model, the central bank remains the only player to be able to provide a central bank money wholesale payment service.