

Network Effects, Compatibility,
and Adoption of Standards:
Essays in Empirical Industrial Economics

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum politicarum
(Doktor der Wirtschaftswissenschaften)

eingereicht an der

Wirtschaftswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von

Magister Ekonomii Michał Grajek,
geboren am 23.05.1975 in Warszawa (Polen)

Präsident der Humboldt-Universität zu Berlin:

Prof. Dr. Jürgen Mlynek

Dekan der Wirtschaftswissenschaftlichen Fakultät:

Prof. Dr. Joachim Schwalbach

Gutachter 1: Prof. Lars-Hendrik Röller, Ph.D.

Gutachter 2: Prof. Dr. Ulrich Kamecke

eingereicht am: 06.10.2004

Tag der mündlichen Prüfung: 10.12.2004

Abstract

This thesis introduces a structural econometric model of demand exhibiting direct network effects. The structural approach we follow allows us to identify the extent of network effects and compatibility between competing networks. To the best of our knowledge, this is the first model that facilitates the identification in the case of *direct* network effects. At the same time, it is the first attempt to assess the degree of compatibility in an econometric framework. The model is then applied to investigate demand for mobile telecommunications service and the link between adoptions of ISO 9000 standard and international trade. The estimation results allow us to formulate some interesting policy conclusions.

Keywords:

Network Effects, Compatibility, Identification, Estimation, Mobile Telecommunications, ISO 9000, International Trade

Zusammenfassung

Diese Dissertation stellt ein strukturelles ökonometrisches Modell der Nachfrage mit direkten Netzwerkeffekten vor. Der strukturelle Ansatz ermöglicht es, den Umfang von Netzwerkeffekten and Netzwerkkompatibilität zu identifizieren. Nach unserem Wissen ist es das erste Modell, das diese Identifikation im Falle von *direkten* Netzwerkeffekten ermöglicht. Gleichzeitig ist es der erste Versuch, den Umfang von Netzwerkkompatibilität mit ökonometrischen Methoden zu bestimmen. Das Modell wird dann angewandt, um die Nachfrage von Mobilfunkdienstleistungen, als auch den Zusammenhang zwischen ISO 9000 Normierung und internationalem Handel zu untersuchen. Die Schätzungsergebnisse führen zu interessanten wirtschaftspolitischen Folgerungen.

Schlagwörter:

Netzwerkeffekten, Netzwerkkompatibilität, Identifikation, Schätzung, Mobilfunktelekommunikation, ISO 9000, Internationaler Handel

Contents

List of Tables.....	iii
List of Figures	iv
Acknowledgements	v
Chapter 1 Introduction	1
1.1 Why is this interesting?	1
1.2 Literature	2
1.3 Contribution of the thesis	5
Chapter 2 Identification of Network Externalities in Markets for Non-Durables	8
2.1 Introduction	8
2.2 Economic model.....	12
2.2.1 Willingness-to-pay function.....	12
2.2.2 Subscription demand	13
2.2.3 Switching costs.....	15
2.2.4 Supply of the network good	16
2.2.5 Dynamics of the network good adoption	17
2.3 Stochastic model	21
2.3.1 Functional specification	21
2.3.2 Identification	23
2.3.3 Stochastic structure	26
2.3.4 Interpretation of the identified structural parameters	27
2.4 Conclusions	29
2.5 Appendices.....	31
2.5.1 Derivatives of the function H with respect to $x(t-\delta)$ and p	31
2.5.2 State equations with firm-specific constants.....	32
2.5.3 Figures.....	34

Chapter 3 Estimating Network Effects and Compatibility in Mobile Telecommunications	35
3.1 Introduction	35
3.2 Mobile telecommunications	37
3.2.1 Network effects	37
3.2.2 Description of the market	39
3.3 A model of demand for mobile telephone service	40
3.4 Empirical evidence	44
3.4.1 Data	44
3.4.2 Estimation issues	45
3.4.3 Results and interpretations	48
3.5 Conclusions	53
3.6 Appendices	55
3.6.1 Direction of bias in the estimated price elasticity of demand	55
3.6.2 Tables	57
3.6.3 Figures	59
Chapter 4 Diffusion of ISO 9000 Standards and International Trade	63
4.1 Introduction	63
4.2 ISO 9000 and its role in international trade	65
4.3 Economic models	68
4.3.1 International trade flows	68
4.3.2 ISO 9000 diffusion	70
4.4 Empirical model	72
4.4.1 Data	72
4.4.2 Implementation of gravity equation for trade	73
4.4.3 Implementation of ISO 9000 diffusion equation	74
4.4.4 Discussion of the results	77
4.5 Conclusions	82
4.6 Appendices	84
4.6.1 Optimal timing of ISO 9000 adoptions	84
4.6.2 Tables	86
Bibliography	92

List of Tables

Table 3.1. Demand in the Polish mobile telephone industry: Estimation results..... 57

Table 3.2. Identified structural parameters of demand in the Polish mobile telephone industry
..... 58

Table 4.1 List of countries and ISO 9000 adoptions..... 86

Table 4.2 Variable description and summary statistics..... 87

Table 4.3 Gravity equation for exports: Fixed-effects estimation results 88

Table 4.4 Gravity equation for exports: First-differencing estimation results..... 89

Table 4.5 Gravity equation for exports (only OECD countries): First-differencing estimation
results 90

Table 4.6 ISO 9000 adoption equation: Arellano-Bond estimation results 91

List of Figures

Figure 2.1 Stable vs. unstable equilibria 34

Figure 3.1 Best-deal prices in the Polish mobile telephone industry 59

Figure 3.2 Number of subscribers in the Polish mobile telephone industry 60

Figure 3.3 Aggregated steady-state demand in the Polish mobile telephone industry 61

Figure 3.4 Simulations of the aggregated steady-state demand in the Polish mobile telephone industry for different values of the structural parameters 62

Acknowledgements

While working on this thesis I received support from many people, whom I would like to thank. I am particularly grateful to my supervisors Lars-Hendrik Röller and Ulrich Kamecke, whose advice and encouragement enormously helped me pursuing the thesis. Lars-Hendrik Röller not only gave me the opportunity to study in stimulating research environment of the Wissenschaftszentrum Berlin (WZB), he also provided me with helpful comments and suggestions throughout. Ulrich Kamecke, as head of the doctoral program “Applied Microeconomics” at the Humboldt University and the Free University Berlin, made the launch of my doctoral project possible. While writing the thesis I very much benefited from his critical comments.

I am also grateful to the fellows at the WZB, who supported me with valuable insights and ideas. In particular I would like to thank Joe Clougherty, Paul Heidhues, Jos Jansen, Oz Shy, and Christian Wey, who read and commented on parts of this thesis.

Finally, I would like to thank my wife Gosia for her patience and understanding and my little son Ignacy for all the sleepless nights.

Chapter 1

Introduction

1.1 Why is this interesting?

The interest in network economics, where network effects, compatibility, and standards play a predominant role, has risen enormously over the last decades. It gave rise to a large body of theoretical literature, which showed that the existence of strong network effects altered market outcomes in many important ways, often leading to market failure. Examples of such market failure include market breakdown due to start-up and hold-up problems, as well as inability of industries to switch to a superior standard, known as “excess inertia”.

In fact, these topics have also attracted a lot of attention outside economics, including business scientists and practitioners, as well as policy makers. One reason for this interest is that the network aspects are more apparent in the fast growing and widely reported industries – the information industries – than in more traditional ones. The famous antitrust cases of AT&T, IBM, and Microsoft contributed to the popular interest in the information industries, as well as challenged theoretical and empirical economists.

It has been claimed that, since we are living in a “New Economy”, a “New Economics” is needed to understand its functioning and to guide business strategy and public policy. Although, this claim seems exaggerated (Shapiro and Varian, 1999, p. ix-x), there is definitely a need for more research in network economics. In particular, this applies to empirical research, since quantification of network effects and compatibility in real markets is still a largely uncovered issue.

1.2 Literature

In general, demand for a good exhibits network effects, if the utility of each consumer increases with the number of other consumers purchasing the same good. The notion of network effects is very closely related to – and in fact often interchangeably used with – the notion of network externalities, positive feedback, and demand-side economies of scale.

Economic theory predicts that network effects can lead to multiple equilibria and particular market shares' dynamics. Multiple equilibria may result even if the network good is offered by a single provider. The seminal contribution in the field is the work by Rohlfs (1974), who studies demand for telephone service. He shows that a monopolist offering the service faces multiple outcomes in terms of equilibrium user sets (network size) at a given price. He then points out the start-up problem for a new communications service stressing that a critical mass of users is needed in order to secure the large network size. In dynamic models, overcoming of the critical mass is reflected by an S-shaped diffusion of network good adoptions (Cabral, 1990).

Katz and Shapiro (1994) provide an overview of market responses to the start-up problem in communication networks discussing various pricing and expectations-management strategies. They extend the discussion to system networks, which consists of two components consumed together: hardware and software. In such systems, indirect network effects can arise, as opposed to direct effects in communication networks. Indirect effects arise, if utility of the hardware users increases with the number of other hardware users, because it positively influences quality and variety of the available software. Providing such systems is potentially vulnerable to hold-up problem, since the buyers of durable hardware could expect to be exploited by the future software price increases (Farrell and Gallini, 1988). Under-utilization of the communication networks and the system networks due to the start-up and hold-up problems is one example of market failure in provision of network goods. In extreme cases (e.g. pure network goods), such under-utilization can lead to complete market breakdown.

Another example of market failure studied in the literature on network economics is “excess inertia” or lock-in by historical events (Farrell and Saloner, 1985; Arthur, 1989). A particularly entertaining reading in this strand of literature is David (1985) in which the author illustrates how QWERTY keyboard standard won the competition with more efficient Dvorak keyboard standard. Industries can be locked-in in an inferior standard, because network

markets are prone to “tipping”, which is the tendency of one standard to drive the competitors out of the market once it gained a lead. Tipping occurs in static (one-shot) models in the form of multiple equilibria in which a single standard dominates (Katz and Shapiro, 1985). In dynamic models, tipping is reflected in equilibria where the adoptions of the losing standard die out once a rival standard is introduced or accepted in the market (Farrell and Saloner, 1986a; Katz and Shapiro 1992, 1994). Tipping does not necessarily, however, lead to excess inertia in introducing new superior standards, as stressed in the earlier works. Indeed, properly defined property rights could alleviate or even reverse the excess inertia problem (Katz and Shapiro, 1986a, 1992; Liebowitz and Margolis 1990).

The above mentioned properties of markets featured by network effects give rise to particularly rich array of strategies and tactics on the side of firms, especially when compatibility is a decision variable (Katz and Shapiro, 1994; Farrell and Klemperer, 2001). Moreover, evaluation of these strategies and tactics by anti-trust authorities is a much more delicate issue in network industries than in more traditional industries (Economides and White, 1994; Röller and Wey, 2003). For instance, in contrast to traditional industries, marginal cost pricing may result in market failure due to the start-up problem in network industries. Hence, introductory pricing below marginal costs may be necessary as a means to achieve large network size. Yet, it is difficult to distinguish introductory pricing from anti-competitive predatory pricing in practice. At the same time, market dominance does not necessarily imply that the market leader earns super-normal profits. The quasi-rents may reflect the costs incurred earlier in order to attract the critical mass of consumers.

An important factor conditioning the impact of network effects is compatibility. If the goods are perfectly compatible, consumers of all the goods make up a common network and contribute to the network effects at the industry level. In the polar case of perfect incompatibility, consumers of each good form separate networks and the network effects operate at the firm level.

The importance of compatibility between competing network goods for performance of network industries is twofold. First, it directly influences the gross consumption benefits of consumers. They directly benefit from compatibility since it expands the size of each network to the total membership of both. The drawback of compatibility, however, is a loss of variety (Farrell and Saloner, 1986b). In general, we might expect that with strong network effects any loss of variety is a minor price to pay to achieve compatibility, as in the case of VHS vs. Beta in videocassette recorders (Cusumano et al., 1992). Second, compatibility alters the nature of

competition between the network good providers thereby indirectly influencing the benefits of consumers. As stated by Besen and Farrell (1994), the providers compete “for the market” under incompatibility, whereas under compatibility they compete in a more traditional manner “within the market”. In other words, winning or monopolizing the market will more often occur under incompatibility. The reason behind it is that positive feedback at the firm level (incompatibility) makes network markets more prone to tipping than positive feedback at the industry level (compatibility). A natural question to ask then is how compatibility affects the overall degree of competition between the providers. There is no simple answer to it. Compatibility tends to relax competition early in the product-life cycle, because the threat of tipping is reduced. However, it tends to intensify competition later on, because it makes monopolization of the market less likely (Katz and Shapiro, 1986b).

Empirical research on network effects confirms their importance in real markets. The following brief overview distinguishes between descriptive and structural econometric analyses (Reiss and Wolak, 2002). The difference is that the former proceeds without reference to an economic model, whereas the essential components of the latter are the theoretical and statistical assumptions that allow a researcher to recover economic primitives from data. As a consequence, estimates of economic magnitudes and the extent of causation can be recovered solely from structural analyses.

The early wave of descriptive econometric studies confirmed that compatibility matters in network markets, including mainframe computer systems (Greenstein, 1993) and computer software (Gandal, 1994, 1995; Brynjolfsson and Kemerer, 1996). Further, empirical evidence of network effects was found in the adoption of automated teller machines (Saloner and Shepard, 1995), microcomputer systems (Gandal, Greenstein, and Salant, 1999; Goolsbee and Klenow, 2002), mobile telephone service (Okada and Hatta, 1999; Kim and Kwon, 2003), and DVD players (Dranove and Gandal, 2003).

The small but growing structural econometric literature focuses on identification of indirect network effects, which arise in system networks. Typically, the identification is achieved by estimating demand for both hardware and software simultaneously and testing for interdependence between them. Substantial extent of indirect network effects was found in markets for compact disc players (Gandal, Kende, and Rob, 2000) and video cassette recorders (Ohashi, 2003). Moderate extent of the effects was reported in Yellow Pages market (Rysman, 2002).

Structural econometric modeling of demand exhibiting direct network effect is still a largely uncovered issue. Economides and Himmelberg (1995) conducted the pioneering study in this field. They estimated demand in the market for facsimiles assuming perfect competition on the supply side. The parameters of utility function that they recovered from the estimation results suggest extremely strong network effects in this market. Our work builds on the Economides and Himmelberg (1995) carefully examining the identification issues and introducing imperfect competition into the model.

1.3 Contribution of the thesis

The thesis proposes a structural econometric model of demand exhibiting direct network effects. The structural approach we follow allows us to identify the extent of network effects and compatibility between competing networks. To the best of our knowledge, this is the first model that facilitates the identification in the case of *direct* network effects. At the same time, it is the first attempt to assess the degree of compatibility in an econometric framework. The model is then applied to investigate demand for mobile telecommunications service and the link between adoptions of ISO 9000 standard and international trade. The estimation results allow us to formulate some interesting policy conclusions.

The first paper of this thesis, “Identification of Network Externalities in Markets for Non-Durables”, introduces the economic structure of the model. Given the economic primitives, we derive the equilibrium diffusion path of the network goods’ adoptions. There are two major assumptions that drive the identification of network effects and compatibility in the model. First, network effects are captured by *lagged* network size in the consumer willingness-to-pay function. Second, providers of the competing network goods set equal hedonic prices.

The first assumption assures uniqueness and continuity of the equilibrium diffusion path. Using lagged instead of current network size in the willingness-to-pay function, we implicitly rule out coordination between consumers. Without coordination, the adoptions follow minimum equilibrium diffusion path, which still will be discontinuous if the lag is infinitely small. In practice, the frequency of data used to estimate the model will determine the lag thereby making the diffusion path continuous. We also discuss the implications of the lag for the extent of identified network effects and rationality of consumers in the model.

The second assumption means that prices “corrected” for the network size are equal across the competing network goods. Hence, without other quality characteristics, which affect value of the goods, nominal (“uncorrected”) prices that are unequal across the goods imply the existence of network effects and incompatibility. At the same time, network effects, as captured by the first assumption, trigger the impact of lagged network size on current network size. This inertia in network size dynamics together with the nominal price differential across goods identifies the extent of network effects and compatibility in the econometric model.

We apply this model to investigate demand for mobile telecommunications service in the second paper of this thesis, “Estimating Network Effects and Compatibility in Mobile Telecommunications”. The firm-level data we use stems from the Polish mobile telephone industry and covers the period 1996-2001. Our empirical analysis points to strong network effects and, despite full interconnection of the mobile telephone networks, low compatibility. These findings suggest intra-network call discounts and transmission of information on quality as main sources of network effects in mobile telecommunications. Finally, we show that ignoring network effects causes overestimation of price elasticity of demand. One policy conclusion from this is that subsequent measure of market power will be overestimated as well.

The last paper, “Diffusion of ISO 9000 Standards and International Trade”, investigates the impact of ISO 9000 adoptions on bilateral trade flows. The ISO 9000 family of standards is designed to provide confidence to buyers that products will meet their expectations thereby enhancing trade and global welfare. However, its critics claim that it is merely a barrier to market entry and a tariff on international trade. Using a panel data set on 101 countries over 1995-2001 we estimate a gravity equation for bilateral exports in which domestic and foreign ISO 9000 adoptions affect bilateral trade barriers. One econometric caveat there is that ISO 9000 adoptions could be endogenous. The first paper of this dissertation offers a way to endogenize the adoption decisions. The basic idea being that ISO 9000 is a common language lowering informational asymmetries between firms. Consequently, we expect value of these standards to increase with the number of adopters giving rise to network effects. Although, formal tests do not reject exogeneity of regressors in our trade equation, we still estimate a diffusion equation of ISO 9000 in order to obtain additional evidence of the role of these standards in international trade.

In general, our empirical results are consistent with the common language hypothesis. We find that the adoptions of ISO 9000 benefit bilateral trade, as well as spur further domestic and foreign adoptions. At the same time, we observe a substitution effect, as certified firms tend to trade more with each other than with uncertified firms. In fact, this creates an effective barrier to trade for less developed countries. The substitution effect works against them, because adoptions of the standards are concentrated in more developed countries. One policy conclusion we draw from this is that the benefits of ISO 9000 will remain in the more developed countries domain unless we find and remove the barriers to its global diffusion.

Chapter 2

Identification of Network Externalities in Markets for Non-Durables

2.1 Introduction

This paper introduces a structural econometric model of consumer demand for non-durable goods or services exhibiting network externalities. Its main contribution is that it allows us to identify structural parameters that determine the extent of externalities in the underlying economic model from the empirical estimation results. The structural parameters' estimates can be employed in turn to test the economic significance of the externalities and the compatibility between competing networks. The structure that we derive is mainly suited to deal with direct network externalities, e.g. as in telecommunication services. One can also think of it, however, as of a reduced form arising from indirect network externalities. E.g. in the case of experience goods, the installed base of consumers could matter, if they transmit information about quality of the good.

Generally, positive network externalities mean that utility, which users derive from consumption of a given good or service, increases with the number of other users.¹ The modern economic literature usually distinguishes two major types: direct and indirect network externalities (see e.g. Katz and Shapiro 1985, 1994; Economides and White 1994; Economides 1996). The first one is related to physical networks, e.g. supported by telecommunication technologies like telephone, telegraph, facsimile or e-mail. Clearly, the utility, which consumers derive from using any of these technologies, increases with the

¹ Throughout the paper, the term network externalities is used interchangeably with the term network effects. The difference between the two is that in addition to network effects, network externalities imply also a market failure (see Liebowitz and Margolis, 1994). In the context of our paper this might depend on interpretation of the economic model.

number of other users. The most obvious reason is that a larger network allows consumers to satisfy more communication needs. The other reason might be the bandwagon effect, which arises because conspicuous consumption gives rise to a conformist behavior as argued by Leibenstein (1950). Blonski (2002) considers another explanation in the context of telecommunication market with competing networks. In his model, it is cheaper for consumers to call within their network, since network suppliers charge access fees for the calls from outside into their networks. As a consequence, consumers benefit from a larger network, because it implies a lower monthly bill. The negative dependence between network size and utility that consumers derive from network good might be justified by congestion or by non-conformism of consumers.

In turn, in a typical virtual network the externality is indirect and comes from the hardware/software paradigm (Katz and Shapiro, 1985). It applies when a good consists of two complementary components: hardware, which is durable, and software, which exhibits supply-side economies of scale. The number of users of a given hardware is relevant since it determines the size of the market for software and influences positively its variety and quality, hence enhancing the utility from using that hardware. This way of reasoning may be applied to computer operating systems, credit cards, video recorders, phonograph equipment etc.

The main difficulty, which our structural econometric model faces, is the multiplicity of equilibria, a common result in theoretical studies of markets featured by network externality. In a one-shot (or static) setting, multiple equilibria are due to coordination problems (Farrell and Klemperer, 2001, pp. 47-50). The simplest example, with one pure network good may be found in Economides and Himmelberg (1995).² They show that consumers' expectations of no network good provision as well as positive levels of the network good sales at a given non-negative price may actually be self-fulfilling equilibrium outcomes.

It is also a common wisdom that network externalities could give rise to S-shaped diffusion of network good's adoptions over time. Multiple steady states in dynamic models of demand with network externalities are analogous to multiple static equilibria. Switching from the low steady state to the high one can be seen as network diffusion (Cabral, 1990). Since there are infinitely many diffusion paths, which are supported by fulfilled consumers' expectations, the question of interest is when and how fast the diffusion occurs. Cabral (1990)

² However, seminal works are Rohlfs (1974) and Katz and Shapiro (1985)

addresses this question in a perfectly competitive setting with one network good. As an equilibrium selection rule, he introduces lagged instead of expected network size into consumers' willingness-to-pay function. By doing so, he obtains a unique network diffusion path, even when the lag length is infinitely small, in which case, consumers are rational. A drawback of this approach is that the infinitely small lag causes at the same time a discontinuous jump in the equilibrium network diffusion path. In other words, the "rational" diffusion process is infinitely fast. Being aware of this counterfactual feature, Cabral (1990) argues that the discontinuous diffusion path in his model can be treated as an approximation to the empirically observed S-shaped diffusion. In fact, assuming small but non-zero perception lag yields such result.

Economides and Himmelberg (1995) propose another solution in the context of perfectly competitive market. The unique network diffusion path in their model results from the assumption that supply of the network good is finitely elastic in the sense that the marginal cost function depends positively on the derivative of network size with respect to time. In other words, the change of network size is costly and these additional (over marginal) costs are passed through to consumers. In this model consumers form expectations and the expectations are fulfilled along the equilibrium network diffusion path. Moreover, the assumption about finitely elastic supply resolves the discontinuity problem discussed above. This is because by construction of the marginal cost function, a discontinuous diffusion would imply an infinitely high price.

In this paper we take the former approach. Following Cabral (1990), we assume that consumers care about the lagged network size in their decision about buying the network good (joining the network). This approach is very appealing from an empirical perspective. First, as we will show, it allows us to identify the structural parameters from the estimation results. Economides and Himmelberg (1995) fail to prove that one can do this in their model. Second, the use of lagged dependent variable is common in econometric practice.

The limitation of the approach we follow is that consumers are rational only when the lag length tends to zero, at least in our simple setup.³ If the lag becomes larger, as it is the case with empirical data time series, consumers do not consider that during the diffusion process the network grows in current period. Another implication of the model is that empirical magnitude of network effect depends to some extent on lag length, hence on data

³ E.g. in the case of experience goods the lag might be justified, because the network externalities arise through information transmission, which takes time. Also in the case of direct network externalities, one might argue that the lag results from costly updating of information about the network size.

frequency. This is because the stronger network effects and more frequent updating of the network size both speed up the diffusion.

The model that we derive is also closely related to the marketing literature on diffusion of innovations. In seminal work of Bass (1969), a structural econometric model of new product diffusion is developed, which is driven solely by the diffusion of awareness of this product. The striking feature of the original Bass' (1969) model is that price does not influence the diffusion. The marketing scientist recognized that puzzle and developed many extended models with the price incorporated (e.g. Horsky, 1990; Jain and Rao, 1990; Bass, Krishnan and Jain, 1994). Our structural econometric model is not an extension of the seminal Bass' (1969) model. We use network externalities instead of imperfect information in order to facilitate the innovation diffusion. The equation to estimate in our example, however, encompasses the equation proposed by Bass (1969).

So far, there is little empirical work that assesses the consumption network externality. Some works follow descriptive approach looking for an empirical evidence of network effects. Greenstein (1993) conducts the first research in that stream. He shows that compatibility with the installed base matters in the choice of the mainframe computer system. Gandal (1994, 1995) estimates hedonic price equations for spreadsheets and data base management systems and finds that consumers are willing to pay significant premium for software supporting a common file compatibility standard. This result is in line with the hypothesis that the software markets exhibit network externalities. Similar findings report Brynjolfsson and Kemerer (1996). Additionally, they find that a product's installed base increases the price of spreadsheets. Gandal, Greenstein and Salant (1999) show the two-way positive feedback between different components in competing microcomputer systems by means of vector autoregressive (VAR) analysis. In this way, they prove empirically the indirect network externality hypothesis. Dranove and Gandal (2003) report the evidence of network externalities in the DVD market. They find that the measure of software availability defined as the percentage of DVD format releases among the top selling films spurs the adoptions of DVD players.

Economides and Himmelberg (1995) conduct the pioneering structural econometric study of demand with network externalities. In the theoretical part of the paper they derive a dynamic model of a perfectly competitive market with consumption network externalities. The possible multiplicity of equilibrium network diffusion paths under fulfilled consumers' expectations is solved by the assumption of finitely elastic supply, as described above. In the

empirical part, however, the assumption that facilitates the estimation is that expected network size is a linear function of the past network size. Hence, it is not clear how to link the empirical results back to the theoretical model. Other examples of structural econometric work in this field include Gandalf, Kende, and Rob (2000) for the CD industry and Rysmann (2002) for the Yellow Pages market. These authors concentrate on the indirect network effect and estimate two interrelated demand equations, one for software and one for hardware. In this way, they model the complementarities between software and hardware in full instead of including the network size in the consumers' utility function as in the direct network externality case.

The paper is organized as follows. Section 2.2 introduces the economic model, which yields the structure for the empirical investigation. Section 2.3 gives an example of the functional specification that leads to a simple stochastic model and discusses the identification and interpretation of the structural parameters. Section 2.4 concludes.

2.2 Economic model

2.2.1 Willingness-to-pay function

The demand model we use is a partial equilibrium, discrete choice, dynamic model. The good being considered is non-durable, *ex ante* homogenous⁴ and subject to network externalities. We refer to the good supplied by different firms as brands. Consumers' willingness to pay for a given brand is influenced by his/her type and by the network size of that brand. We refer to the network as a set of subscribers and to the purchase of the non-durable network good as subscribing to the network.

Denote by i ($i = 1, 2, \dots, I$) the brand of the homogenous good and assume that there is a measure one of infinitely living consumers, each demanding at most one unit of the good. Consumer v 's preference for brand i at time t is represented by the instantaneous willingness-to-pay function $u(v, x_i(t-\delta))$, where v is the individual preference parameter, $x_i(t-\delta)$ is the lagged network size of brand i and the perception lag δ in an arbitrary number. Formally, we assume that the individual preference parameter v is distributed over the interval $[0,1]$ according to the cumulative density function $F(v)$, and that $u(v, x_i(t-\delta))$ is strictly increasing

⁴ By *ex ante* homogeneity we mean that different brands of the good are perceived as intrinsically equal. However, the difference in their valuation is possible *ex post*, when they have different network sizes.

and continuous in v . By construction, the parameter v establishes a rank ordering of the consumers according to their willingness to pay. We assume that the ranking is invariant with respect to changes in $x_i(t-\delta)$. As a matter of convention, the higher v is, the larger is the benefit of using each network. Network externalities are captured by the dependence of each consumer's willingness to pay on the network size $x_i(t-\delta)$.

Introduction of the lagged network size $x_i(t-\delta)$ into the willingness-to-pay function is crucial to our model. As pointed out by Cabral (1990), it is an equilibrium selection device that gives us the unique diffusion path of each network i . However, there appears a natural concern about consumer's rationality in this setting. Cabral (1990) argues that if the lag δ is infinitely small the consumers are rational. This is because their subscription decisions are identical to the ones done by forward-looking consumers. The construction of willingness-to-pay function, however, does not allow them to coordinate in order to switch from the low steady state to the high one, when both are feasible. Section 2.2.5 explains these findings in detail.

From the empirical perspective, the lagged network size in the willingness-to-pay function corresponds to the lagged dependent variable in the estimated equation. Obviously, the lagged dependent variable is easier to work with than with some unobserved expectations of the consumers. The cost of this approach is that we have to give up the rationality of the consumers with respect to the network size in most of the cases. This is because the minimum lag length is naturally defined by the frequency of the data and is usually large. Consequently the approximation of the rational consumers' behavior is poor.

2.2.2 Subscription demand

The network of each brand i constitutes of the set of its subscribers. If the brands are incompatible, each makes up its own network so $x_i(t-\delta) = y_i(t-\delta)$, where $y_i(t-\delta)$ stands for the normalized sales of brand i (the number of subscribers to brand i). If the brands are perfectly compatible, the network is common, which is given by total sales of all brands

$x_i(t-\delta) = \sum_{j=1}^I y_j(t-\delta)$. Because of the assumed homogeneity, the brands with identical

network size (in particular compatible brands) are perceived by consumers as perfect substitutes.

In a more general setting, partial compatibility may prevail. In this case, the network size of a brand is a weighted sum of its own and all other subscribers. Assuming symmetric compatibility, we write it as $x_i(t-\delta) = y_i(t-\delta) + w \sum_{j \neq i} y_j(t-\delta)$, where $w \in [0,1]$ measures the degree of compatibility. $w = 1$ and $w = 0$ correspond to the perfect compatibility and perfect incompatibility respectively and the interior values of w indicate a partial compatibility.

In each instance of time, consumer v decides to buy one of the brands or to stay out of the market in order to maximize her net utility

$$u(v, x_i(t-\delta)) - p_i(t). \quad (2.1)$$

If (2.1) is negative for all brands, than she will not join any of them. This “static” decision rule in our dynamic model is appropriate, as we focus on non-durable goods. In the context of telecommunication service this would mean that consumers could initiate or relinquish their subscription costlessly.

The consumer, for whom (2.1) equals zero, is indifferent between subscribing to and staying out of a given network. Denote $v_{i,t}^* = v^*(x_i(t-\delta), p_i(t))$ the type of the indifferent consumer with respect to brand i in time t . $v_{i,t}^*$ can be obtained from

$$u(v_{i,t}^*, x_i(t-\delta)) = p_i(t). \quad (2.2)$$

The brand i for which $v_{i,t}^*$ is the lowest is the most attractive brand for all subscribers in time t . We define it as

$$v_{L,t}^* = \min_i \{v_{1,t}^*, v_{2,t}^*, \dots, v_{I,t}^*\}. \quad (2.3)$$

By construction, all consumers with higher preference parameter than $v_{L,t}^*$ buy the good. If $v_{i,t}^*$ is equal among some brands, then the subscribers choose among them with equal probability. Define

$$H_i(\mathbf{v}_t^*) \equiv \begin{cases} \frac{1 - F(v_{L,t}^*)}{I_{L,t}} & \text{if } v_{i,t}^* = v_{L,t}^* \\ 0 & \text{otherwise} \end{cases}, \quad (2.4)$$

where $\mathbf{v}_t^* = (v_{1,t}^*, v_{2,t}^*, \dots, v_{I,t}^*)$ is a vector of the indifferent types with respect to brand i in time t , $I_{L,t}$ is the number of brands for which $v_{i,t}^* = v_{L,t}^*$ and F is the distribution function of v .

H_i equals the number of the consumers willing to buy brand i in time t . Now, the state equations, which describe the evolution of each brand's sales over time, are given by

$$y_i(t) = H_i(\mathbf{v}_t^*). \quad (2.5)$$

In the steady-state (given that all prices stay constant) we expect that none of the consumers can increase her utility by changing the subscription decision, so each brand's sales stay constant over time

$$y_i(t) = y_i(t-\delta). \quad (2.6)$$

It is worth noting that the steady-state equilibrium demand of the above model coincides with the standard static model equilibrium with fulfilled consumers' expectations (Rohlf's, 1974; Economides and Himmelberg, 1995). Moreover, the process of achieving equilibrium, which we have described formally, is in line with the logic presented by Rohlf's (1974).

2.2.3 Switching costs

The above model of demand with network externalities is probably the most obvious extension of the Cabral (1990) single brand model. It possesses however some unnatural features. One of them is a particular symmetry. In each instance of time, every active firm has an equal number of subscribers $y_i(t)$, which stays in contrast to the observation that real firms' market shares exhibit persistent differences.

The other feature corresponds to the Bertrand's paradox. If one firm undercuts the others just a little bit it wins immediately the whole market. This creates a strong incentive to undercut and results in fierce price competition. Moreover, with incompatibility, it is extremely difficult to recoup market shares once a firm lost its customers. This is because without installed base it needs to offer far more attractive price than the rival, which just won the whole market. In that case the Bertrand's paradox is even stronger and the market outcome is extremely tippy.

Switching costs offer a solution to the problems mentioned above and are particularly relevant to network markets. In fact, network externalities and switching costs are closely related to each other (Farrell and Klemperer, 2001).

Suppose, switching costs are high enough, such that having bought one brand, consumers will never find it optimal to switch to another one later on. The type of switching costs we have in mind can be observed in mobile telecommunication markets. That is, consumers have to pay a penalty for premature cancellation of a long-term contract. As a security option, they have however the right to relinquish the subscription without any penalty when the firm raises the price. In other words, once the price goes up the switching costs are gone.

The introduction of switching costs of this kind changes the subscription demand described in the previous section to the extent that only the unattached consumers can feed the diffusion of the networks. So, we can rewrite (2.4) as

$$H_i'(v_t^*, v_{t-\delta}^*) \equiv \begin{cases} \frac{F(v_{L,t-\delta}^*) - F(v_{L,t}^*)}{I_{L,t}} & \text{if } v_{i,t}^* = v_{L,t}^* \\ 0 & \text{otherwise} \end{cases} \quad (2.7)$$

Now, H_i' equals the number of the *new* consumers willing to buy the brand i in time t . Accordingly, the state equations are given by

$$y_i(t) = H_i'(v_t^*, v_{t-\delta}^*) + y_i(t-\delta). \quad (2.8)$$

This demand specification allows for persisting differences in market shares of different brands. In particular, the incumbent's installed base of consumers constitutes a persisting competitive advantage over the entrant.

Together with switching costs we introduce new issues concerning the pricing by the firms. First of all, remember that the specification in (2.7) remains valid until the prices go up. Otherwise, we are back to the set-up without switching costs as in (2.4). The switching costs will change also the price setting itself. We discuss that to some extent in the next section.

2.2.4 Supply of the network good

To complete the economic model of the market we would need to model how the prices are determined. Since the paper focuses on the demand side of the market and, in particular, on identification of the network effects, we do not introduce a structure for the supply side. Instead, we discuss some possibilities of extending the economic model to contain the explicit pricing relation as well.

In the simplest case without switching costs, we could plausibly assume that fierce price competition drives prices down to marginal costs. If firms are symmetric regarding their production technology, prices of all brands will be equal and their changes over time will reflect some technological progress and/or economies of scale. As a consequence, market outcomes will be completely symmetric. If the firms start their activity simultaneously with zero network size their brands will remain equally attractive for the consumers. In other words, expression (2.1) will be equal for all i . No firm will drop out of the market ($I_{L,t} = I$) and their networks will grow (or decline) equally fast. This set of assumptions facilitates static, marginal-costs pricing relation, which is well established in the empirical Industrial Organization literature (Bresnahan, 1989).

Once we introduce switching costs, a space for strategic pricing emerges. Indeed, switching costs in our set up tend to reduce competition and give firms the opportunity to earn some mark-ups.⁵ In that case, firms face a trade-off. On the one hand, they want to keep prices high in order to exploit the installed base of consumers. On the other hand, they want to lower prices to attract new subscribers, i.e. to enhance the installed base in the future. Static, marginal-costs pricing is no longer appropriate for modeling this kind of pricing behavior, as it does not take account of this trade-off. Instead, state-space games, in which actions taken in one period shift payoffs in subsequent periods, could be utilized with installed bases of firms as natural state variables. Examples of empirical dynamic pricing models within this framework have been developed in the learning-by-doing literature (e.g. Jarmin, 1994). In these models, cumulative past sales benefit firms in that they lower production costs, what gives rise to a similar trade-off as with network externalities and switching costs.

From the econometric perspective, we do not necessarily need structure for supply relation to be able to correctly estimate the network externalities parameters. Endogeneity problems regarding the price variable can be resolved by instrumental variable technique.

2.2.5 Dynamics of the network good adoption

To get the intuition of dynamics of the network good adoption, a graphical analysis is useful. For simplicity of the presentation we abstract from switching costs and assume the price to be equal across brands, as with perfect competition, and constant over time. We present the common (compatible) network dynamics, so the subscripts i are omitted

⁵ See Klemperer (1995) for an overview of switching costs impact on competition and prices.

throughout this section. In other words, there is one network and one competitive price in this exercise. Because of the symmetry on the supply side, the evolution of the subscriber sets of particular brands is proportional to the common network evolution. To further simplify matters we assume also that the cumulative density function $F(v)$ and the willingness-to-pay function $u(v, x(t-\delta))$ are continuously differentiable in all arguments.

Detailed mathematical treatment of the equilibrium network size path in such model can be found in Cabral (1990). The author proves there that if networks externalities are strong and the lag length δ tends to zero the equilibrium adoption path is unique and discontinuous.

The equilibrium adoption path is described by equation (2.5). It says that the function H maps the network size from time $t-\delta$ to t . Given the assumptions of this section, (2.5) simplifies to

$$x(t) = H(v_t^*) \equiv 1 - F(v_t^*) \quad (2.9)$$

To gain intuition of how network externalities and price affect diffusion we calculate the derivatives of H with respect to the lagged network size $x(t-\delta)$ and price p in appendix 2.5.1. Since H maps the network size from time $t-\delta$ to t , it is convenient to think of it as of a function of the lagged network size $x(t-\delta)$. Examination of (2.25) in appendix 2.5.1 leads to the following lemma

Lemma 2.1: Whenever the solution to equation (2.2) exists, i.e. v_t^* is defined, and the density $f(v_t^*)$ is strictly positive, the extent of network externalities measured by $\frac{\partial u(v_t^*, x(t-\delta))}{\partial x(t-\delta)}$ determine the slope of the function H in the $x(t-\delta)$ domain, such that

- (i) H is non-decreasing if and only if network externalities are non-negative,
- (ii) the slope of H equals zero if there are no network externalities, and
- (iii) the slope of H is larger if network externalities are stronger, other things being equal.

Figure 2.1 illustrates dynamics of the network good adoption. In its upper part we draw H as the function of the lagged network size $x(t-\delta)$. Lemma 2.1 formalizes the link between the extent of network externalities and the slope of H , which in turn determines the dynamics of diffusion. The lower part of figure 2.1 shows the steady-state equilibria of the

network size for each price p , denoted $D(p)$. As it has been already mentioned, it coincides with the analogous static model equilibria with fulfilled consumers' expectations. Dynamic model allows, however, to discriminate among multiple steady-state equilibria. Suppose for the moment that the market price is p^* such as in figure 2.1. Then according to the state equation (2.9) the network size will evolve in the way the upper part of figure 2.1 indicates. If it starts at some size smaller than x_1 it will eventually reach x_0 , if the starting network size is bigger than x_1 it will end up in x_2 . If the network size for some reason equals exactly x_1 , it will stay there, but any arbitrarily small shock will lead to equilibrium at x_0 or x_2 . Therefore we can conclude that x_0 and x_2 are stable steady states, while x_1 is unstable. To apply this way of reasoning to any price p consider the following lemma

Lemma 2.2: Whenever the solution to equation (2.2) exists, i.e. v_t^* is defined, and the density $f(v_t^*)$ is strictly positive, changes in price p determine the shifts of the function H in the $x(t-\delta)$ domain, such that $H(v^*(x(t-\delta), p_1) > H(v^*(x(t-\delta), p_2))$ for every $x(t-\delta)$ if $p_1 < p_2$.

Lemma 2.2 follows directly from examination of (2.28) in appendix 2.5.1. It says that lowering the price shifts the function H up, although it does not need to be a parallel shift. Drawing steady states for each price yields the steady-state demand $D(p)$. We can conclude that

Theorem 2.1: downward-sloping parts of the steady-state demand $D(p)$ consist of stable equilibria, while the upward-sloping parts are unstable, i.e. consist of critical-mass points.

Now, consider a case when price changes over time instead of being a constant. To see how the common network evolves let the price $p(t)$ be a continuous and decreasing function of time and let $p(0) > p_h$ (as in figure 2.1) and $x(p(0))$ be the unique steady-state network size given $p(0)$. As time passes and the price falls, the network size follows the lower steady-state size. Eventually the price reaches p_l and just after that the network size jumps discontinuously

to the higher steady-state and grows further on along it. Formally, this diffusion pattern has been shown to be correct for infinitely small lag δ in Cabral (1990).

If the perception lag is strictly positive, the consumers are myopic with respect to the network size. They do not recognize that the network is going to grow in the current period. As a consequence, the equilibrium network size does not follow exactly but rather tends to the steady-state size. There is no discontinuous jump in the network diffusion either. Instead, the diffusion pattern takes an S-shape.

The dynamic perspective helps to understand the equilibrium selection rule assumed implicitly in our model by the lag structure. It does not allow for coordination among the consumers. Note, that it would be Pareto optimal to jump to the larger steady-state network size before price falls under p_l . However, this would require the coordination of the consumers' subscription decision in order to reach at least the critical mass.

Another insight drawn from the analysis is a sort of substitutability between the extent of network externalities and the lag length, which is described by the following theorem

Theorem 2.2: Network externalities, which extent is measured by $\frac{\partial u(v_t^*, x(t-\delta))}{\partial x(t-\delta)}$

and the perception lag of length δ are substitutes in the sense that both strengthening of network externalities and shortening of the lag length speed up adoption of the network good.

The arrows in the upper part of figure 2.1 indicate the change of the network size from time $t-\delta$ to time t . The length of these arrows reflects the speed of the network size growth (or decline). Now, strengthening the network effects, which implies according to lemma 2.1 a larger slope of the function H , and lowering the lag length (say to $\delta/2$, so between time $t-\delta$ and t there are two “updates” of the network size) one can achieve the same network size growth. In other words, a large extent of network externalities together with a large perception lag may result in the same network diffusion speed as a small extent of externalities and a small lag. One should keep in mind this substitutability when interpreting empirical results. On the other hand, however, manipulating the lag does not influence the steady-state equilibria (the fixed points of the function H), while strengthening of network

externalities does. This observation will be important for the empirical identification of the network effects' strength.

2.3 Stochastic model

2.3.1 Functional specification

The next step towards the structural econometric model is to specify the functional forms in the underlying economic model. This section proposes an example of such specification. It has been chosen because of two reasons. First, the specification yields the demand relation as a simple linear equation (in parameters), which is convenient to work with empirically. Second, the demand relation nests the well-established Bass' (1969) diffusion model.

Assume the consumers' willingness-to-pay function to be

$$u(v, x_i(t-\delta)) = av + bx_i(t-\delta) + cx_i^2(t-\delta), \quad (2.10)$$

where a , b and c are parameters. As before, $x_i(\cdot)$ denotes the network size and v the consumer type. This specification implies that a network of size zero has no other than intrinsic value, which is proportional to the consumer preference parameter v . Network size enters additively into the utility function, which means that consumers are homogenous in their valuation of the network. The square function of the network size catches its non-linear influence on the willingness to pay, e.g. diminishing positive marginal network effect, which is usually assumed in the literature.

The distribution of the individual preference parameters (consumers' types) v is assumed to be uniform on the support $[0,1]$, hence $F(v) = v$ on that support. This distributional assumption corresponds to the linear demand function when the network size is fixed. As pointed out by Economides and Himmelberg (1995), the distribution of types is an a priori assumption, on which the identification of network effects in data critically depends. In that sense the uniform distribution of types is not very fortunate, because it attaches significant proportion of the diffusion S-shape to the network effect arbitrarily. We can,

however, modify the interpretation of the network externalities parameters slightly in order to incorporate some of the distributional effects. Section 2.3.4 discusses that issue more in detail.

Given all the functional assumptions, we calculate the index of the indifferent consumer with respect to each brand i from equation (2.2)

$$v^*(x_i(t-\delta), p_i(t)) = \frac{1}{a} p_i(t) - \frac{b}{a} x_i(t-\delta) - \frac{c}{a} x_i^2(t-\delta). \quad (2.11)$$

The subscription demand of each brand at time t (the state equations) can be obtained from equations (2.3)-(2.5)

$$y_i(t) = \frac{1}{I_{L,t}} \left(1 - \frac{1}{a} p_i(t) + \frac{b}{a} x_i(t-\delta) + \frac{c}{a} x_i^2(t-\delta) \right). \quad (2.12)$$

To get single demand equation (2.12) instead of a switching regime, we assume, that without switching costs the competitive pressure drives the prices down to the marginal costs, so the firms that survive in the market are endowed with the same production technology and set equal prices. This demand relation might be also relevant for the single brand market ($I_{L,t} = 1$). The price then would be of course different from marginal cost.

In the economic model we assumed that there is a measure one of consumers in the market. To be consistent with data we enhance the market to m consumers and call it the market potential⁶. To get actual network size values instead of normalized ones multiply both sides of (2.12) by m

$$Y_i(t) = \frac{1}{I_{L,t}} \left(m - \frac{m}{a} p_i(t) + \frac{b}{a} X_i(t-\delta) + \frac{c}{am} X_i^2(t-\delta) \right), \quad (2.13)$$

where $Y_i(t) = m y_i(t)$ and $X_i(t-\delta) = m x_i(t-\delta)$.

When the switching costs are consider we can still use the indifferent consumer indexes (2.11), but the subscription demand (actual, not normalized) obtained now from (2.3), (2.7) and (2.8) becomes

⁶ The market potential differs in our formulation from the market potential in Bass (1969) in that it does not depend on price.

$$\Delta Y_i(t) = \frac{1}{I_{L,t}} \left(-\frac{m}{a} \Delta p_i(t) + \frac{b}{a} \Delta X_i(t-\delta) + \frac{c}{am} \Delta X_i^2(t-\delta) \right), \quad (2.14)$$

where $\Delta Y_i(t) = Y_i(t) - Y_i(t-\delta)$, $\Delta X_i(t-\delta) = X_i(t-\delta) - X_i(t-2\delta)$ and $\Delta X_i^2(t-\delta) = X_i^2(t-\delta) - X_i^2(t-2\delta)$. Again, to simplify matters we assume that the firms keep equal hedonic (i.e. adjusted for the network size) prices all the time. In other words, they compete for the new subscribers continuously. In principle, it would be also possible that they price low and high interchangeably. So that there would be periods over which one firm attracts the new subscribers and the other extracts a rent from the installed base and periods over which the roles are reversed. Actually, such consecutive pricing pattern is found in Farrell and Shapiro (1988), but it hinges rather on particular assumptions of their model⁷.

2.3.2 Identification

Since data is in discrete time, we need the analogues of (2.13) and (2.14) for estimation purposes. Additionally, we let some stochastic noise enter the equations. This yields

$$Y_{i,t} = \frac{1}{I_{L,t}} \left(\alpha + \beta p_{i,t} + \gamma_1 X_{i,t-1} + \gamma_2 X_{i,t-1}^2 \right) + \xi_{i,t} \quad (2.15)$$

and

$$\Delta Y_{i,t} = \frac{1}{I_{L,t}} \left(\beta \Delta p_{i,t} + \gamma_1 \Delta X_{i,t-1} + \gamma_2 \Delta X_{i,t-1}^2 \right) + \zeta_{i,t} \quad (2.16)$$

respectively, where $Y_{i,t}$ is the discrete analogue of $Y_i(t)$, which is the number of brand i 's customers in time t and $X_{i,t-1}$ is the analogue of $X_i(t-\delta)$, the lagged network size of brand i . $\xi_{i,t}$ and $\zeta_{i,t}$ stand for the error terms and reflect the stochastic noise in the data. Because of the lagged dependent variables the stochastic structure of these equations might be quite complex. We introduce this issue more in detail in section 2.3.3.

⁷ See the discussion in Klemperer (1995).

All four structural parameters a , b , c , and m in (2.13) ($I_{L,t}$ is observable from the market structure) are uniquely identified from the estimates of (2.15). Simple algebra yields the scaling parameter $a = -\alpha/\beta$, the network externalities parameters $b = -\alpha\gamma_1/\beta$ and $c = -\alpha^2\gamma_2/\beta$, as well as the market potential $m = \alpha$.

In the case with switching costs we need some more manipulations. One cannot recover all the structural parameters from the estimates of (2.16) directly. This is because in contrast to (2.15), equation (2.16) is expressed in terms of differences. By differentiating we lose the constant term, so there are only three parameter estimates with four structural parameters to identify. To solve this problem we need to write the sales equation in terms of levels, which yields

$$Y_{i,t} = \alpha_i E_t + \frac{1}{I_{L,t}} (\alpha + \beta p_{i,t} + \gamma_1 X_{i,t-1} + \gamma_2 X_{i,t-1}^2) + \psi_{i,t} \quad (2.17)$$

where α_i is a firm specific constant. E_t is a dummy variable indicating new entry, which is equal to zero in the periods prior to entry and one otherwise. This result is formally derived in appendix 2.5.2 for a general case, i.e. without any functional assumptions. The intuition for this is as follows. Remember, that we have assumed equal hedonic prices among firms each period. Given this assumption, it follows from our economic structure that all active firms attract equal number of new subscribers each period. The only possible source of sustaining differences in total sales is a new entry. Because of the switching costs, the installed base of the incumbent constitutes the competitive advantage over the entrants. This advantage (or disadvantage in case of the entrants) is summarized by the firm specific constants α_i . The formulation in (2.17) indicates that the entry happened only once. It is straightforward to extend it to multiple entries.

Now, we are able to identify all four structural parameters under switching costs as well. The same formulas as before (without switching costs) applied to the estimates of (2.17) yield the desired results. Note, that the interpretation of the parameters in (2.17) is the same as in (2.15). Indeed, as it is shown in appendix 2.5.2, the sum of firm-specific constants equals zero. Switching costs do not influence the market potential or the network effects in our model. They simply allow for persisting asymmetries among firms expressed by nonzero firm specific constants α_i .

Equation (2.17) is convenient also because it nests the two regimes, with and without switching costs. In particular, when the firm specific constants α_i are zero, the sales of the firms are equal, and equation (2.17) boils down to (2.15), which describes the evolution of sales under no switching costs. To get the simple sales equation under switching costs, however, we assumed previously that the prices did not rise. Since we do not need that additional assumption under no switching costs, the validity of equation (2.15) is slightly less restrained. That is why we decided to keep the two cases separately.

Now, let us turn to the question of compatibility of the brands. Our structure allows us to investigate the compatibility, i.e. to check to what extent the network externality operates at the industry level. To test the hypothesis of compatibility empirically, we let

$$X_{i,t-1} = Y_{i,t-1} + wY_{j,t-1}, \quad (2.18)$$

where $Y_{j,t-1} = \sum_{k \neq i} Y_{k,t-1}$ is the sum of all other brands' customers in time $t-1$ and $w \in [0,1]$ measures the degree of compatibility, as described in section 2.2.2. Then (2.17) becomes

$$Y_{i,t} = \alpha_i E_t + \frac{1}{I_{L,t}} \left(\alpha + \beta p_{i,t} + \gamma_1 Y_{i,t-1} + \gamma_{11} Y_{j,t-1} + \gamma_2 Y_{i,t-1}^2 + \gamma_{21} Y_{i,t-1} Y_{j,t-1} + \gamma_{22} Y_{j,t-1}^2 \right) + \psi_{i,t}. \quad (2.19)$$

The identification of the structural parameters a , b , c , and m remains unchanged, since the estimates of α , β , γ_1 , and γ_2 are still available there in (2.19). The new structural parameter w is however overidentified, because there are three new parameters γ_{11} , γ_{21} and γ_{22} in equation (2.19). It can be recovered from $w = \gamma_{11}/\gamma_1$, from $w = \gamma_{21}/2\gamma_2$, and from $w^2 = \gamma_{22}/\gamma_2$. It follows that when the externality operates at the firm level only (incompatible networks, $w = 0$) we expect the estimates of γ_{11} , γ_{21} and γ_{22} to be zero. In the polar case, when the externality operates at the industry level (fully compatible networks, $w = 1$), we expect $\gamma_{11} = \gamma_1$, $\gamma_{22} = \gamma_2$ and $\gamma_{21} = 2\gamma_2$. All the intermediate cases with partial compatibility can be easily obtained from the three equalities as well.

On the one hand, the overidentification of the structural compatibility parameter w gives a scope for a specification test. All three equalities identifying w must hold, otherwise the structural model is not consistent. On the other hand, we could use the overidentification to introduce parameter restrictions and to save on degrees of freedom. For example, recover w

from $w = \gamma_{11}/\gamma_1$ and impose $\gamma_{21} = 2\gamma_{11}\gamma_2/\gamma_1$ and $\gamma_{22} = (\gamma_{11}/\gamma_1)^2\gamma_2$. So we could estimate five instead of seven parameters in (2.19).

Last, as mentioned at the beginning of this section our structure corresponds to the information diffusion models widely studied in the marketing science. In particular, the equation (2.19) nests the original diffusion equation proposed by Bass (1969) for the single product case. When we consider single brand diffusion, (2.19) simplifies to the original Bass model if $\beta = 0$ (i.e. price does not matter for the network diffusion).

2.3.3 Stochastic structure

The final step in the structural econometric model is the introduction of stochastic structure. So far, we have not imposed any assumptions on the error terms in (2.15), (2.16), (2.17), and (2.19). We have not proposed any estimation technique either. This is because it may be far less trivial than the simple, linear in parameters functional form of these equations suggests.

To illustrate the potential econometric pitfalls let us consider the market with network externalities operating at the industry level (fully compatible networks) and two competing brands. The equation (2.19) becomes then

$$Y_{i,t} = \frac{1}{2}(\alpha + \beta p_{i,t} + \gamma_1(Y_{i,t-1} + Y_{j,t-1}) + \gamma_2(Y_{i,t-1} + Y_{j,t-1})^2) + \psi_{i,t}. \quad (2.20)$$

The most obvious way the stochastic noise can enter the empirical relation is a measurement error in the dependent variable. This makes sense in our model, since the price is usually easily observable by an econometrician, while the network size might be not. Suppose, that we observe the brand sales with some noise $Y_{i,t} + \varepsilon_{i,t}$, where $\varepsilon_{i,t}$ is some i.i.d. measurement error. In the estimation we put the observation with noise into equation (2.20) creating the error term of the form

$$\psi_{i,t} = -\varepsilon_{i,t} + \frac{\gamma_1}{2}(\varepsilon_{i,t-1} + \varepsilon_{j,t-1}) + \frac{\gamma_2}{2}(\varepsilon_{i,t-1} + \varepsilon_{j,t-1})^2 + \gamma_2(Y_{i,t-1} + Y_{j,t-1})(\varepsilon_{i,t-1} + \varepsilon_{j,t-1}), \quad (21)$$

which is clearly not an i.i.d. error. The error structure (2.21) received some attention in the time series econometrics⁸. The second and third term on the RHS of (2.21) indicate a multivariate nonlinear moving average. The fourth term points to a multivariate bilinear process. To correctly estimate the structural parameters of this model one needs to take care of the error generating process that is consistent with the assumed structure. Good news is that our economic model gives rise not only to the equations, but to a particular error structure in the econometric model as well. As a consequence, we do not need to rely solely on the statistical procedures to choose the appropriate error structure.

2.3.4 Interpretation of the identified structural parameters

Interpretation of the identified structural network externalities parameters b and c directly is difficult because of two reasons. First, as indicated already in section 2.3.1, the empirical identification of the network effects relies heavily on the functional assumptions. In particular, the distribution of types plays a key role. Another assumption that influences the estimates of the network effects is the consumer perception lag δ that we impose by choosing data frequency. Second, even if we have statistically significant and correct estimates of b and c we still miss some threshold, which tells us which values of the parameters correspond to the economically significant network effects.

Going back to the first problem, our empirical estimates of the network effects can be biased because of the functional assumptions. In particular the uniform distribution of types is likely to bias the network estimates upward, i.e. to attach significant proportion of the diffusion S-shape to the network effect arbitrarily. The natural assumption is that the distribution of types mimics the distribution of consumer income, which is usually log-normal⁹. Section 2.2.5 on the dynamics of network growth helps to understand how any bell-shaped distribution of types contributes to the S-shape of the diffusion curve.

In the case of the perception lag the direction of bias is less clear. In general, we can expect that imposing larger (smaller) lag than the actual one creates an upward (downward) bias. But, the question, how large the actual lag is, remains open. The common sense just tells us that using monthly data (hence the one month perception lag) is more appropriate than

⁸ See Granger and Teräsvirta (1993) for a nice overview of the nonlinear time series analysis.

⁹ Economides and Himmelberg (1995) study the distribution of consumer income in depth to obtain more reliable estimates of network effects.

using yearly data. In section 2.2.5 we formalized the intuitive relationship between the lag length and the strength of network effects. We noted also that the steady-state equilibria are not affected by the lag manipulations.

Being aware of the possible bias in our estimates, how can we infer the economic significance of the identified network effects parameters? We propose to calculate the steady-state inverse demand functions from (2.19) replacing all the parameters with their empirical estimates and imposing steady-state conditions (2.6). All the important economic phenomena driven by the network effects, like multiple equilibria and critical mass of adopters, apply to the case with upward sloping demand. Therefore, the existence of the upward sloping part in the empirical steady-state demand function indicates strong network externalities.

The empirical steady-state demand function seems also more robust to the improper functional assumptions than the identified structural parameters themselves. First, the steady-state equilibria are not affected by the lag manipulations. And second, the intuition suggests that attaching some distribution-of-types effects to the network effects should not change dramatically the shape of the steady-state demand function. Therefore, it seems unlikely that we obtain an upward sloping part of the demand function in the estimation, while in fact it is all along downward sloping in the given market.

Some additional information about the source of the network effect in the market under consideration can be obtained from the estimate of the compatibility parameter w . For example, the ability to satisfy more communication needs with the bigger consumers' pool may give rise to direct network externalities, which operate at the industry level (compatible networks) in telecommunication markets. Whereas endogenous externalities as in Blonski (2002), which are created by firms charging an access fee for the calls from outside into their networks, operate at the firm level (incompatible brands).

Last, the estimate of the market potential parameter m may serve as another specification test. Usually, we have a rough guess of the total number of consumers, which could potentially subscribe to a network. Outstanding values of m signal problems with the estimates.

2.4 Conclusions

This paper introduces a structural econometric model of consumer demand for non-durable goods exhibiting network externalities. Its main contribution is that it allows us to recover the structural parameters capturing the extent of externalities. The structural parameters' estimates can be in turn employed to test the economic significance of the externalities and the compatibility of competing networks.

The identifying assumption that drives our results is that the consumers care about the lagged instead of the current network size in their subscription decision. As Cabral (1990) we argue that when the lag is infinitely small this behavior is rational. It does not allow only for coordination among the consumers.

Empirical implementation of the model leads, however, to a bigger than infinitely small lag, because of the data frequency. Then, if we interpret the economic models in terms of direct network externality, consumers are myopic with respect to the network size. In other words, they do not recognize that during the diffusion process the network grows in current period. Instead, they use the previous period network size in their purchase (subscription) decision. The “myopic” diffusion is slower and smoother (does not exhibit discontinuous jump) than “rational” diffusion.

On the other hand, the installed base of users could matter for the reasons other than the direct network externalities. For example, in the experience good case the installed base could transmit the information about the quality of the good enhancing the willingness to pay for it. In this case the lag in the network size has a natural explanation, since the transmission of information takes time. This reinterpretation of the network effect preserves the rationality of consumers.

Our structure gives also important insights for interpretation of the empirical results. First, as mentioned already in Economides and Himmelberg (1995) the distribution of types is an important a priori assumption that the identification of network effects in data relies on. Second, the lag induced naturally by data frequency influences the empirically identified network effects too. The strength of the network effects and the lag length are in a sense substitutes in generating the diffusion S-shape.

We provide an example of functional specification that yields a simple linear stochastic model of demand. This demand model nests the original Bass' (1969) model of innovation diffusion. Using the economic structure and the functional specification we are able to identify all structural parameters of the model. Interpreting the parameters correctly, we can still investigate the economic significance of the identified network effects.

Last, but not least, we brought stochastic structure into the model. Introducing a measurement error, as the most obvious source of stochastic noise in data, results in a non-trivial error structure in our econometric model. To correctly estimate the structural parameters of the model one needs to take care of the error structure.

2.5 Appendices

2.5.1 Derivatives of the function H with respect to $x(t-\delta)$ and p

First, note that v_t^* is an implicit function of $x(t-\delta)$ and p , what under simplifying assumptions in section 2.2.5 is described by

$$u(v_t^*, x(t-\delta)) = p. \quad (2.22)$$

To calculate the derivative of H with respect to the lagged network size $x(t-\delta)$ we first apply the chain rule to the definition of H given in (2.9). We obtain

$$\frac{\partial}{\partial x(t-\delta)} H(v_t^*) = -\frac{\partial F(v_t^*)}{\partial v_t^*} \cdot \frac{\partial v_t^*}{\partial x(t-\delta)} \quad (2.23)$$

The first term on the RHS of (2.23) is just the density of v at v_t^* . To calculate the second term note that the total derivative of $u(v_t^*, x(t-\delta))$ with respect to $x(t-\delta)$ must stay constant in order to satisfy equation (2.22). This holds for

$$\frac{\partial u(v_t^*, x(t-\delta))}{\partial x(t-\delta)} = -\frac{\partial u(v_t^*, x(t-\delta))}{\partial v_t^*} \cdot \frac{\partial v_t^*}{\partial x(t-\delta)} \quad (2.24)$$

Solving (2.24) for $\frac{\partial v_t^*}{\partial x(t-\delta)}$ and substituting in (2.23) yields the result

$$\frac{\partial}{\partial x(t-\delta)} H(v_t^*) = f(v_t^*) \cdot \left(\frac{\partial u(v_t^*, x(t-\delta))}{\partial v_t^*} \right)^{-1} \cdot \frac{\partial u(v_t^*, x(t-\delta))}{\partial x(t-\delta)}, \quad (2.25)$$

where f is a density function of v .

Analogously, to calculate the derivative of H with respect to the price p we first apply the chain rule to obtain

$$\frac{\partial}{\partial p} H(v_t^*) = -\frac{\partial F(v_t^*)}{\partial v_t^*} \cdot \frac{\partial v_t^*}{\partial p}. \quad (2.26)$$

Then we note that

$$\frac{\partial u(v_t^*, x(t-\delta))}{\partial v_t^*} \cdot \frac{\partial v_t^*}{\partial p} = 1, \quad (2.27)$$

and substitute to get

$$\frac{\partial}{\partial p} H(v_t^*) = -f(v_t^*) \cdot \left(\frac{\partial u(v_t^*, x(t-\delta))}{\partial v_t^*} \right)^{-1}. \quad (2.28)$$

2.5.2 State equations with firm-specific constants

To see how we can nest the two regimes (with and without switching costs) in a single set of the state equations rewrite (2.8) using definitions (2.4) and (2.7) to

$$y_i(t) = H_i(\mathbf{v}_t^*) + y_i(t-\delta) - H_i(\mathbf{v}_{t-\delta}^*) \frac{I_{L,t-\delta}}{I_{L,t}}. \quad (2.29)$$

Remember that we assume equal hedonic prices among firms. One can think of (2.29) as of a decomposition of the total sales of brand i in time t under switching costs. The first term on the RHS of (2.29) gives the total sales of brand i (number of subscribers) if there were no switching costs. The second and the third term adds and subtracts the installed base of brand i respectively in a way that is sensitive to the number of active firms in the market. To see how this can lead to persistent asymmetries among firms expand recursive equation (2.29) to

$$\begin{aligned} y_i(T) &= H_i(\mathbf{v}_T^*) + H_i(\mathbf{v}_{T-\delta}^*) - H_i(\mathbf{v}_{T-\delta}^*) \frac{I_{L,T-\delta}}{I_{L,T}} + H_i(\mathbf{v}_{T-2\delta}^*) - H_i(\mathbf{v}_{T-2\delta}^*) \frac{I_{L,T-2\delta}}{I_{L,T-\delta}} + \\ &+ \dots + y_i(0) - H_i(\mathbf{v}_0^*) \frac{I_{L,0}}{I_{L,\delta}}, \end{aligned} \quad (2.30)$$

where $t = 0$ indicates the time when the market starts up so there are no sales at that time and $T > 0$.

Suppose, there is constant number of firms active in the market such that $I_{L,t} = I_L$ for $t \in (0, T)$. Then the last two terms on the RHS of (2.30) equal zero, because every firm is active from the very beginning of the market, and all the middle terms cancel out. In this case

(2.30) simplifies to (2.5), i.e. the state equations with and without switching costs are the same.

Now suppose, there was an entry into the market in $t = E$, and $0 < E < T$. This means that $I_{L,t}$ rises discontinuously in $t = E$ and stays at the higher level afterwards. The sales equations of the incumbents do not simplify to (2.5) any longer. They become instead

$$y_i^{inc}(T) = H_i(\mathbf{v}_T^*) + \int_E^{E+\delta} [H_i(\mathbf{v}_{t-\delta}^*) - H_i(\mathbf{v}_{t-\delta}^*) \frac{I_{L,t-\delta}}{I_{L,t}}] dt, \quad (2.31)$$

for $T \geq E+\delta$. The integral in (2.31) is positive. It is also invariant with respect to any events in $T > E+\delta$ and can be treated therefor as a firm-specific constant in the post-entry period.

In contrast, the expansion of the recursive equation (2.29) does not go back to $t = 0$ for the entrants. Their history starts at $t = E$ and the sales can be described by

$$y_i^{ent}(T) = H_i(\mathbf{v}_T^*) + \int_E^{E+\delta} [-H_i(\mathbf{v}_{t-\delta}^*) \frac{I_{L,t-\delta}}{I_{L,t}}] dt, \quad (2.32)$$

for $T \geq E+\delta$. To see this result, refer to (2.29) and note that $y_i^{ent}(t-\delta) = 0$ for $t \in (E, E+\delta)$. The integral in (2.32) plays analogous role for the entrants as the integral in (2.31) for incumbents, but it is negative. Therefore, we can conclude that the incumbents have a constant (in terms of the difference in the total sales) competitive advantage over the entrants.

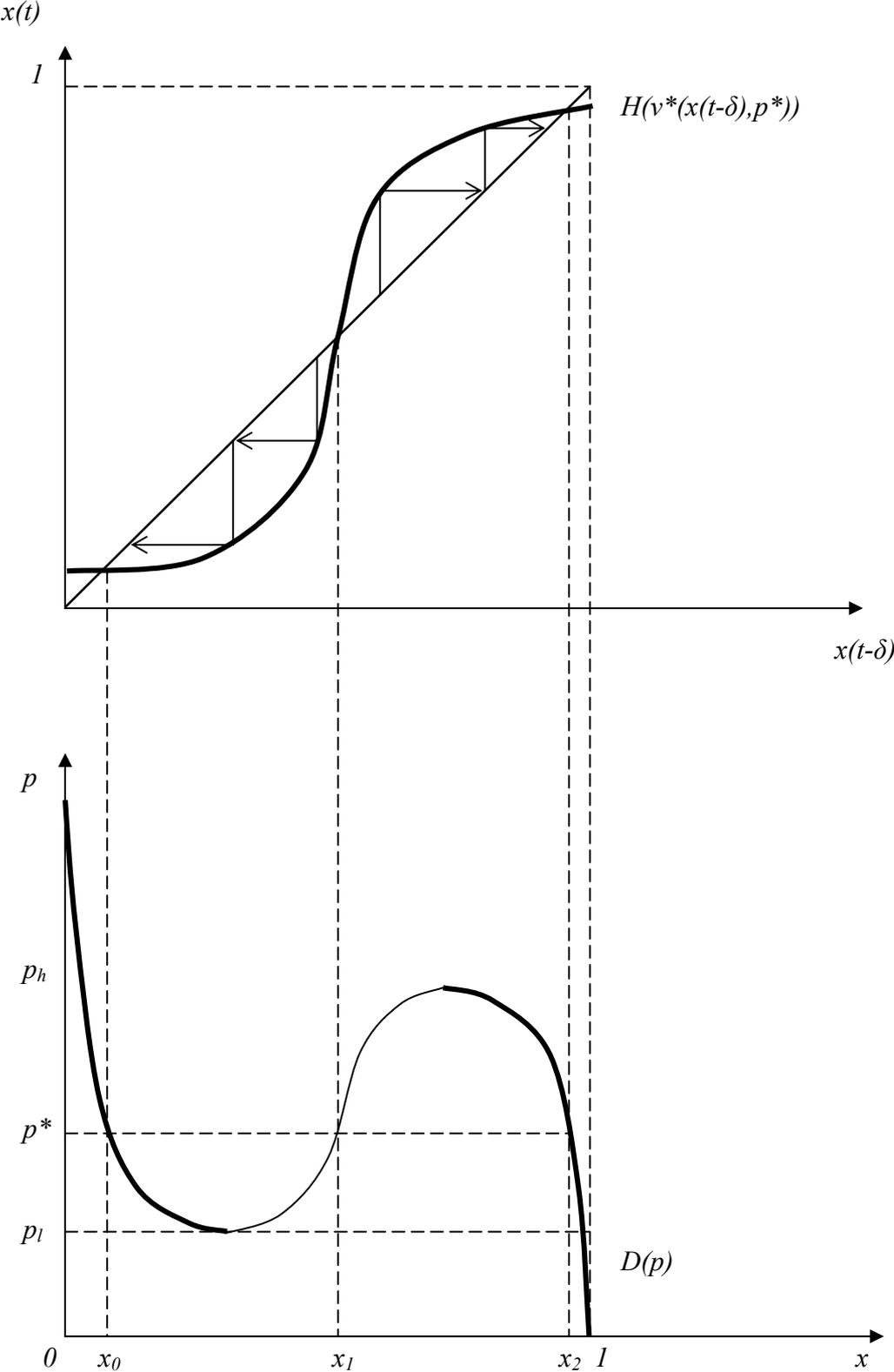
Moreover one can show that the the fixed effects caused by entry sum up to zero. To see that denote the number of incumbents as A and the number of entrants as B . The sum of the effects is then

$$\begin{aligned} & A \int_E^{E+\delta} [H_i(\mathbf{v}_{t-\delta}^*) - H_i(\mathbf{v}_{t-\delta}^*) \frac{I_{L,t-\delta}}{I_{L,t}}] dt + B \int_E^{E+\delta} [-H_i(\mathbf{v}_{t-\delta}^*) \frac{I_{L,t-\delta}}{I_{L,t}}] dt = \\ & = A \int_E^{E+\delta} [H_i(\mathbf{v}_{t-\delta}^*) - H_i(\mathbf{v}_{t-\delta}^*) \frac{A}{A+B}] dt + B \int_E^{E+\delta} [-H_i(\mathbf{v}_{t-\delta}^*) \frac{A}{A+B}] dt = \\ & = \left(A - \frac{A^2}{A+B} - \frac{AB}{A+B} \right) \int_E^{E+\delta} H_i(\mathbf{v}_{t-\delta}^*) dt = 0. \end{aligned} \quad (2.33)$$

One could also investigate the effects of exit in the analogous manner. Since in our economic structure there is no reason for a firm to leave the market we skip this discussion.

2.5.3 Figures

Figure 2.1 Stable vs. unstable equilibria



Chapter 3

Estimating Network Effects and Compatibility in Mobile Telecommunications

3.1 Introduction

Economic literature showed that consumption network effects could have important implications for size and structure of the underlying market, as well as for its competitiveness.¹ A good is characterized by network effects when an increase in the number of users of the good increases its value to other users. Although, the concept gave rise to numerous theoretical developments, it still remains largely unexplored in the empirical literature. This paper aims to fill this gap by providing empirical evidence of the extent of network effects from mobile telecommunications. We specify and estimate a structural model of consumer demand for mobile telephone service, which allows us to identify the extent of network effects and compatibility among competing networks. We emphasize the fact that structural approach, as opposed to descriptive approach, gives a scope for testing economic significance of the effects.

The concept of compatibility is closely linked to that of network effects. In the economic context, we might define compatibility as a measure, which says to what extent utility derived by users of a given network good is influenced by the number of users of competing network goods. As such, it is a very important characteristic of network industries. For instance, it greatly influences the nature of competition. We expect incompatible networks to aggressively compete “for the market”, whereas compatible networks compete more traditionally “within the market”, because the threat of tipping is reduced (Besen and

¹ See e.g. Katz and Shapiro (1994) for a general overview of the properties of markets fetured by network effects .

Farrell, 1994). To the best of our knowledge, this is the first study that assesses compatibility within the econometric framework.

It is important to stress that technological compatibility is not a sufficient condition for economic compatibility defined above. Although, the mobile telephone networks that we study are fully interconnected, it is not clear *a priori* whether they are compatible or not. One reason for incompatibility involves intra-network discounts, which enhance relative attractiveness of own network to its subscribers. In the next section, we put forward several possible sources of network effects in mobile telecommunications. Some of them imply more compatible, while the other more incompatible networks. An estimate of compatibility allows us therefore additionally to discriminate among various possible sources of network effects.

The empirical literature on mobile telecommunications concentrates on determinants of growth and competitiveness of the industry neglecting in general network effects (e.g. Parker and Röller, 1997; Ahn and Lee, 1999; Gruber and Verboven, 2001). In the context of fixed-line telecommunications, the study by Bousquet and Ivaldi (1997) is probably the first one, which tests empirically for existence of network effects. In contrast to our study, which focuses on access to telephone service, they concentrate on usage, which seems more relevant in a saturated market that they consider. Consequently, the concept of network externality they use relies on received calls, which benefit subscribers without paying for them, rather than on installed base of subscribers. Next, Okada and Hatta (1999) specify demand for fixed-line and mobile telephone service adopting Almost Ideal Demand System. They show that the number of mobile subscribers, as a quality measure for telephone service, has significant positive effect on share of telecommunications' expenditures – both mobile and fixed-line – in households' budgets. This result is an empirical evidence for network effects in demand for telephone service. More recently, Kim and Kwon (2003) conduct a conditional logit analysis based on a consumer survey. The analysis reveals that consumers prefer mobile service providers with larger number of subscribers other things being equal. The authors attribute this size effect, which is in line with network effects operating at the firm level we found, to intra-network call discounts and quality signaling effect. Finally, Gruber and Verboven (2001) aim to identify factors responsible for timing and speed of mobile service adoption diffusion. In contrast to our study in which the diffusion's S-shape is driven by network effects they take it as given using a logistic model of diffusion.

The approach we follow draws on the model introduced in chapter 2. The main idea facilitating identification of network effects is that strong network effects might give rise to an

S-shaped diffusion of adoptions. The network effects in our economic framework are captured by the dependence of consumer willingness to pay on installed base of subscribers. Together with price changes, the installed base determines speed of diffusion of mobile telephone service in a population. Estimating structural state equations describing evolution of subscriptions, we are able to disentangle the price effect from the network effects.

We estimate the model using quarterly panel data from the Polish mobile telephone industry for the period 1996-2001. We find strong network effects, which give rise to an upward sloping part in the inverse demand function. The estimated degree of compatibility is surprisingly low. This suggests that subscribers to a given mobile network attach relatively little value to competing networks' subscribers. Simulations of the estimated model with alternative values of network effect and compatibility parameters illustrate their economic significance. We also estimate a restricted model, which does not account for network externalities. Estimated price elasticity of demand is much higher (in absolute terms) in this model, since it falsely attribute changes in demand due to network effects to changes in price. As we also formally show, ignoring network effects in empirical models of emerging network industries can substantially bias downward the estimated elasticity of demand.

The rest of the paper is organized as follows. Section 3.2 proposes several possible sources of network effects in mobile telecommunications and describes the Polish mobile telephone industry. In section 3.3, we specify the model and show that all structural parameters are identified from regression coefficients. Data, estimation issues, and interpretation of the empirical results are discussed in section 3.4. Section 3.5 concludes.

3.2 Mobile telecommunications

3.2.1 Network effects

There are several possible sources of network effects in mobile telecommunications. First, consider a "traditional", direct effect, like in fixed-line telecommunications. Consumers value the installed base of subscribers, because they can satisfy more communication needs. Since installed base of fixed-line subscribers had been already huge when mobile service emerged and mobile customers can call the stationary numbers, it is not clear whether significant network effects arise because of an additional telephone service, which mobile

telecommunications offer. However, short message service (SMS) – available only within mobile network – might help to generate the “traditional” network effects.

A second possible source of network effects is a need of people to buy, consume, and behave like their fellows, which induces bandwagon effect as in Leibenstein (1950). Economic consequences of this desire to join the crowd, which stems from social interactions were also studied in more recent economic literature (e.g. Granovetter and Soong, 1986; Becker, 1991; Lindbeck et al., 1999; Schoder, 2000). We expect that consumption of mobile telephone service is influenced to some extent by such conformist behavior since mobile telecommunications are clearly important media of social interactions.

Intra-network call discounts offer another explanation for network effects in mobile telecommunications. The reason is very intuitive. Assume that customers apply a balanced calling pattern, which means that the number of calls terminating in each network is proportional to their relative size. Since – because of the discounts – intra-network calls are cheaper than inter-network calls, larger installed base of subscribers to a given provider implies *ceteris paribus* lower monthly bill for them. The resulting effect is called endogenous network externality in Blonski (2002). We could extend this argument incorporating fixed-line network into the analysis. Given, that it is cheaper to call a mobile number from mobile network than from stationary network, larger mobile network implies - as before - lower monthly bill, hence higher attractiveness of mobile telephone service in general. By considering different reference network than Blonski (2002), we obtain network effects operating at industry level and not at firm level.

We might also expect that quality of mobile telephone service is *a priori* unknown to consumers. They could learn about the quality from other consumers, who have already subscribed for the service. The installed base of customers would then transmit information to unattached consumers influencing their willingness to pay. Network effects arise in this case, either if the chance of receiving information increases with the size of installed base, or if the network size carries a quality signal.

Additional knowledge about the nature of network effects gives us the extent of compatibility among competing networks, i.e. the level at which the effects operate. The two polar cases are firm-specific network effects on the one hand and industry-specific network effects on the other hand. They correspond to full incompatibility and full compatibility respectively. We expect the “traditional” effects to be industry specific. In contrast, the effects raised by information transmission are rather firm specific. Whether the endogenous effects

are firm or industry specific or both depends on relative prices of mobile intra-network calls, mobile inter-network calls, and fixed-line to mobile calls. The social interaction concept could support both types of network effects, depending on consumers' reference group.

3.2.2 Description of the market

Our empirical investigation regards the second-generation mobile telephone industry that was launched in September 1996 in Poland. The Ministry of Telecommunications (MT) – regulator of that industry – granted initially two licenses for providing the mobile telephone service based on GSM 900 standard. To further intensify competition in the industry, the MT offered a GSM 1800 license to a third provider, which started activity in March 1998. The difference between the two standards – GSM 900 and GSM 1800 – is that the latter operates at a higher frequency allowing providers to serve more connections per unit of area at the same time. GSM 1800 is also more costly to install than GSM 900, since it requires higher density of cellular antenna sites. In practice, because of relatively low density of population in Poland, the service based on GSM 1800 standard was offered exclusively in large cities. The entrant became a countrywide provider in March 2000 only after being granted additionally GSM 900 license. Soon after that, the two incumbents obtained GSM 1800 licenses, and started to offer double-frequency service as well. Licensing remained the only way of regulating the industry.

GSM telecommunications turned out to be very successful in Poland unlike its predecessor based on analog technology standard NMT 450i. The analog service – called also the first-generation mobile telecommunications – was introduced in 1992 and reached the size of around 200 thousand subscribers at the time the second-generation service was introduced in 1996. In contrast, four years after the launch the total number of GSM subscribers amounted to over 3 million and after the next two years it approached 10 million, which roughly correspond to 25% of the total population. Obviously, there are differences in the quality between the analogue NMT and the digital GSM standards, but the extreme market outcomes in terms of network size could indicate presence of network effects.

Demand for mobile telephone service is also likely to be shaped by switching costs. In fact, lack of number portability and long-term contracts, which included a penalty for premature cancellation – a common practice during the period studied – contributed to these costs substantially. The rationale behind keeping high switching costs is twofold. First, they

might enable providers to encourage subscriptions by crediting the handsets, which are necessary for using the mobile telephone service. Second, they might strategically weaken price competition by decreasing cross-price elasticity of demand.² In an extreme case with switching costs high enough, providers compete for the unattached consumers only. Our empirical framework allows us to account for switching costs to some extent. In particular, it follows from our economic model that the entrants will never catch up with the incumbents in terms of the number of subscribers when switching costs are significant. The unwilling-to-switch installed base of incumbents will cause persistent asymmetry. Testing size of the asymmetry empirically is an indirect way to test the significance of switching costs.

Another important feature of emerging mobile telephone industry is rapidly growing coverage, which determines availability and quality of the service. Clearly, it is not reasonable to subscribe to a mobile network unless it covers one's place of residence. Coverage adds also to the quality of the service, since subscribers can use it while traveling. Data on diffusion of the coverage in the Polish mobile telecommunications we have is not detailed enough to use it explicitly in our analysis. Some observations, however, facilitate assumptions of our model. First, the coverage grew way faster than the installed base of subscribers. Already one year after the GSM telecommunications industry had been launched, the coverage reached 60% of the total population, while subscribers amounted just to 1,3%. After two years, the corresponding numbers were 80% and 3%. Second, coverage growth across providers followed basically the same pattern. By ignoring coverage in our analysis, we implicitly assume that the availability constraint does not bind and that consumers do not realize the quality differences among providers due to coverage.

3.3 A model of demand for mobile telephone service

In order to estimate network effects in demand for mobile telephone service, we focus on access to rather than on usage of the service. Rapidly growing installed base of subscribers to mobile networks justifies this approach. One could potentially extend the analysis allowing subscribers to choose usage level as well. This would, however, complicate matters a lot without being crucial to the identification of network effects. The remaining part of this

² This result is discussed in Beggs and Klemperer (1992) and Klemperer (1995).

section outlines the model, which guides our empirical analysis. It adopts basically the structural econometric model proposed in chapter 2.

We assume that there are several mobile telephone networks in a market. Each of them is operated by single provider and associated with a brand denoted by $i = 1, 2, \dots, I(t)$. The number of providers/brands $I(t)$ can vary over time, as reflected by the argument t . Consumer utility derived from subscribing for the service is assumed to depend on private intrinsic value attached to it and, at the same time, on the size of installed base of subscribers. To facilitate empirical analysis, instantaneous willingness to pay of each consumer is specified by

$$u(v, x_i(t-\delta)) = v + cx_i(t-\delta) + dx_i^2(t-\delta), \quad (3.1)$$

where consumer type v represents his/her private valuation of the service, $x_i(t-\delta)$ denotes the lagged network size of brand i , and c and d are parameters determining the extent of network effects. Since v is brand independent, consumers have no taste for a particular brand. We assume that v is uniformly distributed over $(-\infty, a]$ with some density $b > 0$.³ We could equally well introduce a lower bound of the distribution support to limit the population of consumers. We do not do it to avoid the necessity of considering corner solutions, when all consumers subscribe. It is also important to note that network size of brand i is not necessarily equal to its number of subscribers, as it depends on compatibility of networks. Here, compatibility is meant as a measure, which says to what extent subscribers to a given brand value the installed base of other brands. To avoid confusion, we stress that compatibility is not simply implied by interconnection of brands. It depends crucially on the nature of network effects at work, as it has been discussed in section 3.2.1. The relationship between network of a brand and its number of subscribers is expressed by

$$x_i(t) = y_i(t) + wy_{-i}(t), \quad (3.2)$$

where $y_i(t)$ denotes number of subscribers to brand i at time t , $y_{-i}(t) = \sum_{j \neq i} y_j(t)$, and $w \in [0, 1]$ denotes compatibility among brands. $w = 1$ and $w = 0$ correspond to the perfect compatibility and perfect incompatibility respectively and the interior values of w indicate a partial compatibility. For simplicity, the compatibility is denoted by a constant in (3.2), but it is

³ Here, we do not mean a probability distribution, so the integral over the support of density function does not need to equal 1. In fact, the integral equals ∞ , when the lower bound of the support equals $-\infty$, which means that there is infinite number of consumers in the market.

straightforward to condition it on time and/or brand. Actually, we do it at the end of this section to account for changes in intra-network discounts offered by the providers.

Since the focus of this paper is identification of network effects, we do not explicitly model supply side of the market. In order to make our demand analysis tractable, however, we need to restrict pricing behavior of the providers. In particular, we assume that competition in the mobile telephone industry results in setting equal hedonic prices across brands over time.⁴ This assumption seems natural, as consumers' preferences are not brand specific. As a consequence, in each instance of time consumers are indifferent toward brands. Therefore, they do not need to consider potentially costly switching among mobile service providers. Another assumption we make is that providers can credibly commit not to raise their prices.⁵ This assures that mobile service users do never tempt to relinquish subscription. Frequent relinquishing and reopening of subscription might add substantial costs that we do not model in the consumer decision-making.

If consumers correctly foresee that prices will not rise and that competing brands will be equally attractive, their decision to subscribe or not will be fairly simple. They will subscribe when their private valuation of the service exceeds its current hedonic price. In order to determine the number of subscribers, we first calculate the types of indifferent consumers with respect to each brand i from

$$u(v_{i,t}^*, x_i(t-\delta)) = p_i(t), \quad (3.3)$$

where $p_i(t)$ stands for the price of brand i at time t . Assumption of equal hedonic prices assures that the types of indifferent consumers $v_{i,t}^*$ are equal across the brands. All higher-type consumers than $v_{i,t}^*$ subscribe to the service – or continue subscription – at time t . As a tie-breaking rule, we assume that they choose among the brands with equal probability. Given willingness-to-pay function (3.1) and the assumed uniform distribution of types, the total demand for mobile telephone service at time t is then given by⁶

$$\sum_i y_i(t) = ab - bp_i(t) + bcx_i(t - \delta) + bdx_i^2(t - \delta). \quad (3.4)$$

⁴ We use the term hedonic price to depict price corrected for network effects analogously to price corrected for quality, in which context the term hedonic price is usually used.

⁵ The option to cancel subscription free of charge in the case of price increase could serve as a commitment device.

⁶ For details of the derivation see chapter 2.

In case all providers start their activity simultaneously, our model predicts that they will have equal market shares, so the number of subscribers to each brand i is obtained dividing the RHS of (3.4) by the number of brands. Entry may break this symmetry given that significant switching costs are present. In general, we can write the demand for single brand at time t as

$$y_i(t) = l_i E(t) + \frac{1}{I(t)} \left(ab - bp_i(t) + bcx_i(t - \delta) + bdx_i^2(t - \delta) \right), \quad (3.5)$$

where l_i are firm specific constants, $E(t)$ is a dummy variable, which is equal to zero in the periods prior to entry and one otherwise, and $I(t)$ equals the number of providers in the market at time t . The actual values of l_i can be calculated from the size of incumbents' installed base in the entry period. Under switching costs incumbents have an advantage over entrants, which is embodied in positive values of l_i for the former and negative l_i for the latter. Under no switching costs, l_i equals zero for each provider. In any case, the sum of the firm specific constants equals zero, so (3.4) still holds.⁷

Finally, to obtain equation, which we are going to estimate, we put (3.2) into (3.5), then we take a discrete analog of the resulting expression, and we add an error term. This yields

$$y_{i,t} = \lambda_i E_t + \frac{1}{I_t} \left(\alpha + \beta p_{i,t} + \gamma_1 y_{i,t-1} + \gamma_2 y_{i,t-1}^2 + \gamma_{11} y_{-i,t-1} + \gamma_{21} y_{i,t-1} y_{-i,t-1} + \gamma_{22} y_{-i,t-1}^2 \right) + \psi_{i,t}. \quad (3.6)$$

Possible sources of the error term $\psi_{i,t}$ and its properties will be discussed together with estimation issues in section 3.4.2. Once we correctly estimate (3.6), all structural parameters of the model are identified. Simple algebra yields the highest consumer type in the population $a = -\alpha/\beta$ and density of the distribution of types $b = -\beta$. We can interpret also the parameter $\alpha = ab$ as the number of consumers in the population with positive valuation of the mobile telephone service given zero network size. Network effects parameters c and d can be obtained as $-\gamma_1/\beta$ and $-\gamma_2/\beta$ respectively. The compatibility parameter w is overidentified, since it can be recovered as γ_{11}/γ_1 , as $\gamma_{21}/2\gamma_2$, and as $\pm \sqrt{\gamma_{22}/\gamma_2}$. It follows that when the externality operates at the firm level only (incompatible networks, $w = 0$) we expect the estimates of γ_{11} , γ_{21} and γ_{22} to be zero. In the polar case, when the externality operates at the industry level (fully compatible networks, $w = 1$), we expect $\gamma_{11} = \gamma_1$, $\gamma_{22} = \gamma_2$ and $\gamma_{21} = 2\gamma_2$. All

⁷ Ibid.

the intermediate cases with partial compatibility can be easily obtained from the three identifying equalities as well.

The overidentification of the compatibility parameter w gives a scope for a specification test. All three equations identifying w must hold, otherwise either the estimates are not correct, or the data reject the structure. Alternatively, in order to gain in efficiency of estimators, we might identify w from just one relation and build parameter restrictions from the other two.

We could also allow w to vary over time. This might be important in the Polish mobile telephone industry, since providers changed their pricing strategy in the period studied. Initially, the incumbents favored their own subscribers including intra-network call discounts in price schedules. We call it discriminatory pricing, as opposed to non-discriminatory pricing, which does not involve intra-network discounts. The entrant used non-discriminatory pricing from the very beginning and one of the incumbents switched from discriminatory to non-discriminatory regime after the entrant took off. According to endogenous network effects hypothesis described in section 3.2.1, this could increase compatibility of mobile telephone networks. To capture this effect we introduce intersections of the regression

variables $\frac{y_{-i,t-1}}{I_t}$, $\frac{y_{i,t-1}y_{-i,t-1}}{I_t}$, and $\frac{y_{-i,t-1}^2}{I_t}$ with firm-specific pricing dummies $D_{i,t}$ into (3.6).

The pricing dummies $D_{i,t}$ take value of zero in the periods prior to introduction of the “non-discriminatory” pricing by i -th provider and one otherwise. The change in compatibility of networks, denoted by Δw , could be identified from estimates of coefficients on the new intersection variables. The identification of Δw is analogous to that of w .

3.4 Empirical evidence

3.4.1 Data

The firm-level, quarterly panel data used in this study covers prices and number of subscribers in the Polish mobile telephone industry over 1996-2001. The information on prices has been obtained from the Polish Ministry of Telecommunication (MT). It includes all price plans of each provider. In the period studied, all providers used non-linear pricing in the form of multiple price plans, from which the customers self-select the most favorable one. A

plan consists of a monthly fixed charge, a price per minute of usage and often includes some minutes free of charge. A usage price is further diversified according to the time of the day (peak/off-peak hours) and to the termination network (intra-network discounts). Following Parker and Röller (1997), we define the price of a single mobile subscriber as the best-deal monthly bill paid for a given constant calling pattern. That is, we assume some calling pattern (including the overall monthly usage and the proportions of the calls in various times of the day and to various networks) and calculate the monthly bills for all available price plans of each provider.⁸ The single price of each provider is then his lowest bill that reflects the best-deal selection of an average customer. Figure 3.1 shows nominal prices of each provider over the studied period. We display nominal prices instead of real ones, which we used for estimation, to highlight their particular rigidity.⁹ Over the first three years, the incumbents kept basically constant prices. This situation changed first at the end of 1999 when the entrant was additionally granted GSM 900 license, which enabled him/her to offer a countrywide service.

Unfortunately, the MT did not collect data on the number of subscribers in the mobile telephone industry. Therefore, we utilize the numbers, which had been self-reported by the providers in their yearly reports and press announcements. As figure 3.2 shows, these numbers were not regularly appearing, however, they still cover approximately 75% of the quarterly series for each operator.¹⁰ Additionally, there are some monthly observations, which fall inside quarters. To fill the remaining 25% of the quarterly series and still use additional information from the monthly observations we fit a high-degree polynomial time trend to the monthly series and replace missing observations with the fitted trend.

3.4.2 Estimation issues

Before we start the estimation stage, there are two more issues, which have to be addressed. First, since we do not account for simultaneous price setting relation and estimate demand relation alone, a potential endogeneity problem arises. However, we claim that the problem might be ignored, since in our economic framework, prices are not likely to response to short-term dynamics in installed base. Optimal dynamic pricing of a telephone service like

⁸ We apply a calling pattern from the OECD Telecommunications Basket Definitions (2000).

⁹ Because inflation in the late 90's was still relatively high in Poland (on average 10% per annum over 1996-2001), the real prices fall much faster than the nominal once.

¹⁰ By a quarterly observation we mean the number of subscribers at the end of a quarter.

the one studied here, might involve raising price together with the installed base expansion.¹¹ This result is very intuitive, since low initial prices feed the diffusion of networks. In the competitive environment we consider, however, unilateral price increase over the common hedonic price level implies heavy reduction in market share of a given brand. Its provider, would not only attract no new subscribers, but could also loose the installed base if switching costs are low enough. In fact, in case of a price increase, subscribers have the option to quit long-term contract free of charge, which lowers switching costs. Therefore, the feasible dynamic price strategy is likely to involve fixed price over time rather than to react to short-term dynamics in installed base of subscribers. Figure 3.1, which plots the prices in the Polish mobile telephone industry, supports our claim. Over the first three years, the prices were basically constant, while the total number of subscribers reached almost 3 million. Price cuts in the industry started first with the prospect of new entry into the countrywide service market.¹² Nevertheless, to account for possible endogeneity of price we adopt instrumental variable technique using lagged values of price as an instrument. This does not change the results in any important way.

The second issue is error structure of the model. So far, we have not assessed the properties of $\psi_{i,t}$ in (3.6). A common assumption facilitating stochastic components in econometric models is that the researcher knows less about the economic environment than the economic agents.¹³ This might be the case, if dependent variable would be influenced by unobserved factors, which the researcher can only include in an error term. Therefore, this assumption might facilitate an additive error, properties of which depend on the nature of the unobserved factors. Potentially, they could be persistent, common for the whole market and depend on its size. This would imply an autocorrelated, correlated across equations and heteroscedastic error term respectively. Another possibility is that the variables the researcher observes differ from those the agents observe because of data reporting errors. This is likely to be the case with our data. As explained before, we needed to fill the missing values in dependent variable with the interpolated ones in order to obtain a balanced panel. Such measurement error in the dependent variable will much more complicate the error structure in (3.6) than the unobserved factors, because of lagged dependent variables and their squares included as regressors there. For instance, if we assume, that the stochastic noise in the model

¹¹ This result is obtained e.g. in Rohlfs (1974) and in Dhebar and Ohren (1985).

¹² Since the number of licenses for the mobile telephone service is regulated in Poland, we treat entry as an exogenous event.

¹³ See the discussion about introducing stochastic components into a deterministic economic model in Reiss and Wolak (2002).

is just due to an additive i.i.d. measurement error $\varepsilon_{i,t}$ in dependent variable, then the process generating the error term in (3.6) will be a nonlinear function of the measurement error, the dependent variable, and parameters of the model. To save on complicated notation, one could write it as

$$\psi_{i,t} = \varepsilon_{i,t} + g(I_t, y_{i,t-1}, y_{-i,t-1}, \varepsilon_{i,t-1}, \varepsilon_{-i,t-1}; \theta), \quad (3.7)$$

where $\theta = (\gamma_1, \gamma_2, \gamma_{11}, \gamma_{12}, \gamma_{22})$ and $g(\cdot)$ is some known function. The system (3.6) + (3.7) has features of a multivariate bilinear model studied in time series literature.

We consider two estimation techniques in this paper. The first one is generalized method of moments (GMM), which is attractive, because it does not restrict the error covariance matrix. In particular, it allows the error term in our model to be driven by the unobserved factors or the measurement errors, both of which can be heteroscedastic, autocorrelated and correlated across equations. The issue here, however, is a proper set of instruments. Since, we do not have better ones than lagged values of the variables in the model, the usefulness of GMM is limited. The problem is that autocorrelated errors in models involving lagged dependent variable can invalidate any further lags of the dependent variable as instruments. In particular, we will not be able to correctly estimate (3.6) by GMM, when the errors follow an autoregressive process. Another problem is small size of our sample, which might call for a more restricted error structure in order to gain in efficiency of estimates.

To account for these problems, we consider another technique, which Hannan and Rissanen (1982; HR thereafter) proposed for estimating linear ARMA models. Granger and Teräsvirta (1993, p.125) advocate this technique, as relatively simplest, for estimating bilinear models. For the system (3.6) + (3.7), it proceeds as follows. In the first step, we estimate (3.6) by OLS. Then, we insert the lagged OLS residuals into (3.7) as proxies for the lagged measurement error's realizations and estimate the model substituting (3.7) into (3.6). The new least squares residuals (lagged) are inserted as new proxies into (3.7) and the model is re-estimated.¹⁴ The iteration is continued to convergence, which means stabilization of residual variance. The HR technique allows us to take complex error structures that might stem from our assumptions explicitly into account. In principal, we could test validity of the assumed error structure by testing estimated coefficients in (3.7) against their theoretical values θ .

¹⁴ Since each iteration involves the use of lagged residuals from the previous estimation, we would lose observations each time. To avoid this problem, we set the missing (first) lagged residual to zero.

However, because of the small size of our sample relative to the number of coefficients in (3.6), we restrict them to equal θ .

3.4.3 Results and interpretations

Table 3.1 shows the estimation results of equation (3.6). Structural parameters of our economic model identified from the estimated coefficients can be found in table 3.2. We estimate six regressions, which differ in terms of estimation technique, parameter restrictions and error structure. Column labels in both tables identify the regressions. As mentioned before, our economic structure allows us to introduce some parameter restrictions in (3.6). Indeed, we restrict two out of three coefficients in both “Cross-network effects” group and “Pricing regime” group of table 3.1 using overidentification of the compatibility parameters w and Δw . In particular, the coefficients on $y_{i,t-1}y_{j,t-1}$ and $y_{j,t-1}^2$ were set as $\gamma_{21} = 2\gamma_{11}\gamma_2/\gamma_1$ and $\gamma_{22} = (\gamma_{11}/\gamma_1)^2\gamma_2$ respectively. The coefficients on $D_i y_{i,t-1}y_{j,t-1}$ and $D_i y_{j,t-1}^2$ were set analogously. In some regressions, we restrict also the “Fixed effects of entry”. Their theoretical values could be calculated from the data without estimation.¹⁵ We would expect them to be zero, if switching costs in the industry were negligible. In the case of significant switching costs, we expect λ_1 and λ_2 to equal 1,3 and λ_3 to equal -2,6. These numbers mean that persistent advantage of each incumbent over entrant in terms of installed base amounts to 390 thousand subscribers.

The results in the first two columns of table 3.1 are obtained by applying iterated GMM estimation. Regression (a) is a benchmark, where it is assumed that there are no network effects in demand for the mobile telephone service. In this case, our economic structure boils down to a basic linear demand model. For consistency with the remaining five regressions, the benchmark regression (a) is estimated on data aggregated to the industry level. Lagged aggregated price and a constant were used as instruments. With this set of instruments the regression coefficients are exactly identified, i.e. there are no over-identifying moment conditions.

Column (b) of table 3.1 shows the iterated GMM estimates of the full model’s coefficients. As instruments served lagged by two periods number of subscribers, lagged prices – both subscribers and prices at the firm level – and a constant. The over-identifying moment conditions seem valid. The value of the Wald statistics 7,15 falls under the 5%

¹⁵ For details concerning derivation of fixed effects of entry in our economic model we use see chapter 2.

critical value 21,03. The estimated coefficients in the regression (b) are in general highly significant and they have expected signs. The price effect is drastically smaller than that in the benchmark regression. Instead, there is an evidence of positive network effects in the regression (b). As table 3.2 reports, the identified network effects parameters c and d are equal 227,7 and -0,7 respectively. These results imply that the marginal network effects are downward sloping, as it is commonly assumed in the theoretical literature. Compatibility in the Polish mobile telephone market is low, although statistically very significant. The estimate of w equals 0,063, which means that the subscribers value own-brand installed base 16 times more than the installed base of other brands. Abandoning of intra-network discounts slightly increases the compatibility, as indicated by the parameter Δw in table 3.2.

The only insignificant coefficients in regression (b) are the incumbents' fixed effects of entry λ_1 and λ_2 . The entrant's fixed effect λ_3 is negative, but relatively small. As mentioned before, we expect that each incumbent's advantage over entrant amounts roughly to 400 thousand subscribers, when switching costs are significant in the industry. The estimated fixed effects in (b) tell us that the advantage equals just 60 thousand, which suggests that the switching costs are in fact low.¹⁶

As we mentioned in the previous section, the presence of autocorrelation invalidates the GMM estimates in our case. The pooled Durbin-Watson statistic in regression (b) falls into inconclusive region giving no clear indication of autocorrelation.¹⁷ To check robustness of the GMM estimation results, we proceed imposing more structure on the error term and applying the estimation technique devised by HR. First, we assume that the stochastic noise in the model (3.6) is due to an additive i.i.d. measurement error $\varepsilon_{i,t}$ in dependent variable.¹⁸ Then, the error term $\psi_{i,t}$ in (3.6) can be represented by (3.7). Column (c) in table 3.1 reports estimation results of the model (3.6) + (3.7). The coefficients in the regression (c) are not highly significant and the pooled Durbin-Watson statistic indicates autocorrelation.¹⁹

¹⁶ Precisely speaking, our economic structure does not support any intermediate values of fixed effects of entry between those implied by insignificant and significant switching costs. However, the intermediate values intuitively imply a gradual catching up of entrant in terms of number of subscribers. This would happen, if switching costs are in some sense low.

¹⁷ Formally, the Durbin-Watson statistic, which we report in table 3.1, is not an appropriate test in presence of lagged dependent variables, because it is usually biased toward a finding of no autocorrelation.

¹⁸ We also tried to estimate the model assuming that the error term in (3.6) is driven by unobserved factors that are autocorrelated. To account for the autocorrelated error term in presence of lagged dependent variables, we applied approach devised by Hatanaka (1974), which is based on instrumental variable technique. The results were not satisfactory. In particular, the basic demand coefficients were not statistically significant.

¹⁹ To account for possible endogeneity of price, we also estimated (c) replacing least squares with instrumental variable technique at each iteration of the HR procedure. As an instrument for price we took its lagged value.

To improve statistical properties of the model we restrict fixed effects of entry in accordance with our economic structure. Point estimates of the fixed effects of entry in regression (c) suggest that the number of subscribers of the entrant and of the incumbents differ substantially. The magnitude of this difference, which equals roughly 3.9, is in line with the predictions of our economic model given that there are significant switching costs in the industry. However, in comparison with theoretical predictions, the values of all three parameters λ_i are shifted down by 1.3. Given relatively low statistical significance of the coefficients in regression (c), this downward shift might be due to difficulties with disentangling the common intercept from fixed effects of entry by the estimation procedure. Therefore, we restrict the coefficients of the fixed effects of entry to their theoretical values. The statistical significance of the estimated coefficients improves, as we see in column (d) of table 3.1. However, the pooled Durbin-Watson statistic still indicates autocorrelation of the error term, which we interpret as an evidence for misspecification of the model.²⁰

To account for the autocorrelated disturbances $\varepsilon_{i,t}$ in the model (3.6) + (3.7) we include lagged residuals of the model as an additional regressor in (3.6). Columns (e) and (f) in table 3.1 show the results of this exercise without and with restrictions on estimated fixed effects of entry respectively. Both regressions display now good statistical properties. In particular, the pooled Durbin-Watson statistic no longer indicates autocorrelation of the error term. To ensure that the finding of no autocorrelation is correct, we conduct a modification of Breusch-Goldfrey test consisting of regressing residuals from each of the regressions (e) and (f) on their full set of regressors plus lagged residuals and testing significance of coefficients on the lagged residuals. Each time we use one, two and four lags and test their joint significance with the standard F test. In none of the cases we could reject the null hypothesis of no autocorrelation. Although, the regressions (e) and (f) perform well in statistical terms, they differ substantially in their economic relevance. All the variation of the dependent variable in regression (e) is explained by autoregressive factors leaving coefficients on basic demand variables insignificant and with wrong signs. In contrast, regression (f) performs well, both in

The procedure did not converge. Applying the same modification to regressions (d) – (f) we obtained convergence, but the estimates were only slightly changed, so we do not report these results.

²⁰ Another problem with regressions (c) and (d) was convergence of the estimation procedure. Each of them converges to two distinct solutions. Since in each case the solutions are very close to each other, we report just one of them.

statistical and economic terms, providing a useful robustness check for the GMM estimates in regression (b).²¹

In general, the results of regressions (f) confirm those of regression (b) in the way both sets of results differ from the benchmark. The price coefficient in column (f) of table 3.1, which equals $-0,057$, is smaller than in the benchmark regression (a) by an order of magnitude. The marginal network effects are positive²² and diminishing, as indicated by parameters c and d in column (f) of table 3.2. Also, the estimated compatibility measure w , which equals $0,089$, is relatively low. However, the difference in magnitude of price and network effects between regressions (b) and (f) gives rise to very much diverted predictions concerning demand in the Polish mobile telephone industry. To see it, we plot steady-state demand in the industry, which is derived from (3.6) given that $y_i = y_{i-1}$ for all i . On the horizontal axis of figure 3.3, there is aggregated number of subscribers in the industry $y = \sum_i y_i$. For simplicity of the exposition we assume that there are three providers of equal size in the market, hence the prices across providers are also equal and are represented by a common price p . As a reference, we include also demand function implied by the benchmark regression (a) in figure 3.3. We see that the estimated network effects in regressions (b) and (f) give rise to upward-sloping parts in the inverse steady-state demand functions. As discussed in chapter 2, the upward-sloping parts consist of unstable equilibria and correspond to the notion of critical mass. Once the critical mass is reached for a given price p , the number of subscribers in the market tends to and eventually reaches the stable steady state on the downward-sloping part of the demand. In the opposite case, the market shrinks to zero. Focusing on the stable steady states, we note that for each price, regression (b) implies higher demand than regression (f), which in turn implies higher demand than the benchmark regression (a). E.g., for p equal 100 zloty, the demand is estimated at roughly 15, 13 and 7 million subscribers by regression (b), (f) and (a) respectively. The same ranking sustains for the price sensitivity of demand, regression (b) supporting the least sensitive one. Converting the estimated coefficients into price elasticities²³, evaluated at the sample means, we obtain

²¹ We also tried to restrict the fixed effects of entry in regression (e) to zero. The HR estimation procedure did not converge.

²² The marginal network effects are positive for a network (and not installed base) smaller than 30,6 mill. subscribers. Since the population of Poland reaches 40 mill., they are positive roughly over the whole relevant range.

²³ In the case of regressions (b) and (f), we mean long-run price elasticity, i.e. we account for the change in demand between two steady states. Obviously, the short-run elasticities are even smaller than the long-run ones reported in the text.

-0,4, -1,2, and -3,0 for regressions (b), (f) and (a) respectively. In fact, the most important conclusion that we draw from comparison of the three regressions' results is that ignoring network effects leads to significant overestimation of the elasticity. The intuition behind direction of the bias is the following. By ignoring installed base we falsely attribute all changes in current number of subscriptions to changes in price alone. Since both rising installed base of subscribers and falling price lead to more subscriptions, if, loosely speaking, price and installed base are negatively correlated, then the price coefficient is biased downward and consequently the price elasticity overestimated. We formalize this intuition in section 3.6.1 of the appendix.

Another question of interest is to what extent the identified network effects and compatibility alter demand in the industry. To answer this question we simulate the steady-state demand, implied by the estimates of regression (f), manipulating values of the respective structural parameters. The results of this exercise are illustrated in figure 3.4. Note that the demand function under non-discriminatory pricing, in which case compatibility equals $0,089 + 0,036 = 0,125$, indicated by the bold solid line in figure 3.4, corresponds exactly to the demand function implied by regression (f) in figure 3.3. Setting the network effect parameters c and d to zero, we obtain linear demand function indicated by the thin solid line in figure 3.4. Without network effects, the market size shrinks roughly from 15 to 1 million subscribers. The extent of network effects we found might explain why the second-generation telecommunications in Poland performed so much better than the first generation, as we mentioned in section 3.2.2. First, marginal network effects in the analog first generation might diminish much faster than in digital second generation because of congestion effects. Second, provider of the first generation service might have failed to attract the critical mass of subscribers. To assess the economic significance of compatibility, we additionally plot steady-state demand function under discriminatory pricing, i.e. for $w = 0,089$, in figure 3.4. Although, the difference in compatibility between the two pricing regimes seems marginal, it produces a noticeable difference in terms of steady-state demand. In particular, difference in the market size amounts to 2 million subscribers. We conclude that compatibility parameters, we identified in the Polish mobile telecommunications, are not only statistically, but also economically significant.

Finally, given the extent of compatibility we found, one could discriminate among possible sources of network effects in mobile telecommunications. As we discussed in section 3.2.1, some of them are more likely to generate firm specific effects (low compatibility),

while the other industry specific effects (high compatibility). We are also able to test the endogenous network effect hypothesis, as in Blonski (2002), more directly, estimating change in the compatibility due to abandoning of intra-network discounts. Our results provide empirical evidence in favor of the endogenous network effect hypothesis, since Δw is both statistically and economically significant. However, this hypothesis alone is not able to explain the network effects we observe. Even under non-discriminatory pricing, compatibility in the industry remains low. This suggests that the network effects in mobile telecommunications might also be raised by transmission of information about quality of the service and Leibenstein's (1950) conformist behavior of consumers.

3.5 Conclusions

In this paper we specify and estimate a structural model of demand for mobile telephone service. The focus of the study is to provide empirical evidence concerning the extent of network effects and compatibility in mobile telecommunications. The structural approach we follow allows us to identify the parameters of interest from the S-shape of mobile service diffusion. Data for the analysis come from the Polish mobile telephone industry and cover the period 1996-2001, i.e. from introduction of the second generation digital service till recently.

Our results suggest that there are strong network effects in the Polish mobile telecommunications both in statistical and economic terms. The regression coefficients and the identified structural parameters capturing the extent of the effects are statistically highly significant. To illustrate their economic significance, we derive steady-state demand in the industry and compare it to the simulated demand without network effects. The market size that results from setting network effects to zero is 15 times smaller than in the full model. As a benchmark, we also estimate a restricted regression that does not account for network effects. The estimated price elasticity of demand in there is much higher, since we falsely attribute changes in demand due to network effects to changes in price. We formalize this result showing that ignoring network effects in empirical models of emerging network industries can substantially bias downward the estimated elasticity of demand.

We provide also empirical evidence on compatibility of competing networks in mobile telecommunications. In general, we find that the compatibility is low. It amounts to 0,089

under discriminatory pricing, which involves intra-network call discounts. After abandoning of the discounts, the compatibility increases to 0,125. This seemingly marginal increase leads to expansion of the market size by roughly 2 million subscribers, i.e. 15%. Given the compatibility we found, one could discriminate among possible sources of network effects in mobile telecommunications. The significant implications of intra-network call discounts we found support endogenous network effects as one of the sources. The endogenous network effects hypothesis states that larger installed base of subscribers to a given brand implies lower monthly bill for them, hence network effects at the firm level, if intra-network calls are cheaper than inter-network ones. Since without such discriminatory pricing, the compatibility is still relatively low, we conclude that endogenous effects alone are not able to explain the network effects we find. Transmission of information about quality of the service and conformist behavior of consumers offer alternative explanations.

3.6 Appendices

3.6.1 Direction of bias in the estimated price elasticity of demand

This section formalizes the intuition behind the direction of bias in the estimated price elasticity of demand when we ignore network effects. Suppose that correctly specified linear demand model reads

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon, \quad (3.8)$$

where y is an $N \times 1$ vector of quantities, X_1 is an $N \times K$ matrix, columns of which are N observations on each out of K basic explanatory variables and β_1 is a $K \times 1$ vector of corresponding parameters. X_2 denotes an $N \times L$ matrix of variables responsible for network effects and β_2 is an $L \times 1$ vector of parameters. ε stands for an $N \times 1$ vector of disturbances and $E[\varepsilon] = 0$. We allow for heteroscedasticity and autocorrelation of disturbances letting the covariance matrix $E[\varepsilon\varepsilon'] = \Sigma$. Suppose further that some variables in X_1 are not orthogonal to ε and a valid set of instruments for them is included in Z , which is an $N \times M$ matrix, and $M \geq K$.

If we regress y on X_1 without including X_2 our GMM estimate of β_1 is then

$$\begin{aligned} b_1^{\text{GMM}} &= [X_1' Z S^{-1} Z' X_1']^{-1} [X_1' Z S^{-1} Z' y] = \\ &= [X_1' Z S^{-1} Z' X_1']^{-1} [X_1' Z S^{-1} Z' (X_1\beta_1 + X_2\beta_2 + \varepsilon)] = \\ &= \beta_1 + [X_1' Z S^{-1} Z' X_1']^{-1} [X_1' Z S^{-1} Z' X_2] \beta_2 + [X_1' Z S^{-1} Z' X_1']^{-1} [X_1' Z S^{-1} Z' \varepsilon], \end{aligned} \quad (3.9)$$

where S is a matrix from the sample data that converges in probability to the same matrix as $\frac{1}{N} Z' \Sigma Z$. In particular, if disturbances ε are uncorrelated, S could be the White (1980) estimator, otherwise it could be the Newey and West (1987) estimator.

To assess asymptotic properties of the estimator in (3.9) let us first assume that $\frac{1}{N} Z' \Sigma Z$, $\frac{1}{N} X_1' X_1$ and $\frac{1}{N} X_2' X_2$ converge in probability to positively definite matrixes, and that $\frac{1}{N} Z' X_1$ and $\frac{1}{N} Z' X_2$ converge in probability to nonzero matrixes. Then

$$\text{plim } b_1^{\text{GMM}} = \beta_1 + \text{plim} \left[\left(\frac{1}{N} X_1' Z \right) S^{-1} \left(\frac{1}{N} Z' X_1' \right) \right]^{-1} \left[\left(\frac{1}{N} X_1' Z \right) S^{-1} \left(\frac{1}{N} Z' X_2 \right) \right] \beta_2 +$$

$$+ \text{plim} \left[\left(\frac{1}{N} X_1' Z \right) S^{-1} \left(\frac{1}{N} Z' X_1 \right) \right]^{-1} \left[\left(\frac{1}{N} X_1' Z \right) S^{-1} \left(\frac{1}{N} Z' \varepsilon \right) \right]. \quad (3.10)$$

If Z is indeed a proper set of instruments for X_1 in model (3.8), then $E[z_n \varepsilon_n] = 0$, where the subscript $n = 1, 2, \dots, N$ depicts the n -th row of the corresponding matrix, so $\text{plim} \frac{1}{N} Z' \varepsilon = 0$ and the third term on the RHS of (3.10) vanishes. However, the second term on the RHS of (3.10) does not, which means that the estimator is inconsistent. Multiplying out the second term on the RHS of (3.10) we obtain a vector of persistent biases. The direction of bias on the price coefficient is of particular interest for us. Suppose for simplicity that X_2 contains just one variable – the lagged dependent variable, i.e. installed base in our setting – so that β_2 is a scalar. If network effects are positive on average then β_2 will be positive. The vector of biases in (3.10) is then a product of a positive scalar and probability limit of a string of matrix computations, which we recognize as an estimator of some parameter vector γ from the following relation

$$X_2 = X_1 \gamma + v. \quad (3.11)$$

If Z again contains proper set of instruments than the estimator of γ is consistent and we can finally state that the direction of bias on the price coefficient in our original demand model ignoring network effects is determined by the sign of parameter on price in (3.11). Negative correlation – negative partial correlation if X_1 contains more variables than just constant and price – between price and installed base implies then downward bias on price coefficient, hence on price elasticity, in the misspecified demand. This direction of the bias is especially likely in the emerging network markets, where installed base increases and price tends to decrease over time.

3.6.2 Tables

Table 3.1. Demand in the Polish mobile telephone industry: Estimation results

Variables	(a) ^{ab}	(b)	(c)	(d)	(e)	(f)
Basic demand						
<i>Intercept</i>	143,6 ^{***} (23,2)	2,620 ^{**} (2,50)	13,97 [*] (1,73)	12,31 ^{***} (2,89)	-4,646 (-1,23)	9,340 ^{***} (4,68)
<i>p_{i,t}</i>	-0,735 ^{***c} (-16,6)	-0,014 ^{**} (-2,64)	-0,065 (-1,68)	-0,072 ^{***} (-3,35)	0,022 (1,22)	-0,057 ^{***} (-5,62)
Own network effects						
<i>y_{i,t-1}</i>	-	3,185 ^{***} (37,7)	2,756 ^{***} (4,73)	2,886 ^{***} (15,7)	3,202 ^{***} (14,1)	2,850 ^{***} (29,7)
<i>y_{i,t-1}²</i>	-	-0,010 ^{***} (-6,40)	-0,0101 [*] (-1,83)	-0,0097 ^{**} (-2,42)	-0,0055 ^{**} (-2,07)	-0,0093 ^{***} (-4,36)
Cross-network effects ^d						
<i>y_{-i,t-1}</i>	-	0,200 ^{***} (6,87)	0,298 (1,66)	0,182 [*] (1,95)	0,141 [*] (1,97)	0,254 ^{***} (5,77)
Pricing regime ^d						
<i>D_iy_{-i,t-1}</i>	-	0,052 ^{**} (2,92)	0,064 (0,88)	0,141 ^{***} (2,93)	-0,012 (-0,40)	0,102 ^{***} (4,27)
Fixed effects of entry						
<i>λ₁</i>	-	0,167 (0,64)	0,614 (0,66)	1,3	1,421 ^{***} (3,74)	1,3
<i>λ₂</i>	-	0,183 (1,03)	-0,031 (-0,03)	1,3	1,018 ^{***} (2,77)	1,3
<i>λ₃</i>	-	-0,766 ^{**} (-2,47)	-3,949 (-1,65)	-2,6	0,786 (0,76)	-2,6
Lagged disturbances						
<i>ε_{i,t-1}</i>	-	-	-	-	0,933 ^{***} (5,92)	1,164 ^{***} (8,19)
N	20	54	57	57	57	57
pooled D-W	0,52	1,71	0,81	0,79	1,99	1,99

*** denotes significance at 1% level, ** at 5% level, * at 10% level; t-statistics in parentheses.

^a (a) and (b) are estimated by iterated GMM, (c) – (f) are estimated by technique devised in Hannan and Rissanen (1982).

^b For consistency with the results of other estimations the observations in (a) are aggregated to the market level.

^c Aggregated price is calculated as average price across providers weighted by their market shares.

^d Restrictions on regression coefficients, which over-identify structural parameters, are imposed.

Table 3.2. Identified structural parameters of demand in the Polish mobile telephone industry

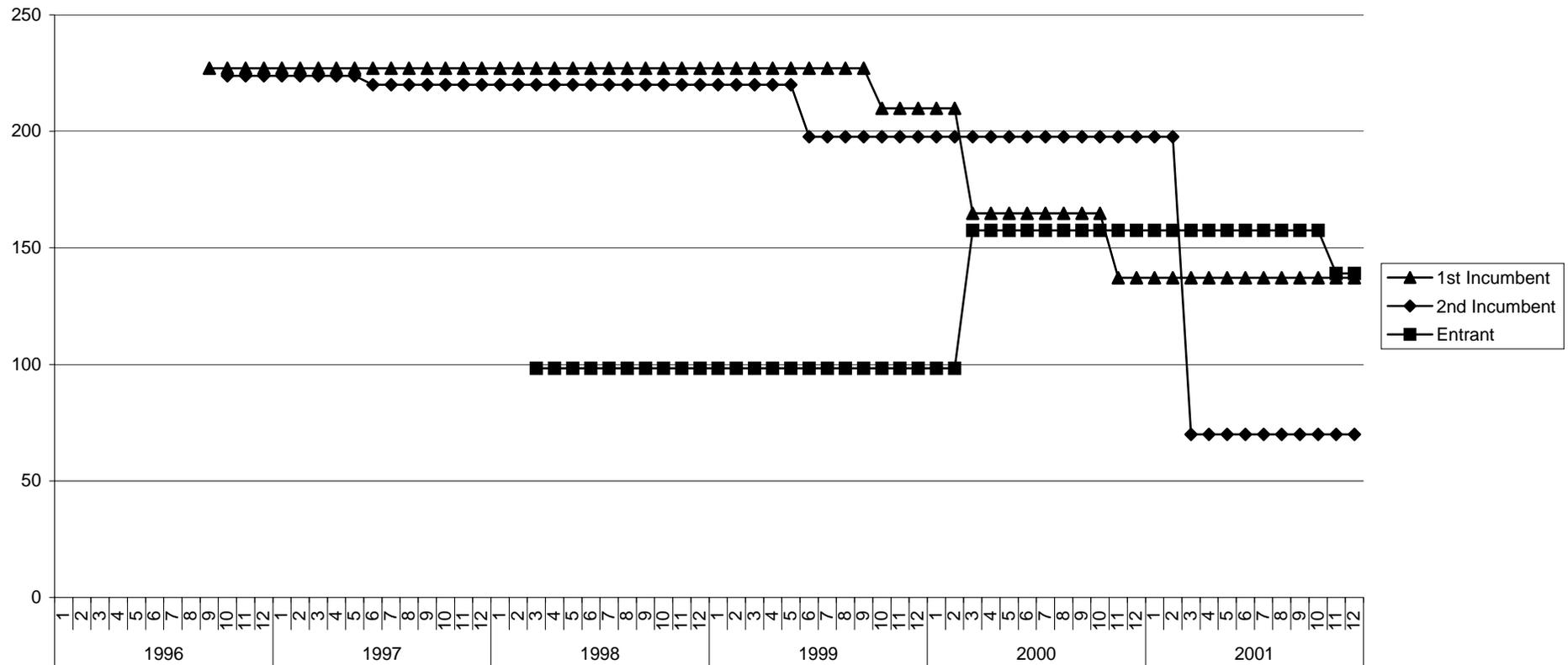
Parameters	(b) ^a	(c)	(d)	(e)	(f)
Distribution of consumer types					
<i>a</i>	187,3*** (14,7)	215,0*** (17,8)	170,6*** (12,5)	209,9*** (13,1)	163,5*** (17,9)
<i>b</i>	0,014** (2,64)	0,065 (1,68)	0,072*** (3,35)	-0,022 (-1,22)	0,057*** (5,62)
Network effects					
<i>c</i>	227,7** (2,67)	42,42 (1,33)	40,00*** (3,34)	-144,7 (-1,28)	49,91*** (5,59)
<i>d</i>	-0,706*** (-3,17)	-0,156 (-1,29)	-0,135* (-1,94)	0,250 (0,98)	-0,163*** (-3,45)
Compatibility					
<i>w</i>	0,063*** (6,00)	0,108 (1,27)	0,063* (1,86)	0,044* (1,77)	0,089*** (5,36)
Δw	0,016*** (3,08)	0,023 (0,78)	0,049*** (3,05)	-0,004 (-0,40)	0,036*** (4,40)

*** denotes significance at 1% level, ** at 5% level, * at 10% level; t-statistics in parentheses.

^a Column labels correspond to these in table 1.

3.6.3 Figures

Figure 3.1 Best-deal prices in the Polish mobile telephone industry



i) Best-deal prices are given in current zlotys (prices used for estimation are in September 1996 Polish zlotys).

ii) In the period studied 1 U.S. dollar \approx 4 Polish zlotys.

Figure 3.2 Number of subscribers in the Polish mobile telephone industry

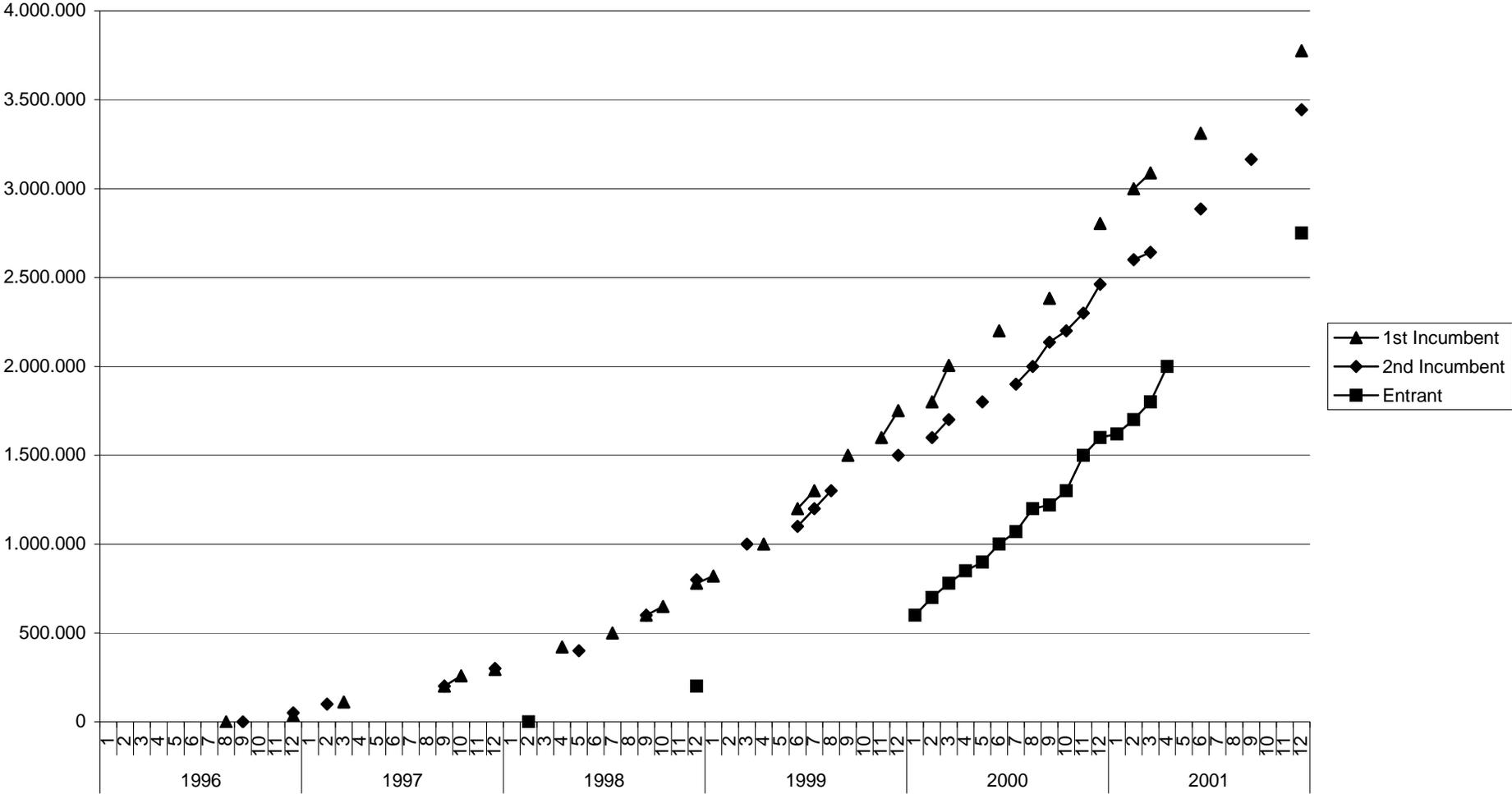
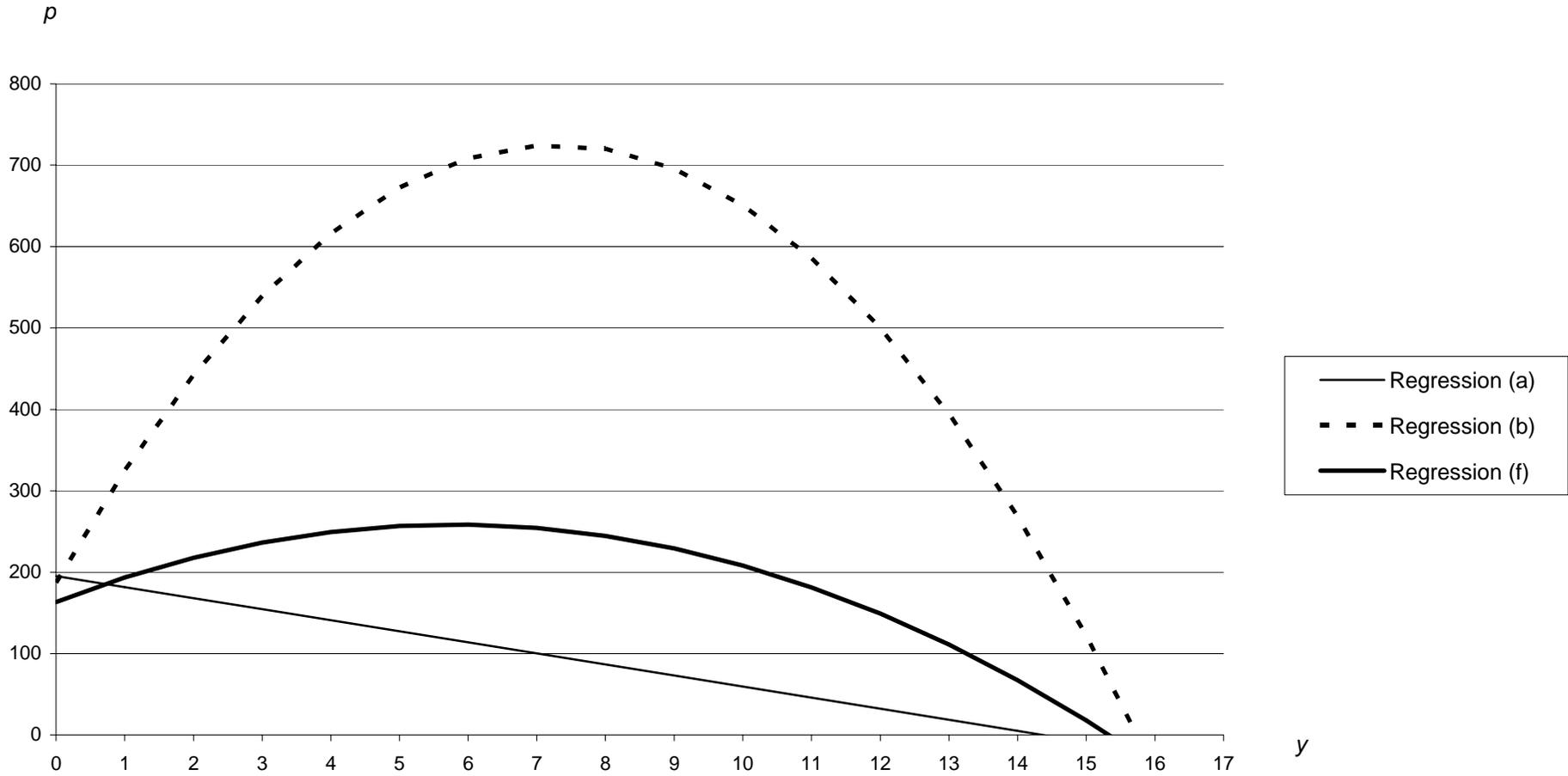
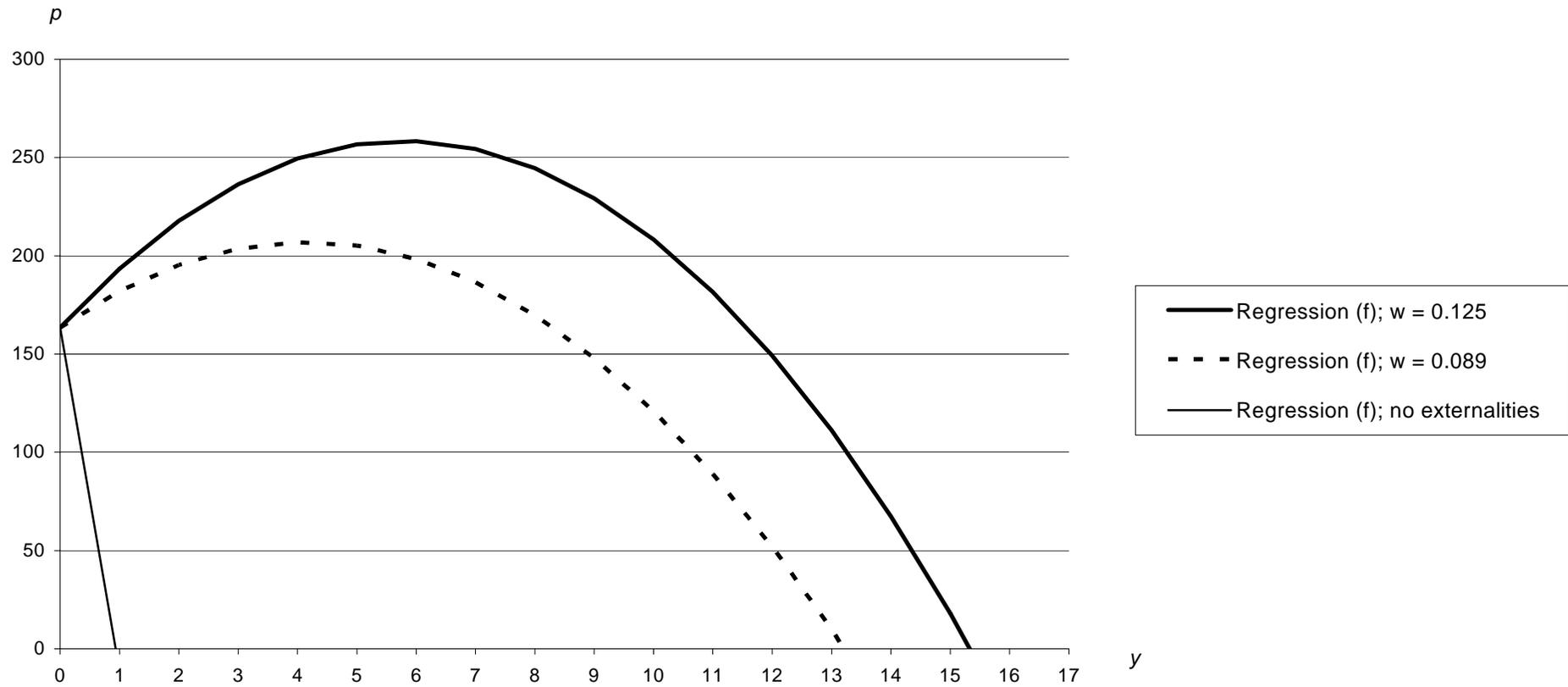


Figure 3.3 Aggregated steady-state demand in the Polish mobile telephone industry



i) Aggregated number of subscribers y in millions

Figure 3.4 Simulations of the aggregated steady-state demand in the Polish mobile telephone industry for different values of the structural parameters



i) Aggregated number of subscribers y in millions

Chapter 4

Diffusion of ISO 9000 Standards and International Trade

4.1 Introduction

The main goal of the International Organization for Standardization (ISO) is to harmonize standards around the world, which, as it is widely claimed, promotes trade and therefore global welfare. A prominent example of the work done by ISO is the ISO 9000 family of standards, often referred to as generic quality management standards. The vision of the developers is that

(...) through its worldwide acceptance and use, the ISO 9000 family of standards will provide an effective means for improving the performance of individual organizations and providing confidence to people and organizations that products (goods and services) will meet their expectations thereby enhancing trade, global prosperity and individual well-being.¹

Critics of ISO 9000 claim, however, that it is merely a barrier to market entry and a tariff on international trade. There are valid arguments on both sides. On the one hand, ISO 9000 might be a common language, which lowers informational asymmetry between firms and allows them to organize trade more efficiently. Indeed, the standards emphasize clear and open communication with customers, as well as with suppliers.² Furthermore, they provide a tool facilitating screening and performance evaluation. Learning this common language offers

¹ <http://www.tc176.org/About176.asp>

² See ISO (2001).

an alternative for establishing vertical relations based on long-term relationship and brand reputation. On the other hand, ISO 9000 has been used as a standard against which to assess performance in government procurements and in setting of minimum quality requirements for imports. This raises a concern that the standard is mainly a tool for protecting domestic markets.

This paper empirically investigates the impact of ISO 9000 on international trade. We estimate a gravity equation for bilateral trade incorporating ISO 9000 adoptions in each country as a factor affecting bilateral trade barriers. As it has been pointed out in the literature, the causality might go both ways. International trade might benefit from standards' harmonization, as trade barriers decrease, and standardization process might in turn be determined by intensity of foreign trade, which indicates openness of an economy (see e.g. Casella, 1996; Moenius, 2000). As a consequence, empirical models of international trade using standardization as an explanatory variable may suffer from endogeneity bias. Formal tests, however, do not reject strict exogeneity of ISO 9000 adoptions in our gravity equation. Nevertheless, we estimate ISO 9000 diffusion equation to obtain additional insights into the role of ISO 9000 for international trade.

The empirical literature investigating the impact of common standards on trade is scarce. In particular, we are not aware of any study that investigates the performance of ISO 9000 in this context. Few existing empirical analyses of ISO 9000 focus on managers' motivation to seek the certification and on market reaction to it. Examples include Anderson et. al. (1999), who find that after controlling for regulatory and customer pressures, providing credible signals of quality assurance to external parties motivates the adoption decision. Further, Docking and Downen (1999) examine the reaction of the firms' stock price to the announcement of ISO 9000 registration. They find that, for the smaller firms, investors react positively to the announcement and that there was no significant reaction for the larger firms.

The work by Moenius (2000), who looks at the impact of country-specific and bilaterally shared product and process standards on international trade, is probably closest related to our work. He finds that both country-specific and shared standards are favorable to trade. Similarly, Swann et al. (1996) report that both international and country-specific product standards promote imports into the U.K. Further, Blind (2001) analyses Switzerland's trade of measurement and testing products with Germany, France, and the U.K. He finds that the stock of both national and international standards in this sector has a positive impact on

the trade flows. In turn, Blind (2002) investigates factors responsible for intensity of standardization in 20 industrial sectors of seven countries. He reports empirical evidence on the positive relation between the stock of national and international standards in a sector and the ratio of exports to total production in that sector.

To a great extent, our work relates also to the strand of literature considering the role of networks in reducing information costs associated with international trade (e.g. Rauch, 1999; Rauch and Trindade, 2002). In our view, the role of ISO 9000 for international trade very much overlaps with the role of ethnic Chinese networks for trade studied by Rauch and Trindade (2002). To the extent that ISO 9000 lowers information and search costs, it also relates to the role of Internet for trade studied by Freund and Weinhold (2004). Neither of these studies, however, discusses the potential endogeneity of networks' formation.

The rest of the paper is organized as follows. Section 4.2 describes the ISO 9000 family of standards and its role in international trade more in detail. Section 4.3 provides theoretical models of bilateral trade flows and standard diffusion, which guide our empirical analysis. Data, empirical implementation of the theoretical models, and discussion of the results are presented in section 4.4. Section 4.5 concludes.

4.2 ISO 9000 and its role in international trade

The ISO 9000 family of standards is often referred to as generic quality management standards. They are generic in the sense that they can be implemented by any organization regardless of its size, the sector of its activity, and its managerial or national culture. Quality management reflects what the organization does to enhance customer satisfaction by meeting his/her requirements and expectations.³ Compliance with ISO 9000 indicates consistent use of documents and standardized procedures to produce a good or service, for which the customer contracts. In other words, ISO 9000 certifies that the firm's products conform to the specification.

The history of ISO 9000 started in 1987 with publication of the ISO 9000 Quality Assurance Standards by a Technical Committee (TC 176) of the International Organization

³ See ISO (2002).

for Standardization (ISO). By the end of 2001, the number of ISO 9000 certificates exceeded half a million in 161 countries around the world, contributing to its reputation as an international reference for quality requirements in business-to-business dealings.⁴ We treat ISO 9000 as a uniform standard although it consists of a series of nested standards, which evolved over time. Originally, the core members of the family, with which firms could actually be certified, were ISO 9001, ISO 9002, and ISO 9003. They differed in terms of the quality system elements they covered. The nested nature of these standards allowed firms to accommodate differences in the scope of their operations.⁵ The 2000 edition of the ISO 9000 family replaced these three standards with a single one labeled ISO 9001: 2000. As supplementary standards, the 2000 edition included ISO 9000: 2000, which describes fundamentals and specifies vocabulary for a quality management system, and ISO 9004: 2000, which provides guidelines for performance improvements. Both of them were developed on the basis of previous standards, which they replaced. Given that the core members of the ISO 9000 family were finally replaced by a single one, our simplifying assumption treating ISO 9000 as a uniform standard seems justified.

ISO 9000 adoption is a sovereign decision of each firm, however, they can seek certification only in their home countries. Each country has one government-designed accrediting agency, which certifies the competency of third party registrars to conduct ISO 9000 quality audits. The registrars are also charged with issuing of certificates.⁶ In general, motivations behind the implementation of ISO 9000 could be divided into three main categories: i) compliance with government regulations, ii) entering new form of vertical relations due to use of a common language, and iii) internal efficiency gains. In fact, all factors influencing managers to seek ISO 9000 certification that Anderson et al. (1999) identify after a comprehensive review of practitioners journals fall into one of the three categories.

The first category stems from the fact that ISO 9000 has been used as a standard against which to assess performance in government procurements and in setting of minimum quality requirements for products that affect public safety. The Single Market Initiative of 1992 initiated by the European Community involves the most noticeable example of such regulations. The public safety argument obliged firms to attain a uniquely designed EC Mark

⁴ Ibid.

⁵ See Anderson et al. (1995, 1999) for details.

⁶ Ibid.

in order to get access to certain markets.⁷ ISO 9000 was selected as a means to attain the mark in most of the cases. This raises the concern that ISO 9000 can be a barrier to market entry and a tariff on international trade. We relate our empirical findings to potential effects of the regulations.

The second category of motivations is the focus of this study. As noted by Bénézech et al. (2001, p. 1396), “the ISO 9000 series can be viewed as a code, a language used by firms to extend their industrial relationship”. Thanks to standardized documentation flow and organizational procedures within certified firm, ISO 9000 provides a screening device that allows other firms to observe and to evaluate its performance. This naturally lowers informational asymmetries between firms. Consequently, ISO 9000 proxies for conformance of the firm’s product to the specification, for which the customer contracts. This leads to lower transaction and search costs in vertical relations between firms. To realize the benefits of ISO 9000, however, both contracting parties should have adopted (i.e. learned) it in the first place. This is why the common language analogy is appropriate.⁸ Learning this common language could be viewed as an alternative for establishing vertical relations based on long-term relationship and brand reputation. This explains the potential of ISO 9000 for reducing barriers to market entry and non-tariff barriers to trade.

Finally, firms seek ISO 9000 certification to realize efficiency gains. The discipline of documentation and organizational procedures could reduce waste, lower costs, and improve productivity. For example, relying on a survey of ISO 9000 certified firms in the U.S. (sales from \$100 million to \$500 million), Anderson et al. (1999) report average annual savings of \$200 000 due to the certification.

The same authors further report that obtaining ISO 9000 certification at a manufacturing site in the U.S. takes from 9 to 28 months and approximately 35-40% of all sites fail the first audit. The costs of the standard adoption and certification are substantial. A medium size manufacturing facility employing 100 people can expect to spend \$50 000. For larger firms (sales from \$100 million to \$500 million), the average cost that the authors report is \$300 000. We relate to these efficiency gains, costs, and timing of adoption modeling ISO 9000 diffusion in section 4.3.2.

⁷ Ibid; depending on the product category, the EC Mark must have been attained till 1992-1995.

⁸ We give the common language hypothesis a slightly different spin than Bénézech et al. (2001). They concentrate rather on the role of ISO 9000 as a means to codify the knowledge within a firm.

4.3 Economic models

4.3.1 International trade flows

The standard empirical framework used to predict international trade flows is the gravity equation. In a simple form, which explains its name, the equation reads

$$V_{ij} = A \frac{Y_i Y_j}{D_{ij}}, \quad (4.1)$$

where V_{ij} is value of exports from country i to country j , Y_i and Y_j are their economic masses often measured by GDP or GNP, D_{ij} is a measure of the distance between them and A is a constant of proportionality.⁹ Attractiveness of the gravity equation originally stems from its empirical explanatory power. Recent developments, however, show that the equation can also be theoretically motivated, in particular, in the context of classical Heckscher-Ohlin framework, as presented by Deardorff (1998).

The following discussion is going to selectively overview the theoretical foundations of the gravity equation in order to facilitate selection of the explanatory variables, as well as to help the interpretation of our empirical results.

Suppose first that preferences of consumers in every country are identical and homothetic. Suppose further that trade is balanced – so that each country's expenditures equal its income – and frictionless, i.e. there are no transport costs and no other impediments to trade. Then, as shown by Deardorff (1998), the value of i 's total exports to j is

$$V_{ij} = \frac{Y_i Y_j}{Y^w}, \quad (4.2)$$

where Y^w stands for world income. Thus, an even simpler gravity equation than (4.1) emerges with constant of proportionality $A = 1/Y^w$. Deardorff (1998) further shows that if preferences are not identical and /or not homothetic, then bilateral trade flows are *on average* given by the simple frictionless gravity equation (4.2) under some additional regularity conditions. He

⁹ Some empirical studies, in which gravity equation for trade is not theoretically motivated, often define V_{ij} as bilateral trade flows rather than exports alone.

argues also that the tendency of high-income consumers to consume larger budget share of capital-intensive goods will lead high-income capital-abundant countries to trade more than average with each other and less than average with low-income labor-abundant countries. This would motivate inclusion of per capita income in the gravity equation, as it is often the case in empirical literature.

Now, we turn to the more realistic case where trade is impeded. That is, we allow for transportation costs, tariffs and other non-tariff barriers to trade that increase the price of domestically produced goods in foreign markets. A particularly elegant gravity equation for this case is derived by Anderson and Wincoop (2003). In their model goods are differentiated by the place of origin and each region is specialized in the production of only one good. Moreover, preferences in each region are identical and homothetic, approximated by a constant elasticity of substitution (CES) utility function. Their gravity equation then reads

$$V_{ij} = \frac{Y_i Y_j}{Y^w} \left(\frac{t_{ij}}{P_i P_j} \right)^{1-\sigma}, \quad (4.3)$$

where σ is the elasticity of substitution between all goods in the CES utility function. The trade barriers between i and j are captured by the trade cost factor t_{ij} such that the price of region i goods for region j consumers equals $p_i t_{ij}$ and p_i denotes the exporter's supply price, net of trade costs.¹⁰ ¹¹ P_i and P_j are the consumer price indices of i and j . The key implication of the model is that trade between regions is determined by *relative* trade barriers, as the price indices $\{P_i\}$ depend on all trade cost factors $\{t_{ij}\}$. Unfortunately, empirical implementation of the model is troublesome. As discussed by the authors, the price indices $\{P_i\}$ are not observable, since they cannot be interpreted as consumer price levels.¹² Moreover, the authors claim uniqueness of their implicit solution for the price indices only for the case with symmetric trade cost factors.

Given the difficulties of the structural approach, we are going to model the impact of ISO 9000 adoptions on trade in a semi-structural way. As we have already argued, ISO 9000

¹⁰ Note that the trade cost factor t_{ij} negatively affects the value of exports V_{ij} in (4.3) only if the substitution between goods is elastic ($\sigma > 1$).

¹¹ The theoretical gravity equation (4.3) is valid under the assumption that the trade cost factors are symmetric, i.e. $t_{ij} = t_{ji} \forall i, j$.

¹² While $\{P_i\}$ are consumer price indices in the Anderson and Wincoop's (2003) model, they should be interpreted much broader, as indicated by the authors. E.g., a home bias in preferences instead of trade costs might lead to exactly the same gravity equation for exports as (4.3).

could be understood as a common language, adoption of which allows firms to lower the transaction and search costs. We will assume that, by lowering these costs, adoptions of the standard in both country i and country j will decrease their bilateral trade barriers. Section 4.4.2 on empirical implementation of the model presents that more in detail.

4.3.2 ISO 9000 diffusion

The main hypothesis, which drives our adoption process, is that ISO 9000 exhibits significant network effects. Network effects naturally arise in this context, as in the context of any other language. The number of ISO 9000 adopters determines the size of the pool of potentially more efficient business contacts, hence the value of the standard for each adopter. Since we are interested in the link between ISO 9000 and international trade, the relevance of foreign adoptions of the standard for the home country adopters is of crucial importance. In particular, we would like to test whether economic distance between countries in terms of trade related factors matters for the relevance of foreign adoptions. Finding such relationship would give us additional insight about the role of ISO 9000 in international trade. The model developed in the second chapter of this thesis facilitates such analysis. It allows us to derive structural country-specific diffusion equations from adoption decisions of individual firms in each country. The model in chapter 2, however, is formulated to study the adoptions of competing network goods by consumers in one country or region. With ISO 9000, we face the situation where a single network good is supposed to be adopted by consumers in various countries. Therefore, we need to slightly reformulate the model.

We assume that in each country $i = 1, 2, \dots, I$ there is an infinite number of heterogeneous firms, which instantaneously decide whether to adopt ISO 9000 or not. The adoption decision is influenced by the firm-specific intrinsic valuation of the standard, denoted by v_i , which corresponds to the efficiency gains that the firm realizes after the adoption. Another factor influencing the adoption decision is the network size of certified firms at time t denoted by $x_i(t)$. Network effects arise in the adoption of ISO 9000 due to potentially more efficient contracting among the certified firms. The efficiency gains and network effects together shape the net instantaneous benefits of ISO 9000 adoption, which for simplicity takes the following functional form

$$u(v_i, x_i(t)) = v_i + cx_i(t) + dx_i^2(t), \quad (4.4)$$

where c and d are parameters that capture the extent of network effects. The costs of ISO 9000 adoption consist of a sunk investment in the reorganization of the firm q_i and an instantaneous audit fee p_i , both of which are assumed to be constant over time.

Firms maximize the net benefits adopting ISO 9000 when the present value of the stream of future benefits exceeds the present value of the costs. As shown in appendix 4.6.1, we can calculate the net-of-cost intrinsic valuation of indifferent firm in country i at time t , denoted as $\tilde{v}_{i,t}^*$, from the following first order condition

$$\tilde{v}_{i,t}^* + cx_i(t - \delta) + dx_i^2(t - \delta) = 0, \quad (4.5)$$

The lag of network size δ in (4.5) is crucial for the dynamic properties of the model. It can be motivated by excessive optimism of the firms regarding time needed to implement the standard in their own sites. In fact, as we point out in section 4.2, the adoption process takes time and large share of firms fails the first audit.

Assume that the net-of-cost intrinsic valuations in each country – $\tilde{v}_{i,t}$ – are uniformly distributed over an interval $(-\infty, a_i]$ with some density b_i . Integrating over all firms with intrinsic valuation higher than $\tilde{v}_{i,t}^*$, we obtain the equilibrium number of adopters in each country

$$y_i(t) = (a_i + cx_i(t - \delta) + dx_i^2(t))b_i. \quad (4.6)$$

The final step in setting up the model is to define the relevant network for ISO 9000 adopters in each country. We are not going to differentiate the firms in a given country in terms of their cooperation prospects with each other. Instead, we want to emphasize the difference between the foreign and the domestic firms. Therefore, we define the network size of ISO 9000 adopters as

$$x_i(t) = y_i(t) + \sum_{j \neq i} w_{ij} y_j(t), \quad (4.7)$$

where w_{ij} reflects the relative importance of country j adopters for country i adopters. Since we expect that foreign markets are relevant for the adoption decisions of domestic firms, we are going to relate w_{ij} to bilateral trade between i and j in the next section. The general idea is that the intensity of trade indicates closeness of the economies, hence the relevance of foreign firms as business partners for the domestic firms.

Finally, substituting (4.7) in (4.6), we arrive at the following adoption equation

$$y_i(t) = a_i b_i + b_i c \left(y_i(t - \delta) + \sum_{j \neq i} w_{ij} y_j(t - \delta) \right) + b_i d \left(y_i(t - \delta) + \sum_{j \neq i} w_{ij} y_j(t - \delta) \right)^2, \quad (4.8)$$

which guides the empirical analysis in section 4.4

4.4 Empirical model

4.4.1 Data

Data on ISO 9000 adoptions comes from ISO (2002). Bilateral exports are taken from the UN Commodity Trade Statistics Database (Comtrade) and GDP and population figures come from the International Financial Statistics published by the IMF. The data ranges over 1995-2001 and covers 101 countries. Table 4.1 lists the countries included in this study together with the number of firms that adopted ISO 9000 standard till the end of 2001 in each of the countries. Summary statistics of the variables used in the estimations are given in table 4.2.

A distinctive feature of this study is its particularly wide coverage of countries. In fact, our sample covers approximately 80% of the world trade and 99% of the world ISO 9000 adoptions. A primary reason for this is that the adoptions of ISO 9000 in each country depend on the world diffusion of the standard, as predicted by equation (4.8). Therefore, the smaller the coverage of the sample is, the more severe is the concern about omitted variable bias in the estimates. Additionally, inclusion of the less developed countries, for which the trade barriers aspect of ISO 9000 is potentially more severe, is important for the generality of our results. In

the context of gravity equation for trade, the wide coverage of countries might seem problematic. The theories of trade in imperfect substitutes, which were the first to justify the gravity model, were thought to apply only to the industrialized countries. However, as found by Hummels and Levinsohn (1995), the model works equally well for the larger set of countries.¹³

4.4.2 Implementation of gravity equation for trade

In our empirical implementation of gravity equation, we apply panel data techniques, which have the advantage over cross-section estimations, that they can capture all time invariant trade determinants by means of country-pair specific effects. This is very useful, since the trade barriers in gravity models are usually difficult to quantify, as they might consist of tariff barriers, transportation costs, information costs, etc., some of which are not even observable. Among many variables proposed by researchers to approximate the trade barriers, there are geographical distance, linguistic and colonial ties, membership in trade agreements and monetary unions, and common border. To the extent that these variables are time invariant, which is very likely given the relatively short time span of our data, the country-pair specific effects will account for them. A particular ingredient of the trade barriers, we consider, are search and transaction costs. According to the main hypothesis of this paper, they can be lowered for firms that invested in learning of the common language – ISO 9000. Following other authors we assume that the trade barriers can be approximated by a log-linear function. Then, our specification of the trade barriers in (4.1) reads¹⁴

$$D_{ijt}^{-1} = e^{\eta_{ij}} (1 + ISO_{it})^{\delta_1} (1 + ISO_{jt})^{\delta_2}. \quad (4.9)$$

According to (4.9) the lack of ISO 9000 awarded firms in both countries i and j simplifies the measure of the bilateral trade barriers to a function of country-pair specific effect. Since, we do not impose symmetry on the trade barriers, η_{ij} is in fact importer-exporter specific. This asymmetry is going to be important later on, because it allows us to distinguish the distance to

¹³ See the discussion in Frankel et.al. (1997), p. 55.

¹⁴ As it is clear from the discussion in previous chapter, the measure of distance between two countries D_{ij} could be broader interpreted as the measure of trade barriers between them. However, according to the model by Anderson and Wincoop (2003), this interpretation implicitly assumes that the elasticity of substitution between goods produced in different countries is bigger than one.

foreign customer (importer) from the distance to foreign supplier (exporter) from the domestic firm viewpoint. Unilateral adoptions of ISO 9000 affect average trade barriers only marginally.¹⁵ Bilateral adoptions multiply this effect, as follows from our interpretation of ISO 9000 as a common language. For the sake of generality, we do not restrict the parameters δ_1 and δ_2 to be equal, although this would fit our common-language hypothesis.

Then, after substituting the measures of economic masses and the constant of proportionality in (4.1) with specific functions of the observables, the gravity equation that we estimate becomes

$$\ln X_{ijt} = \alpha + \beta_1 \ln GDP_{it} + \beta_2 \ln GDP_{jt} + \gamma_1 \ln POP_{it} + \gamma_2 \ln POP_{jt} + \delta_1 \ln(1 + ISO_{it}) + \delta_2 \ln(1 + ISO_{jt}) + \lambda_t + \eta_{ij} + \varepsilon_{ijt}, \quad (4.10)$$

where X_{ijt} denotes exports from country i to country j in year t .¹⁶ We allow the parameters on countries' GDP – β_1 and β_2 – to be different from one and from each other. We also include *POP* variable, which measures the countries' population in millions. The reason for that is to capture the tendency that reach countries trade more than average with each other, as explained in section 4.3.1.¹⁷ Finally, λ_t stands for time effects, which are meant to capture changes in the world income, and ε_{ijt} is a usual i.i.d. error term.

4.4.3 Implementation of ISO 9000 diffusion equation

Now, we turn to the empirical implementation of the ISO 9000 adoption equation (4.8). As already mentioned, we are mainly interested in the relevance of international markets for the adoption decisions of domestic firms. Our assumption, which follows from the common-language hypothesis, is that ISO adoptions in close economies reinforce each other relatively more than ISO adoptions in distant economies. There are two natural candidates for measuring this economic closeness/distance; intensity of bilateral trade and bilateral trade barriers. In the context of gravity equation (4.10), the difference between the two is that intensity of bilateral trade depends on the countries' economic masses and some

¹⁵ We would like to stress that specification (4.9) describes *average* bilateral trade barriers, since the trade barriers faced by ISO 9000 certified firms and non certified firms differ.

¹⁶ In equation (4.10), a small technical difficulty arises, because of zero-valued entries in bilateral exports X_{ijt} . Following Frankel et al. (1997), we treat these observations as missing.

¹⁷ Actually the inclusion of $\ln POP$ or $\ln(GDP/POP)$ in (4.6) is mathematically equivalent. What changes is only the interpretation of the parameters. See Frankel et al. (1997) for a simple exposition of this point.

unobserved factors captured by the error term on top of the bilateral trade barriers. We assume that the measure of economic distance relevant for the ISO 9000 adoptions are the bilateral trade barriers. In other words, we assume that the relevance of a foreign ISO adoption for domestic firms depends neither on the size of the foreign economy in which the adoption took place, nor on the other unobserved factors captured by the error term in (4.10). Our specification of the economic closeness then reads

$$w_{ij} = w_1 e^{\eta_{ij}} + w_2 e^{\eta_{ji}}, \quad (4.11)$$

where w_1 and w_2 are some constant weights and $\{\eta_{ij}\}$ are the importer-exporter specific effects in (4.10). Since η_{ij} reflects the barriers for exports of i to j and η_{ji} reflect the barriers for imports of i from j , specification (4.11) distinguishes the distance to foreign customers – weighted by w_1 – from the distance to foreign suppliers – weighted by w_2 . In fact, estimation of the parameters w_1 and w_2 drives our interest in the ISO 9000 adoption equation. Positive values of w_1 and/or w_2 would suggest that the firms' adoption decisions are indeed affected by the number of potential foreign customers and/or suppliers certified with ISO 9000.

We also need to relate country-specific parameters of the types' distribution a_i and b_i to some observables in order to avoid estimation of excessive number of parameters in (4.8). One could expect that b_i , which reflects the number of firms with a given efficiency gains prospects from ISO 9000 adoption, depends on the total number of firms in country i , which in turn positively correlates with the GDP of that country. Similarly, one could argue that a_i , the maximum realizable efficiency gains across firms in country i , depends on the country's GDP. The rationale is that the efficiency gains due to ISO 9000 adoptions increase with the firm size – so a_i most probably reflects the efficiency gains of the largest firms in the country – and the world largest firms are located in the richest countries. We are going to relate a_i and b_i to the country's GDP one at a time to keep the model as simple as possible. This leads us to the following specifications

$$a_i = a \quad \wedge \quad b_i = b_0 + b_1 GDP_i \quad (4.12)$$

and

$$a_i = a_0 + a_1 GDP_i \quad \wedge \quad b_i = b. \quad (4.12')$$

Then, after applying specifications (4.11) and (4.12) and substituting theoretical values in the equation (4.8) with observables, the ISO 9000 adoption equation that we estimate becomes

$$ISO_{it} = a(b_0 + b_1 GDP_{it}) + c(b_0 + b_1 GDP_{it}) \left(ISO_{i(t-1)} + \omega_1 \Sigma ISO_{i(t-1)}^C + \omega_2 \Sigma ISO_{i(t-1)}^S \right) + d(b_0 + b_1 GDP_{it}) \left(ISO_{i(t-1)} + \omega_1 \Sigma ISO_{i(t-1)}^C + \omega_2 \Sigma ISO_{i(t-1)}^S \right)^2 + \psi_{it}, \quad (4.13)$$

where the variables $\Sigma ISO_{i(t-1)}^C$ and $\Sigma ISO_{i(t-1)}^S$ are the indices of foreign customers' adoptions and foreign suppliers' adoptions respectively.¹⁸ The indices make use of the asymmetry in the importer-exporter specific effects $\{\eta_{ij}\}$, which reflect the asymmetry in bilateral trade barriers. They are defined as $\Sigma ISO_{i(t-1)}^C \equiv \sum_{j \neq i} e^{\hat{\eta}_{ij}} ISO_{j(t-1)}$ and $\Sigma ISO_{i(t-1)}^S \equiv \sum_{j \neq i} e^{\hat{\eta}_{ji}} ISO_{j(t-1)}$,

where $\{\hat{\eta}_{ij}\}$ are the estimates of $\{\eta_{ij}\}$. ψ_{it} is an i.i.d. error term, which capture the influence of unobserved factors. Alternatively, applying specification (4.12') instead of (4.12), we obtain

$$ISO_{it} = a_0 b + a_1 b GDP_{it} + cb \left(ISO_{i(t-1)} + \omega_1 \Sigma ISO_{i(t-1)}^C + \omega_2 \Sigma ISO_{i(t-1)}^S \right) + db \left(ISO_{i(t-1)} + \omega_1 \Sigma ISO_{i(t-1)}^C + \omega_2 \Sigma ISO_{i(t-1)}^S \right)^2 + \psi'_{it}. \quad (4.13')$$

Note that ω_1 replaced w_1 and ω_2 replaced w_2 in both equations (4.13) and (4.13'). This is because the importer-exporter specific effects $\{\eta_{ij}\}$ cannot be identified separately from the constant α in equation (4.10) without additional assumptions. To obtain $\{\hat{\eta}_{ij}\}$ we will arbitrarily assume that $\max_{i,j} \{\hat{\eta}_{ij}\} = 0$, i.e. we will normalize the smallest net-of-ISO-9000

measure of trade barriers in our sample to 1.¹⁹ In case of any such normalization, the relation between w_k and ω_k can be shown to be $\omega_k = w_k e^{\Delta} \quad \forall k = 1, 2$, where Δ is some unknown

¹⁸ Note that in contrast to (4.8), time in (4.13) is treated as discrete variable. As a consequence, δ in (4.8) becomes "one period" in (4.13).

¹⁹ The reason why we choose this particular normalization is merely that the order of magnitude of coefficients in (4.13) and (4.13') is not too diverse. This makes exposition of the estimation results look nicer.

constant. This means that we are still able to make some inference about w_k having estimated ω_k . In particular, they have the same signs and the ratio of ω_1 and ω_2 equals the ratio of w_1 and w_2 .

The assumption, which gives rise to the specification (4.11), that the bilateral trade barriers are the relevant measure of economic distance in the context of ISO 9000 diffusion, has an important implication for the econometric treatment of our model. It implies that ISO 9000 adoptions and bilateral trade are not interdependent. So, we avoid the necessity of estimation simultaneous-equation model. An indirect way to test this assumption is to test the exogeneity of ISO variables in (4.10). If they would indeed depend on the bilateral trade itself rather than on bilateral trade barriers, the test should fail.

4.4.4 Discussion of the results

First, we estimate (4.10) by fixed effects (FE). Typically, researchers report FE estimation results along with random-effects (RE) results. The advantage of FE estimation over RE estimation is that consistency in the former does not rely on orthogonality between the country-pair specific effects η_{ij} and all the other explanatory variables.²⁰ We skip the RE estimation, since we expect the adoptions of ISO in country i to depend on the economic distance to each trade partner indicated by η_{ij} . The FE estimation results are presented in table 4.3.

The first three columns (1) – (3) contain results of the regressions, in which we exclude some of the explanatory variables and column (4) corresponds to the regression with the full set of covariates. We see that estimated coefficients vary only marginally across these different specifications. In general, we find the coefficient on countries' own income β_1 and the coefficient on partner's income β_2 to lie about 0,3 and 0,7 respectively. Moreover, the coefficients on population γ_1 and γ_2 tend to be of the same magnitude with reversed sign as income coefficients. These estimates imply that, in contrast to theoretical prediction and the results from cross-sectional studies (see e.g. Frankel et al., 1997, table 4.2 and table B6.6), trade is almost entirely driven by countries' income per capita. Glick and Rose (2002), who also apply FE estimation, report findings, which are similar to ours (see table 4 in there).

²⁰ We follow the approach expressed by Wooldridge (2002, p. 251–252), that FE and RE correspond to the assumptions we are willing to impose on unobserved effects η_{ij} in order to estimate the model rather than to their deterministic or stochastic nature.

The next four regressions (5) – (8) in table 4.3 augment the first four regressions by inclusion of the leading explanatory variables, as suggested by Wooldridge (2002, p. 285), in order to test the strict exogeneity assumption. The null hypothesis of the Wald test reported in table 4.3 is that coefficients on the leading explanatory variables equal zero. We see that the Wald test rejects the null – and thereby strict exogeneity of the explanatory variables – in all four cases at very high significance. Without strict exogeneity, FE estimator becomes inconsistent, so the estimates of the gravity equation coefficients in table 4.3 can be misleading.

We believe that the reason why strict exogeneity failed in the FE estimations is that macroeconomic indicators like exports and GDP tend to be trending variables. This is also true for population and ISO adoptions variables. As it is well known from the time series literature, simple OLS is likely to deliver spurious correlations between trending variables. The most straightforward remedy for that is to use first differencing (FD) estimation.

Table 4.4 reports the results of the same exercise, as in table 4.3, using FD estimation. Again, we see that the estimated coefficients of the gravity equation are stable across different specifications. However, the estimates changed in comparison to the FE results. The most noticeable change concerns population variables. The coefficient on country's own population γ_1 changed sign to positive and the coefficient on partner's population γ_2 decreased in magnitude. Both coefficients are now statistically insignificant, which is in contrast to the findings of cross-sectional studies. The reason for this discrepancy might be that the intended effect of countries' income per capita on trade is already captured by the country-pair specific effects η_{ij} . This is in line with the Anderson and Marcouiller's (2002) alternative explanation of why capital-abundant countries trade disproportionately with each other. They argue that the reason are strong institutions to support trade security, which can be plausibly treated as constant in our data.

Given the interpretation, that country-specific effects absorb the impact of per capita income, the coefficient on partner's GDP in table 4.4, which equals roughly 0,77, corresponds well to the findings of cross-sectional studies. The coefficient on own GDP is however about 10 times smaller and only marginally significant.²¹ The most important indicator, which gives us some confidence about plausibility of our results, is the Wald test. It does not reject strict

²¹ Freund and Weinhold (2004) also report insignificant FD estimate of own GDP coefficient in their gravity equation for exports. Moreover they find the coefficient on partner's GDP to lie just slightly above 0,10 (see table 3 in there).

exogeneity of the covariates in first differences, at least in the regressions without population variables (column 13 and 15 in table 4.4).

Finally, we turn to the impact of ISO-9000-adoptions' variables, which are the focus of our study. Their FD estimates are lower than the FE estimates, but still significant at 5% level. The coefficients on own and partner's adoptions equal 0,027 and -0,026 respectively. This means that 10% increase in the number of firms awarded with ISO 9000 certificate in a country leads on average to 0,27% increase in bilateral exports and 0,26% decrease in bilateral imports of that country. These results provide an empirical evidence for the role ISO 9000 plays in international trade, although they are not fully in line with our expectations. The common language hypothesis, as we stated it, suggests that bilateral trade flows should rise with both, exporter's and importer's, adoptions, since they both contribute to the number of potentially more efficient business links. However, this line of argument does not take into account the possibility of substitution between suppliers, i.e. exporters in this case. Freund and Weinhold (2004) argue along the same lines interpreting their findings on impact of the Internet on bilateral trade. Also, Anderson and Wincoop (2003) point to the fact that bilateral trade flows depend on trade barriers between all trading parties.²² In our specification of the gravity model, concentration of ISO 9000 adoptions in few countries around the world could explain falling average bilateral imports with the number of adoptions. In fact, the OECD members (30 out of 101 countries in our sample) account for 75% of ISO 9000 adoptions in table 4.1.

To explore further the substitution effect of ISO 9000 on international trade we restrict our sample to the OECD countries and repeat the FD estimations. The results are reported in table 4.5. Again, the population variables proved to be insignificant and cause endogeneity problems. Coefficients on own and partner's GDP are higher than for the whole sample roughly by 0,1 and 0,2 respectively and highly significant. Coefficients on own and partner's ISO 9000 adoptions are also significant and amount to 0,063 and 0,036 respectively. This supports the substitution effects hypothesis. Bilateral exports between ISO 9000 abundant countries indeed increase with both domestic and foreign adoptions. These results, however, should be treated with caution, because the Wald test rejects strict exogeneity even in the regressions without population variables (columns 21 and 23 in table 4.5). Now, the GDP variables are responsible for the rejection. There are good reasons to believe that the

²² See the discussion in section 4.3.1.

endogeneity of GDP in these regressions is merely a statistical phenomenon and that it does not significantly bias the results. First, the inclusion of variables in levels does not significantly change the coefficients on variables in differences. Second, Hummels and Levinsohn (1995) report that correcting for the endogeneity of GDP with instrumental variables makes very little difference.

As we mentioned in section 4.2, critics of ISO 9000 reasonably argue that the standard is actually a barrier to trade, since it has been used as a tool for introducing import restrictions. In fact, the positive effect of domestic ISO 9000 adoptions on exports, that we found, could be explained by increasing access to the regulated markets. However, the impact of domestic adoptions on imports, that we found, cannot be explained by the trade barrier hypothesis. In particular, the finding that imports increase with domestic ISO 9000 adoptions within OECD countries cannot be supported by this hypothesis. The reversed relation in the full sample might be due to the use of ISO 9000 in introducing import restrictions, however, under additional assumption that the restrictions increases with domestic ISO 9000 adoptions.²³

Estimation results of the ISO 9000 diffusion equation provide us additional insights into the link between the standards and international trade. To estimate (4.13) and (4.13') we follow the general method of moments (GMM) approach for linear dynamic panel data models, which was proposed by Arellano and Bond (1991). By doing so, we allow for additional unobserved heterogeneity on top of (4.13) and (4.13'). This unobserved heterogeneity accounts for institutional factors, like the national accrediting agencies, which could either spur or hamper the diffusion process.

To obtain linear models we multiply out the terms in the structural equations (4.13) and (4.13'). By multiplying out the terms, the parameters ω_1 and ω_2 – the transformed weights in our measure of economic closeness (4.11) – become overidentified. This allows us to test the underlying structure of the empirical diffusion equations. The estimated coefficients of the linearized equations (4.13) and (4.13') are reported in table 4.6, in columns (25) and (25') respectively.²⁴ Except for the lagged dependent variable, all the regressors are

²³ Since the panel data estimation techniques we use provide time-series evidence, because cross-section differentiation is captured by country-pair specific effects, this assumption is much more heroic than it might seem.

²⁴ In the regression (25) we additionally assumed that $d = 0$ in (4.13). Without this assumption the main coefficients of the model turned out to be insignificant and the Arellano-Bond test indicated second-order autocorrelation in the residuals. The likely reason for this is the multi-collinearity of the explanatory variables.

treated as strictly exogenous. We see that both regressions perform reasonably well in statistical sense, as indicated by the Sargan and the Arellano-Bond test statistics at the bottom of the table. However, the underlying economic structure of both regressions (25) and (25') is rejected by the data. According to the structural equation (4.13), the ratio of coefficients on $\Sigma ISO^S_{i(t-1)}$ and $ISO_{i(t-1)}$ in the regression (25) should be equal to the ratio of coefficients on $GDP_{it}\Sigma ISO^S_{i(t-1)}$ and $GDP_{it}ISO_{i(t-1)}$, because both ratios identify the same parameter ω_2 . Yet, the ratios have opposite signs. Similarly, the equation (4.13') predicts that the ratio of coefficients on $(\Sigma ISO^C_{i(t-1)})^2$ and $ISO^2_{i(t-1)}$ in the regression (25') should be equal to the squared ratio of coefficients on $\Sigma ISO^C_{i(t-1)}$ and $ISO_{i(t-1)}$, since they identify $(\omega_2)^2$. But, the first ratio is negative.²⁵ Therefore, we are going to treat the results as coming from reduced form approach and limit the discussion to pointing out major correlation patterns.

The results of the regressions (25) and (25') suggest in general that ISO 9000 adoptions in each country are positively related to the market size measured by the country's GDP. The adoptions exhibit also significant inertia, as indicated by the coefficients on the lagged dependent variable. Since we are mostly interested in the foreign trade considerations in the adoption decisions of firms, the crucial result in table 4.6 are the coefficients on the indices of foreign customers' adoptions and foreign suppliers' adoptions. We find that domestic ISO 9000 adoptions are positively related to foreign customers' adoptions, as expected. In contrast, foreign suppliers' adoptions are not or weakly negatively correlated with domestic adoptions depending on the specification. In other words, the diffusion of ISO 9000 seems to proceed from customers to suppliers and not the other way around, at least in the international context. This asymmetry does not necessarily contradict our common language hypothesis. It might reflect the fact that business customers are able to benefit from ISO 9000 without actually being certified, as someone, who understands a language without having passed exams that certify that. At the same time, certified suppliers might not care about certification of the customers, if their relations are already established.

To check the robustness of these findings, we perform additional two regressions (26) and (26'), in which the indices of foreign ISO 9000 adoptions in (25) and (25') are replaced with simple unweighted sum of foreign adoptions. The coefficients on the unweighted foreign adoptions are much less significant than those on the indices of foreign adoptions in the previous regressions. In other words, our indices are much better predictors of domestic

²⁵ We also carried out formal Wald-type tests, which confirm this intuitive argumentation.

adoptions than is the unweighted sum. This gives us some confidence in the effects of foreign trade considerations on domestic adoptions we found in (25) and (25').

4.5 Conclusions

In this paper, we empirically assess the link between the ISO 9000 family of standards and international trade. According to the vision of its developers, ISO 9000 should provide confidence to people and organizations that products will meet their expectations, thereby enhancing trade and global welfare. In contrast, its critics claim that it is merely a barrier to market entry and a tariff on international trade.

Our modeling strategy is to look at the impact of ISO 9000 adoptions on bilateral trade flows between countries. We estimate a gravity equation for bilateral exports using data on 101 countries over 1995-2001. To obtain additional evidence, we estimate an international ISO 9000 diffusion equation and test whether the number of certified foreign trade partners plays a role in the domestic firms' adoption decisions.

Using the full sample, we find that domestic ISO 9000 adoptions spur bilateral exports and hamper bilateral imports. In the sample restricted to the OECD countries, however, the domestic adoptions are positively related to both bilateral exports and imports. Additionally, we find empirical evidence on the positive effect of foreign customers' adoptions of ISO 9000 on domestic adoptions.

In general, these findings suggest that the ISO 9000 standards have indeed significant positive impact on international trade. They are consistent with the common language hypothesis, which states that ISO 9000 lowers informational asymmetry between firms and allows them to organize vertical relations more efficiently. The negative effect of ISO 9000 adoptions on bilateral imports in the full sample could be explained by the substitution effect. Since domestic certified firms are likely to choose other certified firms as suppliers, the average bilateral imports from all trade partners will fall with domestic adoptions if the concentration of adoptions around the world is high enough.

The hypothesis that ISO 9000, as a tool for introducing import restrictions, is a barrier to international trade is not able to explain our empirical findings, although we cannot reject

it. The substitution effect we pointed out might be, however, another way, in which ISO 9000 constitutes an effective trade barrier. If the sluggish diffusion of ISO 9000 in the less developed countries continues, the benefits of ISO 9000 envisioned by its developers will remain in the developed countries' domain.

4.6 Appendices

4.6.1 Optimal timing of ISO 9000 adoptions

We assume that once the standard is adopted, it yields the infinite stream of future benefits. The present value of the stream of benefits can be written as

$$V(v_i, x_i(t), t) = \int_t^{\infty} e^{-\rho s} u(v_i, x_i(s)) ds, \quad (4.14)$$

where ρ is a common instantaneous discount rate. The cost of adoption consists of the sunk investment in reorganization of the firm q_i and the instantaneous audit fee p_i , both of which are assumed to be constant over time. We write the present value of the cost as

$$C(q_i, p_i, t) = e^{-\rho t} q_i + \int_t^{\infty} e^{-\rho s} p_i ds. \quad (4.15)$$

The optimal timing of adoption involves solving the following maximization problem

$$\max_t [V(v_i, x_i(t), t) - C(q_i, p_i, t)]. \quad (4.16)$$

Assuming that problem (4.16) is concave, we obtain the intrinsic valuation of indifferent firm $v_{i,t}^*$ as the solution to

$$\frac{d}{dt} V(v_{i,t}^*, x_i(t), t) - \frac{d}{dt} C(q_i, p_i, t) = 0, \quad (4.17)$$

which simplifies to

$$u(v_{i,t}^*, x_i(t)) - \rho q_i - p_i = 0. \quad (4.18)$$

Now, substituting benefit function (4.4) in (4.18) we finally obtain

$$\tilde{v}_{i,t}^* + cx_i(t) + dx_i^2(t) = 0, \quad (4.19)$$

where $\tilde{v}_{i,t}^* = v_{i,t}^* - \rho q_i - p_i$ denotes the net-of-cost intrinsic valuation of indifferent firm. Further, we modify solution (4.19) introducing a lag δ into the network size. The lag is crucial for identification in this model, as explained in chapter 2. In the context of ISO 9000 adoption

decision, the lag can be motivated by an excessive optimism of firms regarding the time needed to implement the standard. The fact that firms obtain certification later than implied by (4.19) leads to the modified first order condition

$$\tilde{v}_{i,t}^* + cx_i(t - \delta) + dx_i^2(t - \delta) = 0. \quad (4.20)$$

Note that deriving (4.19) we assumed that the firms have perfect knowledge about the future diffusion of network. This assumption carries over to (4.20). In other words, the firms know that the *others* are overoptimistic.

4.6.2 Tables

Table 4.1 List of countries and ISO 9000 adoptions

Country/Region	# ISO 9000 certifications Dec. 2001	Country/Region	# ISO 9000 certifications Dec. 2001	Country/Region	# ISO 9000 certifications Dec. 2001
Argentina	2324	Guatemala	18	Panama	33
Australia	26750	Honduras	11	Paraguay	46
Austria	4000	Hong Kong	3814	Peru	200
Bahrain	59	Hungary	6362	Philippines	961
Bangladesh	38	Iceland	30	Poland	2622
Barbados	11	India	5554	Portugal	2474
Belarus	78	Indonesia	1395	Qatar	52
Belgium	4670	Iran	618	Romania	1670
Belize	4	Ireland	3700	Russian Federation	1517
Bolivia	42	Israel	6447	Senegal	8
Botswana	5	Italy	48109	Seychelles	6
Brazil	9489	Japan	27385	Singapore	3513
Bulgaria	469	Jordan	402	Slovak Republic	827
Cameroon	8	Kazakhstan	41	Slovenia	1026
Canada	11635	Korea	17676	South Africa	2263
Chile	229	Latvia	67	Spain	17749
China	63188	Lithuania	202	Sri Lanka	155
Colombia	1117	Luxembourg	108	St. Lucia	3
Costa Rica	60	Macao	39	Sudan	3
Croatia	415	Malawi	1	Swaziland	11
Cyprus	334	Malaysia	3195	Sweden	4652
Czech Republic	5627	Malta	207	Switzerland	8605
Denmark	2163	Mauritius	175	Thailand	3870
Dominica	3	Mexico	2233	Trinidad and Tobago	29
Dominican Republic	25	Moldova	7	Tunisia	302
Ecuador	33	Mongolia	2	Turkey	2949
Egypt	546	Morocco	158	Uganda	60
El Salvador	17	Namibia	24	United Kingdom	66760
Estonia	202	Netherlands	12745	United States	37026
Finland	1870	New Zealand	2069	Uruguay	241
France	20919	Nicaragua	5	Venezuela	373
Germany	41629	Norway	1703	Zambia	10
Greece	2325	Oman	67	Zimbabwe	134
Grenada	3	Pakistan	539		

Source: ISO(2002)

Table 4.2 Variable description and summary statistics

Variable	Description	Mean	Std. dev.	Min	Max
X_{ij}^a	Exports from country i to country j in billion current US\$	0,50	4,12	0	239,95
GDP_i^b	GDP of country i in trillion current US\$	0,28	1,03	0,0002	10,08
POP_i^b	Population of country i in millions	49,06	160,05	0,07	1285,23
ISO_i^c	Number of country i 's ISO 9000 adopters in thousands	2,93	8,33	0	66,76
ΣISO_i^d	The sum of ISO 9000 adopters in all countries except country i in thousands	287,20	124,37	74,46	505,54
$\Sigma ISO_i^C^d$	Index of ISO 9000 adoptions by all foreign business customers of firms in country i	3,49	7,38	0,002	60,94
$\Sigma ISO_i^S^d$	Index of ISO 9000 adoptions by all foreign suppliers of firms in country i	10,78	10,98	0,45	103,56

Note: All variables are on yearly basis.

Sources: ^a UN Comtrade databank; ^b IMF International Financial Statistics; ^c ISO(2002); ^d own calculations based on all previous sources

Table 4.3 Gravity equation for exports: Fixed-effects estimation results

Dependent variable: $\ln \text{Exports}_{ijt}$								
Independent variables ^a	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\ln GDP_{it}$	0,300*** (8,51)	0,303*** (8,56)	0,296*** (8,46)	0,299*** (8,52)	0,117** (2,38)	0,118** (2,38)	0,097** (2,01)	0,091* (1,87)
$\ln GDP_{jt}$	0,697*** (20,21)	0,711*** (20,45)	0,722*** (21,16)	0,736*** (21,41)	0,736*** (14,60)	0,746*** (14,70)	0,758*** (15,41)	0,773*** (15,57)
$\ln POP_{it}$		-0,230 (-1,13)		-0,295 (-1,43)		1,072 (0,82)		-2,948** (-2,01)
$\ln POP_{jt}$		-0,651*** (-3,28)		-0,704*** (-3,49)		-1,263 (-1,06)		-0,541 (-0,37)
$\ln ISO_{it}$			0,043*** (5,18)	0,043*** (5,16)			0,021* (1,69)	0,023* (1,88)
$\ln ISO_{jt}$			-0,042*** (-5,09)	-0,043*** (-5,25)			-0,059*** (-5,37)	-0,059*** (-5,33)
<i>const</i> ^b	-8,598*** (-7,00)	-6,903*** (-5,08)	-9,098*** (-7,47)	-7,094*** (-5,22)	-12,165*** (-7,71)	-11,332*** (-6,53)	-12,325*** (-7,84)	-10,815*** (-6,14)
$\ln GDP_{i(t+1)}$					0,316*** (6,22)	0,316*** (6,21)	0,326*** (6,48)	0,326*** (6,49)
$\ln GDP_{j(t+1)}$					-0,027 (-0,53)	-0,023 (-0,46)	-0,031 (-0,63)	-0,030 (-0,60)
$\ln POP_{i(t+1)}$						-1,041 (-0,77)		2,906* (1,92)
$\ln POP_{j(t+1)}$						0,731 (0,60)		-0,165 (-0,11)
$\ln ISO_{i(t+1)}$							0,033** (2,57)	0,030** (2,31)
$\ln ISO_{j(t+1)}$							0,000 (0,01)	-0,001 (-0,10)
Wald test (χ^2)	-	-	-	-	39,02***	40,00***	48,24***	51,06***
Number of:								
observations	46909	46909	45467	45467	39023	39023	37459	37459
groups	8803	8803	8724	8724	8616	8616	8426	8426
Observations per group:								
min	1	1	1	1	1	1	1	1
avg	5,3	5,3	5,2	5,2	4,5	4,5	4,4	4,4
max	7	7	7	7	6	6	6	6
R ² :								
within	0,026	0,027	0,030	0,030	0,025	0,025	0,028	0,029
between	0,445	0,211	0,486	0,167	0,517	0,496	0,559	0,352
overall	0,417	0,197	0,461	0,155	0,482	0,469	0,533	0,334

^a Year-dummies' coefficients suppressed.

^b The constant term is defined here as the average of importer-exporter-specific effects.

*** denotes significance at 1% level, ** at 5% level, * at 10% level; t-statistics in parentheses.

Table 4.4 Gravity equation for exports: First-differencing estimation results

Dependent variable: $\Delta \ln \text{Exports}_{ijt}$								
Independent variables ^a	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
$\Delta \ln GDP_{it}$	0,076 (1,56)	0,074 (1,53)	0,073 (1,52)	0,072 (1,50)	0,072 (1,48)	0,071 (1,46)	0,071 (1,47)	0,077 (1,57)
$\Delta \ln GDP_{jt}$	0,760*** (15,34)	0,764*** (15,39)	0,776*** (15,96)	0,778*** (15,98)	0,758*** (15,25)	0,762*** (15,31)	0,776*** (15,74)	0,784*** (15,85)
$\Delta \ln POP_{it}$		0,338 (0,69)		0,304 (0,61)		-0,062 (-0,12)		0,046 (0,09)
$\Delta \ln POP_{jt}$		-0,539 (-1,14)		-0,343 (-0,71)		-0,770 (-1,55)		-0,561 (-1,10)
$\Delta \ln ISO_{it}$			0,027** (2,27)	0,027** (2,27)			0,025** (2,09)	0,018 (1,48)
$\Delta \ln ISO_{jt}$			-0,026** (-2,32)	-0,026** (-2,31)			-0,026** (-2,36)	-0,031*** (-2,70)
$\ln GDP_{it}$					-0,004 (-1,39)	-0,011*** (-2,72)	-0,005 (-0,97)	-0,020** (-2,54)
$\ln GDP_{jt}$					-0,002 (-0,93)	-0,007* (-1,67)	-0,004 (-0,79)	-0,014* (-1,79)
$\ln POP_{it}$						0,011** (2,45)		0,012*** (2,62)
$\ln POP_{jt}$						0,005 (1,20)		0,009* (1,85)
$\ln ISO_{it}$							0,001 (0,28)	0,006 (1,25)
$\ln ISO_{jt}$							0,001 (0,30)	0,004 (0,84)
Wald test (χ^2)	-	-	-	-	2,46	10,04**	3,32	13,63**
Number of observations	36583	36583	35250	35250	36583	36583	35250	35250
R ²	0,013	0,013	0,014	0,014	0,013	0,013	0,014	0,014

^a Year-dummies' coefficients suppressed.

*** denotes significance at 1% level, ** at 5% level, * at 10% level; t-statistics in parentheses.

Table 4.5 Gravity equation for exports (only OECD countries): First-differencing estimation results

Dependent variable: $\Delta \ln \text{Exports}_{ijt}$								
Independent variables ^a	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
$\Delta \ln GDP_{it}$	0,322*** (5,47)	0,316*** (5,33)	0,311*** (5,28)	0,306*** (5,16)	0,319*** (5,42)	0,259*** (4,34)	0,308*** (5,22)	0,261*** (4,36)
$\Delta \ln GDP_{jt}$	0,879*** (15,19)	0,883*** (15,10)	0,871*** (15,03)	0,874*** (14,95)	0,879*** (15,22)	0,874*** (14,88)	0,859*** (14,70)	0,863*** (14,60)
$\Delta \ln POP_{it}$		0,461 (0,73)		0,390 (0,62)		0,696 (1,10)		0,698 (1,08)
$\Delta \ln POP_{jt}$		-0,272 (-0,42)		-0,299 (-0,47)		-0,319 (-0,50)		-0,561 (-0,86)
$\Delta \ln ISO_{it}$			0,063*** (3,72)	0,063*** (3,70)			0,060*** (3,55)	0,018 (0,97)
$\Delta \ln ISO_{jt}$			0,036** (2,07)	0,036** (2,08)			0,036** (2,09)	0,035* (1,85)
$\ln GDP_{it}$					-0,011*** (-3,69)	-0,044*** (-7,54)	-0,011* (-1,92)	-0,044*** (-5,42)
$\ln GDP_{jt}$					-0,008** (-2,56)	-0,013** (-2,23)	-0,001 (-0,10)	-0,001 (-0,14)
$\ln POP_{it}$						0,041*** (6,58)		0,038*** (5,64)
$\ln POP_{jt}$						0,007 (1,08)		0,001 (0,21)
$\ln ISO_{it}$							0,000 (0,07)	0,002 (0,36)
$\ln ISO_{jt}$							-0,008 (-1,46)	-0,009 (-1,60)
Wald test (χ^2)	-	-	-	-	19,48***	63,64***	20,20***	52,62***
Number of observations	4559	4559	4559	4559	4559	4559	4559	4559
R ²	0,085	0,085	0,089	0,089	0,089	0,098	0,093	0,099

^a Year-dummies' coefficients suppressed.

*** denotes significance at 1% level, ** at 5% level, * at 10% level; t-statistics in parentheses.

Table 4.6 ISO 9000 adoption equation: Arellano-Bond estimation results

Dependent variable: ΔISO_{it}				
Independent variables ^a	(25)	(25')	(26)	(26')
Market size				
GDP_{it}	6,261 ^{***} (10,90)	1,0371 ^{***} (3,39)	4,3072 ^{***} (7,29)	0,4972 (1,51)
Domestic adoptions				
$ISO_{i(t-1)}$	1,397 ^{***} (41,49)	1,386 ^{***} (40,71)	1,4737 ^{***} (56,65)	1,251 ^{***} (47,22)
$ISO^2_{i(t-1)}$		0,0129 ^{***} (15,15)		0,0027 ^{***} (7,57)
Foreign customers' adoptions				
$\Sigma ISO^C_{i(t-1)}$	0,2309 ^{***} (7,50)	0,1697 ^{***} (5,43)		
$(\Sigma ISO^C_{i(t-1)})^2$		-0,0046 ^{***} (-2,99)		
Foreign suppliers' adoptions				
$\Sigma ISO^S_{i(t-1)}$	-0,0383 ^{***} (-2,76)	-0,0277 (-1,15)		
$(\Sigma ISO^S_{i(t-1)})^2$		0,00006 (0,09)		
Unweighted foreign adoptions				
$\Sigma ISO_{i(t-1)}$			-0,0033 ^{**} (2,60)	-0,0041 (-0,89)
$(\Sigma ISO_{i(t-1)})^2$				0,000003 (0,97)
Interaction terms				
$GDP_{it} ISO_{i(t-1)}$	-0,0362 ^{**} (-2,09)		-0,1136 ^{***} (-20,26)	
$GDP_{it} \Sigma ISO^C_{i(t-1)}$	-0,0020 (-0,17)			
$GDP_{it} \Sigma ISO^S_{i(t-1)}$	-0,2369 ^{***} (-10,27)			
$GDP_{it} \Sigma ISO_{i(t-1)}$			0,0023 ^{**} (1,97)	
$ISO_{i(t-1)} \Sigma ISO^C_{i(t-1)}$		-0,0101 ^{***} (-4,19)		
$ISO_{i(t-1)} \Sigma ISO^S_{i(t-1)}$		-0,0150 ^{***} (-6,37)		
$\Sigma ISO^C_{i(t-1)} \Sigma ISO^S_{i(t-1)}$		0,0041 [*] (1,69)		
$ISO_{i(t-1)} \Sigma ISO_{i(t-1)}$				-0,0004 ^{***} (-10,84)
Sargan test (χ^2) ^b	17,98 (14)	16,78 (14)	17,11 (14)	21,03 (14)
Arellano-Bond m_2 test ^c	1,45	1,41	1,47	1,34
Number of observations	475	475	475	475

*** denotes significance at 1% level, ** at 5% level, * at 10% level; t-statistics in parentheses.

^a All variables are in first differences.

^b Sargan test of over-identifying restrictions; degrees of freedom in parentheses.

^c Arellano-Bond test of second-order serial correlation in residuals.

Bibliography

- Ahn H., M.-H. Lee (1999), "An econometric analysis of the demand for access to mobile telephone networks", *Information Economics and Policy*, 11(3), 297-305.
- Anderson J.E., D. Marcouiller (2002), "Insecurity and the Pattern of Trade: An empirical Investigation", *Review of Economics and Statistics*, 84(2), 342-252.
- Anderson J.E., E. van Wincoop (2003), "Gravity with Gravitas: A Solution to the Border Puzzle", *American Economic Review*, 93(1), 170-192.
- Anderson S.W., Daly J.D., Johnson M.F. (1995), "The Value of Management Control Systems: Evidence on the Market Reaction to ISO 9000 Quality Assurance Certificates", *University of Michigan Business School, Working Paper #9501-16*.
- Anderson S.W., Daly J.D., Johnson M.F. (1999), "Why Firms seek ISO 9000 Certification: Regulatory Compliance or Competitive Advantage?", *Production and Operations Management*, 8(1), 28-43.
- Arellano M., S. Bond (1991), "Some Test of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", *Review of Economic Studies*, 58(2), 277-297.
- Arthur W.B. (1989), "Competing Technologies, Increasing Returns, and Lock-in by Historical Events", *Economic Journal*, 99(394), 116-131.
- Bass F.M. (1969), "A new product growth model for consumer durables", *Management Science* 15(5), 215-227.
- Bass F.M., T.V. Krishnan, and D.C. Jain (1994), "Why the Bass Model Fits without Decision Variables", *Marketing Science* 13(3), 203-223.
- Becker G.S. (1991), "A Note on Restaurant Pricing and Other Examples of Social Influence on Price", *Journal of Political Economy* 99(5), 1109-1116.
- Beggs A., P. Klemperer (1992), "Multi-period competition with switching costs", *Econometrica* 60(3), 651-666.

- Bénézech D., Lambert G., Lanoux B., Lerch Ch., Loos-Baroin J. (2001), "Completion of knowledge codification: an illustration through the ISO 9000 standards implementation process", *Research Policy*, 30(9), 1395-1407.
- Besen S.M., J. Farrell (1994), "Choosing how to Compete: Strategies and Tactics in Standardization", *Journal of Economic Perspective*, 8(2), 117-131.
- Blind K. (2001), "The Impact of Innovations and Standards on Trade of Measurement and Testing Products: Empirical Results of Switzerland's Bilateral Trade Flows with Germany, France and the UK", *Information Economics and Policy*, 13(4), 439-460.
- Blind K. (2002), "Driving Forces for Standardization at Development Organizations", *Applied Economics*, 34(16), 1985-1998.
- Blonski M. (2002), "Network externalities and two-part tariffs in telecommunication markets", *Information Economics and Policy* 14, 95-109.
- Bousquet A., M. Ivaldi (1997), "Optimal pricing of telephone usage: An econometric implementation", *Information Economics and Policy* 9(3), 219-239.
- Bresnahan T.F. (1989), "Empirical Studies of Industries with Market Power" in Schmalensee R. and R.D. Willig (eds.), "Handbook of Industrial Organization", Volume II, *Elsevier Science Publishers*, Amsterdam.
- Brynjolfsson E., C.F. Kemerer (1996), "Network Externalities in Microcomputer Software: An Econometric Analysis of the Spreadsheet Market", *Management Science* 42(12), 1627-47.
- Cabral L.M.B. (1990), "On the adoption of innovations with 'network' externalities", *Mathematical Social Sciences* 19, 299-308.
- Casella A. (1996), "Free Trade and Evolving Standards", in Bhagwati J., Hudec R., eds., *Fair Trade and Harmonization: Prerequisites for Free Trade?*, vol.1, MIT Press, Cambridge, 119-156.
- Cusumano M.A., Y. Mylonadis, S.R. Rosenbloom (1992), "Strategic Maneuvering and Mass-Market Dynamics: The Triumph of VHS over Beta", *Business History Review*, 66(1), 51-94.
- David P.A. (1985), "Clio and the Economics of QWERTY", *American Economic Review*, 75(2), 332-337.
- Deardorff A.V. (1998), "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?" in J.A. Frankel, ed., *The regionalization of the world economy*, University of Chicago Press, Chicago, 7-22.

- Dekimpe M.G., P.M. Parker, and M. Sarvary (1998), "Staged Estimation of International Diffusion Models", *Technological Forecasting and Social Change* 57, 105-132.
- Dhebar A., S. S. Oren (1985), "Optimal dynamic pricing for expanding networks", *Marketing Science* 4(4), 336-351.
- Docking D.S., R.J. Downen (1999), "Market Interpretation of ISO 9000 Registration", *Journal of Financial Research*, XXII(2), 147-160.
- Dranove D., N. Gandal (2003), "The DVD-vs.-DIVX Standard War: Empirical Evidence of Network Effects and Preannouncement Effects", *Journal of Economics and Management Strategy*, 12(3), 363-386.
- Economides N. (1996), "The Economics of Networks", *International Journal of Industrial Organization* 14(6), 673-699.
- Economides N., Ch. Himmelberg (1995), "Critical Mass and Network Size with Application to the US Fax market", Stern School of Business, New York University *Discussion Paper* EC-95-11.
- Economides N., L.J. White (1994), "Networks and Compatibility: Implications for Antitrust", *European Economic Review* 38(3/4), 651-662.
- Farrell J., C. Shapiro (1988), "Dynamic competition with switching costs", *RAND Journal of Economics* 19(1), 123-137.
- Farrell J., G. Saloner (1985), "Standardization, Compatibility, and Innovation", *RAND Journal of Economics*, 16(1), 70-83.
- Farrell J., G. Saloner (1986a), "Installed Base and Compatibility: Innovation, Product Preannouncements and Predation", *American Economic Review*, 76(5), 940-955.
- Farrell J., G. Saloner (1986b), "Standardization and Variety", *Economics Letters*, 20(1), 70-74.
- Farrell J., N. Gallini (1988), "Second-Sourcing as a Commitment: Monopoly Incentives to Attract Competition", *Quarterly Journal of Economics*, 103(4), 673-694.
- Farrell J., P. Klemperer (2001), "Coordination and Lock-In: Competition with Switching Costs and Network Effects", mimeo.
- Frankel J.A, Stein E., Wei S.-J. (1997), *Regional Trading Blocks in the World Economic System*, Institute for International Economics, Washington, DC.
- Freund C.L., D. Weinhold (2004), "The effect of the Internet on international trade", *Journal of International Economics*, 62, 171-189.

- Gandal N. (1994), "Hedonic Price Indexes for Spreadsheets and an Empirical Test for Network Externalities", *RAND Journal of Economics* 25(1), 160-170.
- Gandal N. (1995), "Competing Compatibility Standards and Network Externalities in the PC Software Market", *Review of Economics and Statistics* 77(4), 599-608.
- Gandal N., Greenstein S., D. Salant (1999), "Adoptions and Orphans in the Early Microcomputer Market", *Journal of Industrial Economics* 47(1), 87-105.
- Gandal N., Kende M., and R. Rob (2000), "The dynamics of technological adoption in hardware/software systems: the case of compact disc players", *RAND Journal of Economics* 31(1), 43-61.
- Glick R., A.K. Rose (2002), "Does currency union affect trade? The time-series evidence", *European Economic Review*, 46, 1125-1151.
- Goolsbee A., P.J. Klenow (2002), "Evidence on Learning and Network Externalities in the Diffusion of Home Computers", *Journal of Law and Economics*, 45(2), 317-344.
- Granger C.W.J., T. Teräsvirta (1993), "Modelling Nonlinear Economic Relationships", *Oxford University Press*, New York.
- Granovatter M., R. Soong (1986), "Threshold model of interpersonal effects in consumer demand", *Journal of Economic Behavior and Organization* 7, 83-99.
- Greenstein S. (1993), "Did installed base give an incumbent any (measurable) advantage in federal computer procurement?", *RAND Journal of Economics* 24(1), 19-39.
- Gruber H., F. Verboven (2001), "The diffusion of mobile telecommunications services in the European Union", *European Economic Review* 45, 577-588.
- Hannan E. J., J. Rissanen (1982), "Recursive estimation of mixed autoregressive-moving average order", *Biometrika* 69(1), 81-94.
- Hatanaka M. (1974), "An efficient two-step estimator for the dynamic adjustment model with autoregressive errors", *Journal of Econometrics* 2, 199-220.
- Horsky D. (1990), "A Diffusion Model Incorporating Product Benefits, Price, Income and Information", *Marketing Science* 9(4), 342-365.
- Hummels D., J. Levinsohn (1995), "Monopolistic Competition and International Trade: Reconsidering the Evidence", *Quarterly Journal of Economics* 110(3), 799-836.
- ISO (2001), *Quality Management Principles*, ISO Central Secretariat, Geneva.
- ISO (2002), *The ISO Survey of ISO 9000 and ISO 14000 Certificates*, Eleventh cycle – 2001, ISO Central Secretariat, Geneva.

- Jain D.C., R.C. Rao (1990), "Effect of Price on the Demand for Durables: Modeling, Estimation, and Findings", *Journal of Business & Economic Statistics* 8(2), 163-170.
- Jarmin R.S. (1994), "Learning by Doing and Competition in the Early Rayon Industry", *Rand Journal of Economics*, 25, 441-454.
- Katz M. L., C. Shapiro (1994), "Systems Competition and Network Effects", *Journal of Economic Perspectives* 8(2), 93-115.
- Katz M.L., C. Shapiro (1985), "Network Externalities, Competition, and Compatibility", *American Economic Review* 75(3), 424-440.
- Katz M.L., C. Shapiro (1986a), "Technology Adoption in Presence of Network Externalities", *Journal of Political Economy*, 94(4), 822-841.
- Katz M.L., C. Shapiro (1986b), "Product Compatibility Choice in a Market with Technological Progress", *Oxford Economic Papers*, Special Issue on the New Industrial Economics.
- Katz M.L., C. Shapiro (1992), "Product Introduction with Network Externalities", *Journal of Industrial Economics*, 40(1), 55-84.
- Katz M.L., C. Shapiro (1994), "Systems Competition and Network Effects", *Journal of Economic Perspectives* 8(2), 93-115.
- Kim H.-S., N. Kwon (2003), "The advantage of network size in acquiring new subscribers: a conditional logit analysis of the Korean mobile telephone market", *Information Economics and Policy* 15(1), 17-33.
- Klemperer P. (1995), "Competition when Consumers have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics and International Trade", *Review of Economic Studies* 62, 515-539.
- Leibenstein H. (1950), "Bandwagon, snob, and Veblen Effects in the theory of consumers' demand", *Quarterly Journal of Economics* 64, 183-207.
- Liebowitz S.J., S.E. Margolis (1990), "The Fable of the Keys", *Journal of Law and Economics*, 33(1), 1-25.
- Liebowitz S.J., S.E. Margolis (1994), "Network Externalities: An Uncommon Tragedy", *Journal of Economic Perspectives* 8(2), 133-150.
- Lindbeck A., S. Nyberg, and J.W. Weibull (1999), "Social norms and economic incentives in the welfare state", *Quarterly Journal of Economics* 114(1), 1-35.
- Moenius J. (2000), "Information versus Product Adaptation: The Role of Standards in Trade", *Northwestern University*, mimeo.

- Newey W., K. West (1987), "A Simple Positive Semi-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica* 55(3), 703-708.
- OECD Telecommunication Basket Definitions (2000), *Teligen*, Richmond, U.K.
- Ohashi H. (2003), "The Role of Network Effects in the US VCR Market, 1987-1986", *Journal of Economics & Management Studies*, 12(4), 447-494.
- Okada Y., K. Hatta (1999), "The interdependent telecommunications demand and efficient price structure", *Journal of Japanese and International Economies* 13(4), 311-335.
- Parker P.M., L.-H. Röller (1997), "Collusive conduct in duopolies: multimarket contact and cross-ownership in the mobile telephone industry", *RAND Journal of Economics* 28(2), 304-322.
- Rauch J.E. (1999), "Networks versus markets in international trade", *Journal of International Economics*, 48(1), 7-35.
- Rauch J.E., V. Trindade (2002), "Ethnic Chinese Networks in International Trade", *Review of Economics and Statistics*, 84(1), 116-130.
- Reiss P. C., F. A. Wolak (2002), "Structural econometric modeling: rationales and examples from Industrial Organization", mimeo.
- Rohlfs J. (1974), "A theory of interdependent demand for a communications service", *Bell Journal of Economics* 5(1), 16-37.
- Röller L.-H., C. Wey (2003), "Merger Control in the New Economy", *Netnomics*, 5(1), 5-20.
- Rysman M. (2002), "Competition Between Networks: A Study of the Market for Yellow Pages", Boston University, *Industry Studies Project Working Paper* #104.
- Saloner G., A. Shepard (1995), "Adoption of Technologies with Network Effects: An Empirical Examination of the Adoption of Automated Teller Machines", *RAND Journal of Economics*, 26(3), 479-501.
- Schoder D. (2000), "Forecasting the success of telecommunication services in the presence of network effects", *Information Economics and Policy* 12, 181-200.
- Shapiro C., H.R. Varian (1999), "Information Rules", *Harvard Business School Press*, Boston.
- Swann P., Temple P., Shurmer M. (1996), "Standards and Trade Performance: The UK Experience", *Economic Journal*, 106(438), 1297-1313.
- White H. (1980), "A Heteroscedasticity Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity", *Econometrica* 48(4), 817-838.

Wooldridge J.M. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge and London.

Hiermit erkläre ich, dass ich außer der im Literaturverzeichnis angeführten Literatur keine weiteren Hilfsmittel benutzt habe. Außer den unter „Acknowledgements“ genannten Personen habe ich keine weitere Hilfe erhalten. Ich bezeuge durch meine Unterschrift, dass meine Angaben über die bei der Abfassung meiner Dissertation benutzten Hilfsmittel, über dir mir zuteil gewordene Hilfe sowie über frühere Begutachtungen meiner Dissertation in jeder Hinsicht der Wahrheit entsprechen.

Berlin, den 06. Oktober 2004

Michal Grajek