Schade, Christian; Köllinger, Philipp

Working Paper

Adoption of e-business: patterns and consequences of network externalities

Papers, No. 2004,05

Provided in Cooperation with:
CASE - Center for Applied Statistics and Economics, Humboldt University Berlin

Suggested Citation: Schade, Christian; Köllinger, Philipp (2004) : Adoption of e-business: patterns and consequences of network externalities, Papers, No. 2004,05, Humboldt-Universität zu Berlin, Center for Applied Statistics and Economics (CASE), Berlin

This Version is available at:
http://hdl.handle.net/10419/22179

Adoption of e-business:

patterns and consequences of network externalities

Philipp Köllinger (1) and Christian Schade (2)

(1) German Institute for Economic Research (DIW Berlin), Königin-Luise-Str. 5, 14195 Berlin, Germany; Email: pkoellinger@diw.de, Tel.: +49-30-89789-618

(2) CASE – Center for Applied Statistics and Economics, Institute for Entrepreneurial Studies and Innovation Management, School of Business and Economics, Humboldt-University Berlin, Spandauer Str. 1, 10178 Berlin, Germany; Email: schade@wiwi.hu-berlin.de, Tel.: +49-30-2093-5904

We thank the European Commission, DG Enterprise, for granting us access to the e-business watch database, DIW Berlin for their support, as well as Maria Minniti, Sönke Albers, and Pio Baake for their helpful comments and suggestions. All errors are ours.

Adoption of e-business:

patterns and consequences of network externalities

Abstract

The paper analyzes the adoption of various e-business technologies. Strong empirical evidence is found for the existence of increasing returns to adoption due to indirect network externalities between related technologies. If a company is close to the technological frontier, its probability of adoption increases. The empirical analysis is based on more than 5,000 observations from a cross-sectional European enterprise survey conducted in June 2002. A classification and regression tree (CART) is used to illustrate technological complementarities and their effect for the adoption probability of a firm.

Keywords

TECHNOLOGY ADOPTION, PATH DEPENDENCE, CLASSIFICATION TREES

JEL Classifications

O30; L29; C14

Introduction

Technological progress is often associated with the invention of new technologies. However, only those innovations that are finally used lead to the realization of economic benefits. Already Schumpeter (1934) recognized that the diffusion process of major innovations is the driving force behind the business cycle, in particular the long run Kondratieff cycle (Stoneman, 1986). Accordingly, the diffusion of Internet-based technologies in firms has recently received much attention. Applications such as online sales, e-procurement, or supply chain management are expected to enable process innovations and efficiency gains on the user side and thus reduce variable costs and improve productivity. Adopters are frequently believed to gain competitive advantage over their rivals, which in turn can result in changes in market structures and profit levels (OECD, 2000; Brynjolfsson and Hitt, 2003). In addition, there is already evidence that investments into information and communication technologies spur long term growth (Jorgenson, 2001; Oliner and Sichel, 2000; Nordhaus, 2002).

However, the determinants of the adoption process are still somewhat unclear. As a matter of fact, even potentially beneficial technologies are not adopted by all firms immediately. Instead, diffusion is a dynamic process that features pioneer users, followers, and also a number of non-adopters. Various theories have been suggested to explain this. In the literature, the most prevalent are rank, stock, order, and epidemic effects. Also, uncertainty and technological interdependencies have recently been discussed. For an overview, see Stoneman (1983, 1986), Karshenas and Stoneman (1993), or Hall and Khan (2003).

Our research focuses on two related concepts: rank effects and technological complementarities. Both concepts can be explicitly linked to the literature on network effects.

Rank effects are a general concept that relates firm heterogeneity to adoption probability. The basic idea is that firms differ from each other in at least one relevant dimension such that the net present value of a technological innovation is higher for some firms than for others. This makes it possible to rank firms in terms of the benefit to be obtained from the use of the new technology. Firms that rank higher are expected to adopt more rapidly. Important dimensions of heterogeneity are e.g. firm size, R&D intensity, and market power (David, 1969, 1991; Davies, 1979).

One factor leading to rank effects are network externalities. Generally, a technology is said to have a network effect when the value of the technology increases with the number of components in the network. In the case of direct network effects, each user is identified with a component of the network and provides a direct externality to all other users by adding complementary links to the existing links (Economides, 1996; Shy, 1996; Katz and Shapiro, 1985). E.g., the value of a firm's internal e-business technology may increase with the number of employees that are connected to and make use of the technology. Examples are local area network (LAN), and knowledge management solutions. In this case, a technology will be more valuable to a large firm with many employees, potentially leading to early adoption of the technology by large firms.

In addition, there can be indirect network externalities between complementary technologies. In this case, the components of the network are the technologies themselves. The size and the value of the network is determined by the number of

connected, complementary technologies and the number of users connected to each of the technologies (Economides, 1996). If such indirect network externalities prevail, the installed base of technologies in a firm will have an influence on the gross return expected from an additional technology and consequently on the likelihood of adoption. Complementarities arise if technologies are either directly or indirectly compatible. Direct compatibility can e.g. be observed between hardware and software. In this case, one technology is a prerequisite for the functioning of another, or at least makes the additional technology more efficient. Indirect compatibility exists if technologies require similar, complementary inputs to function properly. This could be e.g. the general level of know-how among employees of a firm or their experience with computer-supported processes. In both cases, direct and indirect compatibility, the existence of one technology provides a positive externality for the adoption of another technology and gives rise to indirect network effects.

A number of authors have recently dealt with the influence of interactions between different technologies on the diffusion process (Arthur, 1989; Church and Gandal, 1993). The two articles closest to our research are the empirical studies conducted by Stoneman and Kwon (1994) and Colombo and Mosconi (1995). Stoneman and Kwon (1994) analyze the simultaneous diffusion of multiple process technologies, using a probit model on survey data from the UK engineering and metalworking industries that includes the date of adoption of five different technologies. Their results suggest that significant cross technology effects may exist and need to be taken into account in modeling the diffusion of either technology. The authors differentiate between complementary and substitute technologies. Their results indicate that the more complementary the technologies are, the greater the likelihood

that firms will adopt both technologies simultaneously. Along the same line of thought, Colombo and Mosconi (1995) also analyze the diffusion of multiple technologies, employing a hazard rate model on a sample of firms from the Italian metalworking industry. The technologies considered originate from the Flexible Automation paradigm and the design/engineering spheres (CAD/CAM etc.), respectively. They pay particular attention to technological complementarities and learning effects associated with experience of previously available, related technologies. Their study confirms that technological synergies and cumulative learning by using effects are key determinants to a firm's adoption behavior. The legacy of a firm's technological history is found to greatly affect adoption choices. Colombo and Mosconi imply that the diffusion of innovations should be studied as a path dependent, evolutionary phenomenon, where firm heterogeneity is both a cause and an effect of technology adoption.

We extend this line of research, focusing on technological innovations that are based on the Internet. We pay particular attention to the influence of indirect network effects that emerge as a consequence of technological complementarities on the adoption behavior of firms. The main objective of this paper is to provide new empirical results and insights that can serve to enrich our understanding of the phenomenon. Also, our results contribute to the growing economic literature on ICT and link it to the literature on innovation. In addition, we introduce a classification and regression tree (CART) as a sophisticated, yet intuitively appealing method for analyzing adoption patterns when interdependencies between covariates exist.

Technological paradigms and trajectories

Rank effects may not be perceived as an explicit and sufficient concept of technological progress. To understand technological progress, we need both a concept for the rate and the direction of development. In addition to rank effects, which concern the rate of development, we find it useful to also consider the literature on technological paradigms and trajectories (Dosi, 1982), which provide a concept for the direction of development. Dosi's (1982) theory of technological paradigms is related to our above thoughts on network externalities. Dosi suggests that in broad analogy to the Kuhnian definition of a scientific paradigm (Kuhn, 1962), technological paradigms can be defined. A technological paradigm is a model or pattern of a solution to *selected* technological problems, based on *selected* principles derived from the natural sciences and on *selected* material technologies. A cluster of related concrete technological solutions can be associated with each technological paradigm, such as nuclear technologies, biotechnologies, or Internet technologies. Dosi calls the pattern and direction of progress based on a technological paradigm a *trajectory.* Technology, in this view, includes a perception of a limited set of possible technological alternatives and of notional future developments. We can think of the outer limits of a trajectory as the optimal combination of all relevant technological and economic variables, so to speak the "production possibility frontier" with respect to a given technological paradigm.

Note that numerous technological trajectories can exist in parallel. Also, trajectories can be more or less general and more or less powerful. In addition, there might be complementarities among trajectories because they require complementary forms of knowledge, experience, skills etc. Dosi points out that progress along a trajectory is

likely to retain some cumulative features, i.e. network externalities. The probability of future advances is hence related to the position that one (a firm or a country) already occupies vis-à-vis the existing technological frontier.

Following this conceptual framework, we define e-business as a cluster of related technological innovations that are jointly based on the Internet. In this sense, e-business is a technological paradigm with a very general scope, because its “normal problem solving tools” are applicable in various sectors, firms, and regions. The normal course of development along the e-business trajectory starts with the non-availability of any technology from the e-business cluster within a firm or country, progresses with the adoption of various technologies, and possibly ends with reaching the possibility frontier, i.e. the optimal combination of all technological and economically relevant parameters. Note that this is not a deterministic process. Not all firms need necessarily reach the production possibility frontier with respect to a given technological paradigm.

Firms that invest in a technology from a certain cluster usually also have to invest in complementary inputs, such as human capital (training, hiring, accumulation of experience and know-how) or re-organization of processes and structures. Brynjolfsson and Hitt (2003) have confirmed the importance of complementary investments for the case of computerization in firms. Thus, as far as complementarities prevail, the marginal benefits from adoption of a technology are greater for firms that have previously adopted other related technologies. This should result in a more rapid diffusion of technologies in firms that are already experienced users of related technologies. The closer a firm is to the technological frontier, the higher the likelihood that it will make future advances along the technological

trajectory. In this view, technological development is a path dependent process where current choices of technologies become the link through which prevailing economic conditions may influence the future dimensions of technology, knowledge, and economic opportunities (Ruttan, 1997).

The remaining parts of the paper provide new empirical evidence and further insights into the phenomenon outlined above. First, we illustrate the data on which the analysis is based and define an appropriate cluster of related Internet technologies. Then, we describe a micro-level adoption model that incorporates rank effects and includes the influence of other related technologies. Subsequently, we present the regression results for a number of technologies from the cluster. Finally, we introduce a classification tree (CART) for one particular technology, e-learning, to explore potential reasons for different adoption probabilities of different firms that may emerge as a consequence of rank effects. CART allows us to identify clusters of firms that exhibit significantly different adoption probabilities and characteristics. Also, CART detects cumulative patterns among the predictor variables, providing us with intuitively appealing and insightful details about technological interdependencies.

The Data

The data used for this analysis originates from the first enterprise survey of the E-business Market W@tch, a research project sponsored by the European Commission. The first survey round was conducted in summer 2002 among almost 10,000 firms, covering 15 industry sectors across 15 member states of the European

Union[1]. The purpose of the questionnaire was to measure the uptake and impact of e-business technologies.

The dataset contains basic information about each company, e.g. size class, sector, country, and turnover development. The majority of variables relates to the availability and usage of various Internet-based technologies. In addition, companies were asked about their IT training efforts. Also, various questions relate to the perceived importance and impact of e-business at the firm level. The extensive coverage of e-business technology parameters in the survey makes the dataset predestined to test for the existence of technological complementarities and rank effects.

The survey was conducted in all 15 sectors only in the four largest European member states (France, Germany, Italy, and the UK). In the smaller countries only five to six sectors were included in the survey. Therefore, our analysis is limited to the EU4 which enables a homogeneous sector coverage to eliminate sample selection bias. This reduces the number of relevant observations to 5,917.

Also, we focus our analysis only on companies that fulfill the basic technological requirements for engaging in any kind of e-business activity. Firms that do not have computers or Internet access and do not use the WWW and email are filtered out. This reduces the number of relevant observations to 4,852.

---

[1] The precise definition of the sectors included in the survey can be found on the website of the project at http://www.ebusiness-watch.org. The questionnaire and information about how to obtain the dataset can be requested from the authors.

The data we use consists of qualitative variables only. We recoded all relevant technology and control variables as dummies[2]. Unfortunately, the data has no time dimension. It only measures the degree of e-business uptake in summer 2002. However, for the purpose of identifying factors and patterns that influence adoption at that point in time, the data proves useful.

Given the data from the e-business watch, we define a cluster of 25 presumably complementary technologies that are jointly based on the Internet (see Table 1). Each of these solutions serves a different purpose for supporting processes and information flows within a company, or between a company and its environment, including customers, suppliers, co-operation partners and the general public. Some of the technologies are part of the IT infrastructure of a company and can be used for various purposes (e.g. LAN or Extranet). Others are special software solutions that support specific processes (e.g. e-learning, CMS, SCM).

[2] YES=1, NO=0. For some questions firms could also answer "don't know". The proportion of "don't know" answers was usually around 5 per cent. To avoid missing value problems, we also coded these answers as "0".

Table 1 – Cluster of related technologies, based on the Internet

Wide Area Network (WAN)
Local Area Network (LAN)
Extranet
Intranet
Content Management System (CMS)
Online Banking
E-Learning
Internet-based Human Resource Management (HRM)
Tracking working hours online
Automating travel cost reimbursement online
Sharing documents / performing collaborative work online within the company
Use of an Application Service Provider (ASP)
Knowledge Management System (KMS)
Customer Relationship Management (CRM)
Posting job vacancies online
Supply Chain Management (SCM)
Negotiating contracts online
Exchanging documents with costumers online
Exchanging documents with suppliers online
Managing capacities online
Forecasting product demand online
Designing new products online
Purchasing online
Selling online
Participation in e-marketplaces

Increasing Returns to Technology Adoption

To illustrate the existence of the "cumulative features" of development along the trajectory of e-business, we introduce a simple formal framework based on investment-theoretic considerations.

Let *N* be the number of heterogeneous firms that compete in a market with perfect information. We focus on the initial purchase of a new technology and abstract from intra-firm diffusion and from the level of use of the technology by the acquirer. Each firm $i = 1...N$ is characterized by a vector of $\overline{x_i}$ individual covariates. A cluster of *K* related, non-substitutable technologies exists. The acquisition of technology $j = 1...K$ from this cluster yields a present value of g_{ij} to firm *i*. We analyze the adoption of

each technology *j* from the cluster separately. The purchase of the technology is costly and consists of two components:

- the cost of technology p_{ij} (e.g. hardware, software);
- the cost for complementary investments into human capital, process re-engineering, and organizational change c_{ij}.

For each technology *j*, denote the total number of other adopted related technologies from the cluster in the firm by k_i. A higher position on the trajectory can simply be defined by $k_i' > k_i$. The total cost of adoption C_{ij} can vary among firms and is specified as

(1) $C_{ij} = p_{ij}(\bar{x}_i, k_i) + c_{ij}(\bar{x}_i, k_i)$.

It is a function of the individual characteristics of the firm $\overline{x_i}$ and the position of the firm upon the relevant technological trajectory k_i. If technologies require similar complementary investments that lead to a reduction in c_{ij}, or if bulk discounts on p_{ij} can be achieved, then technologies are said to be indirectly compatible and

(2) $C_{ij}(\overline{x_i}; k_i') < C_{ij}(\bar{x}_i; k_i)$.

It follows that the net present value G_{ij} of the technology depends on the individual characteristics of the firm, and explicitly also on the number of other installed, related technologies in the firm:

(3) $G_{ij} = g_{ij}(\bar{x}_i; k_i) - C_{ij}(x_i; k_i)$.

The present value g_{ij} could also depend on k_i if technologies are directly compatible, where the use of one technology makes the other directly more efficient (e.g. LAN and Knowledge Management). A profit maximizing firm will adopt technology *j* at time *t* if $G_{ij} > 0$. If increasing returns to adoption exist, $G_{ij}(\bar{x}_i; k_i) < G_{ij}(\bar{x}_i; k_i')$. This means that, ceteris paribus, a firm that is on a higher position upon the trajectory will expect a higher present value and / or lower cost of implementation from the adoption of an additional technology from the associated cluster.

Define y_{ij} to indicate whether a firm *i* has adopted technology *j* at time *t* as

(4) y_{ij} = 1 if firm has adopted and y_{ij} = 0 otherwise.

Firms adopt, if the non-observable latent variable y_{ij}^* exceeds a critical value:

(5) $y_{ij}^* > \bar{y}^* \longrightarrow y_{ij} = 1$.

Given (3), we can specify

(6) $y_{ij}^* = \alpha + \overline{\beta x_i} + \gamma k_i + \varepsilon_i$.

If increasing returns to adoption exist, γ should be significant and positive. Given that diffusion processes can often be well described by a logistic function (Mansfield, 1961), we assume that the error terms ε_i are identically and independently distributed following a logistic probability density function. Then we can write the probability of firm *i* to adopt technology *j* as

(7) $y_{ij}^* = P(y_{ij} = 1 \mid \overline{x_i}, k_i) = \dfrac{1}{1 + \exp(-\alpha - \overline{\beta x_i} - \gamma k_i)}$.

The unknown parameters α, $\overline{\beta}$ and γ can be estimated with a logit regression. The considered cluster of *K* related technologies is given by Table 1. We ran regressions for each of the 25 technologies. Table 2 shows regression results for five arbitrarily chosen technologies from the cluster.[3] In each regression, k_i is the total number of installed technologies from the cluster within a firm, excluding the technology under examination. In addition to k_i we control for home country, size class, sector membership, and whether a firm has more than one establishments. The reference categories are Germany, 50-249 employees, tourism sector, and one establishment. The regression results clearly confirm the existence of technological complementarities and increasing returns to adoption. Also, it appears that this phenomenon is quite persistent. In all 25 cases, γ is clearly positive and significant on the 99% confidence level or above. This means that the adoption probability for each of these technologies increases significantly with the total number of other related technologies being used in the firm. In other words, the more advanced a firm already is on the trajectory of e-business, the more likely it is to "go another step" and vice versa. On the grounds of this observation, we can hypothesis a growing "digital divide" among firms: There are pioneers with very timely adoption of many e-business technologies and other firms that never adopt any such technology. Keeping in mind that IT and e-business applications are usually associated with lower variable costs and thus higher productivity, this could have important economic consequences. It would be desirable to test this hypothesis in a longitudinal dataset once available.

[3] The remaining regression results are available from the authors upon request.

Table 2 – Factors influencing adoption probabilities: significant logit regression results

	E-Learning	**Online Sales**	**Online Purchasing**	**CRM**	**KMS**
Co-variables					
Sector					
Food		-1.929 (60.978)			
Publishing		- .970 (30.082)			
Chemicals	- .542 (4.615)	-1.936 (69.824)	-.395 (4.913)	.639 (3.369)	1.235 (4.297)
Metal Products		-2.233 (82.022)	-.297 (3.018)		
Machinery	- .468 (3.910)	-2.276 (88.306)			
Electronics		-1.966 (86.795)			1.147 (3.948)
Transport Equipment		-1.718 (58.939)			1.261 (4.464)
Retail		- .942 (22.795)			
Monetary Services	.386* (3.415)	-1.223 (43.979)			1.212 (4.399)
Insurances		-1.266 (35.881)	- .459 (6.116)		1.156 (3.530)
Real Estate		-1.688 (64.749)			
ICT Services		-1.440 (62.191)	1.142 (40.212)		1.050 (3.474)
Business Services		-1.966 (84.350)	.273 (2.719)		1.375 (5.951)
Health Services		-2.304 (72.795)		-1.146 (4.037)	
Country					
France	- .404 (9.409)	-.721 (33.262)	-1.432 (231.860)	- .343 (3.327)	
Italy	.256 (4.755)	-.618 (26.310)	-1.443 (248.316)	- .431 (5.272)	
UK	.464 (17.75)	-.253 (5.514)	- .411 (22.545)		
Size class					
1-49 empl.	.431 (17.697)	.298 (8.262)	.284 (13.594)	- .732 (21.050)	- .471 (4.994)
> 250 empl.	.244 (3.534)		- .348 (9.467)		
> 1 establishments					
Number of other related technologies (k_i)	.253 (396.847)	.165 (179.322)	.200 (339.860)	.111 (38.341)	.168 (55.513)
Constant	-3.673 (286.961)	-1.178 (47.503)	- .592 (14.931)	-3.318 (99.976)	-5.334 (85.341)

Model Diagnostics					
N	4,852	4,852	4,852	4,852	4,852
Nagelkerke R2	.217	.177	.248	.088	.101
-2 Loglikelihood	3775	3681	5712	1848	1200
Significance	.000	.000	.000	.000	.000
df	21	21	21	21	21
Table displays significant coefficients at the 95% confidence level, 90% confidence denoted by *. Wald statistic in ( ).					

Complementarities in detail: rank-effects reconsidered

Up to this point, we demonstrated the existence of increasing returns to adoption because of network externalities between related technologies. The remaining part of the paper presents empirical evidence for adoption patterns that emerge as a consequence of these effects. We explore possible combinations of technological and structural variables (rank effects) that lead to a higher or lower probability of adoption for one particular technology. We chose to present the results for e-learning because they are especially intuitive and thus easily interpreted. Naturally, the method we use can also be applied to all other technologies from the cluster.[4]

We use a classification and regression tree (CART) for this purpose. CART was first introduced by Breiman et. al. (1984). It can loosely be codified as a combination of non-parametric regression and cluster analysis. It is particularly well suited for our purposes because it detects higher order interdependencies between co-variables and avoids the problem of multicollinearity. In addition, by simultaneously identifying significant predictors and clusters that exhibit significant differences with respect to the dependent variable, CART provides us with a unique insight into adoption

[4] Additional results are available from the authors upon request.

patterns that can be identified in the data. The result, a “tree” presented in graphic form, is both parsimonious and easy to interpret.

CART has recently been used in numerous studies in the medical sciences (Zhang and Bracken, 1995; Zhang and Singer, 1999). However, to our knowledge, its application in an economic context is still novel. Therefore, we include a short description of the method at this point and a short technical introduction to CART in the appendix.

The basic idea of CART is to systematically split the dataset into homogeneous groups with respect to the dependent variable based on the best set of predictors. We derive the final tree in four steps.

In the first step, called *recursive partitioning*, the sample of subjects is systematically sorted into completely homogeneous subsets until a *saturated tree* is found. In our case, complete homogeneity means that a node contains either only adopters or non-adopters. The root node of a tree contains the sample of subjects from which the tree is grown. Then, based on the parameter value that is most predictive for the outcome, the root node is split into two daughter nodes that now form a second layer of the tree. All nodes in the same layer constitute a partition of the root node. The process of splitting nodes is continued and the partition becomes finer and finer as the layer gets deeper and deeper. For each split, CART considers the entire set of available predictor variables to determine which one maximizes the homogeneity of the following two daughter nodes. This is a hierarchical process that reveals interdependencies between covariates. Also, a predictor might show up numerous times in different parts of the tree. Each case of the sample is sorted into one of the daughter nodes at each layer of the tree, according to the splitting rule that was used.

Those subsets that are not split are called terminal nodes. When a case finally moves into a terminal subset, its predicted class is given by the class label attached to that terminal subset (e.g. “adopter {Y=1}” or “non-adopter {Y=0}” for node *t*). The process is continued until the nodes are completely homogeneous and cannot be split any further. This is the saturated tree. The saturated tree is usually too large to be useful. In the worst case, it is trivial because each terminal node could consist of just one case. The resulting model is obviously subject to severe over-fitting problems. Therefore, we must find a nested sub-tree of the saturated tree that exhibits the best “true” classification performance and satisfies statistical inference measures.

To proceed, we generate a series of nested optimal sub-trees of the saturated tree in the second step. This process is called *pruning*. We use the cost-complexity pruning algorithm suggested by Breiman et. al. (1984), which ensures that a uniquely best sub-tree can be found for any given tree complexity.

In a third step, we must select one of the trees from the pruning sequence. The solutions lies in finding an honest estimate for the true classification performance and selecting the sub-tree that minimizes the estimated true misclassification costs. This is usually done with an independent test sample, boot-strapping, or cross-validation. We choose a 20-fold cross validation procedure because it makes better use of the information contained in the original dataset than the independent test sample method and outperforms bootstrapping in terms of reduced bias (Breiman et. al., 1984, pp. 72-78, 311-313).[5]

[5] We are using the software CART 5.0 by Salford Systems for the analysis.

Following these steps, we identify the best classifying tree. However, because we are mainly interested in interpreting the revealed structures, we must also ensure that the model satisfies the usual significance tests. We arrive at the final tree by calculating significance tests for all splits in the tree, dropping those splits (and their successors) that are not significant at the 95% confidence level or above.

For the analysis we are using the same dataset as before. In the e-business watch, e-learning is defined as the usage of online, Internet-based technologies to support employee training. We focus on firms that fulfill the necessary technological and organizational requirements to eventually be an e-learning adopter. Thus, firms that do not have the necessary basic infrastructure and ability to use the worldwide web and e-mail can be filtered out again. In addition, we also exclude firms from this analysis that do not offer any kind of computer training support to their employees. Firms that do not care about the basic computer skills of their work force obviously do not qualify for the rather advanced application of e-learning. The working sample for this analysis thus includes 4,098 firm observations, 801 of which are e-learning users (19.5%).

The results of our analysis are displayed in Figure 1, a detailed description of the relevant predictor variables is given in Table 3. Table 4 summarizes the average k_i's for each tree cluster. All variables that relate to rank effects in the dataset are included in the CART analysis (country, size class, sector, number of establishments, turnover development, equipment with other Internet technologies, IT training efforts, and some opinion statements reflecting the general attitude of firms towards e-business). In each tree node the number of e-learning adopters (top) and non-adopters (bottom) is given, as well as the ratio of adopters (percentage figure above

the node). The variable names below the nodes are the predictors that provide the best split for the node. Because all variables in the data set are of binary format (with 0=no and 1=yes), the split of each node is according to whether the predictor occurs or not.

The terminal nodes can be ordered according to the ratio of e-learning adopters they contain. The numbers below the terminal nodes indicate this order, with 1 being the most and 8 being the least e-learning affine segment in the data. We refer to these order number of the segments to describe and interpret them.

Figure 1 – CART for the adoption of e-learning

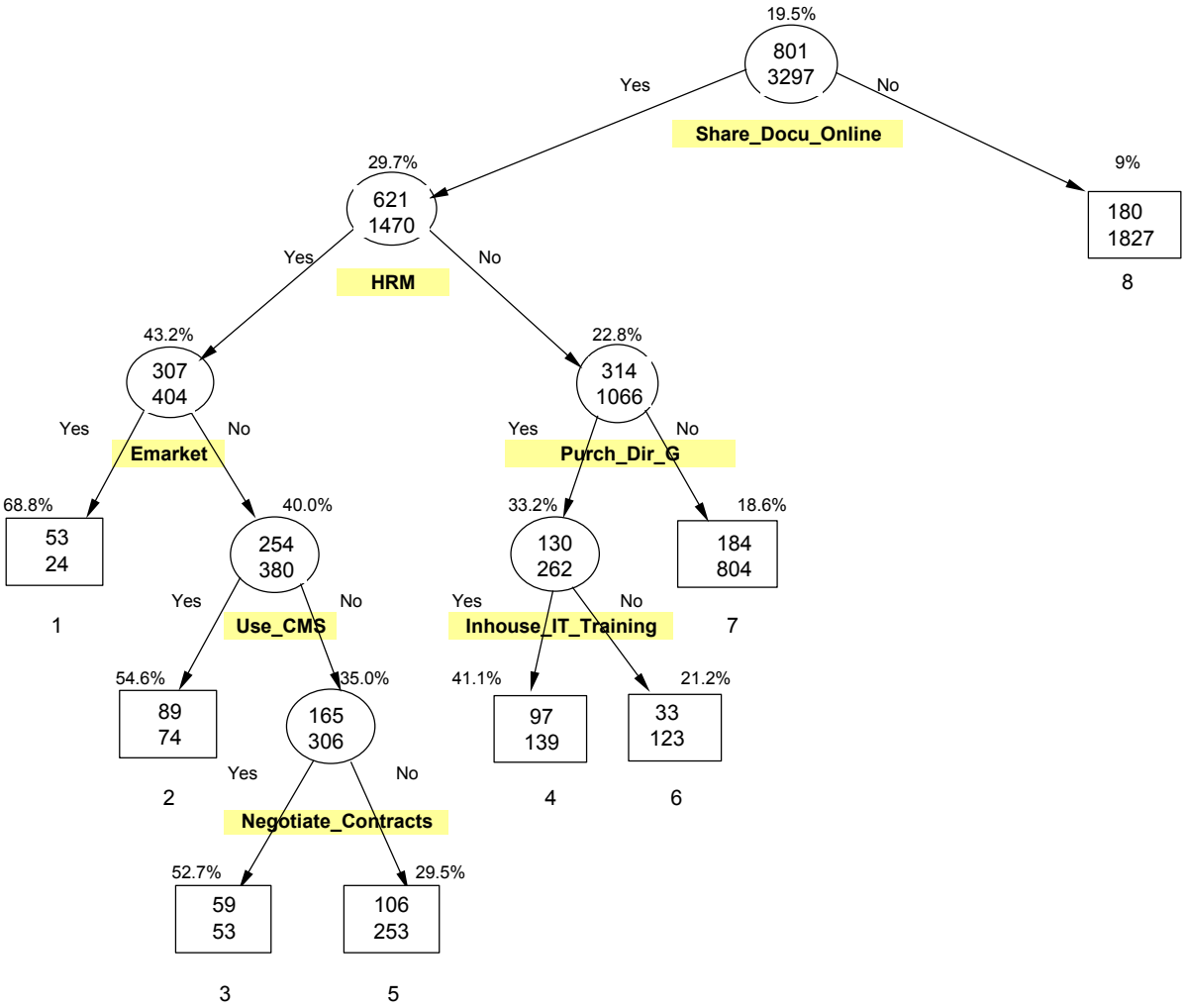

Table 3 – Description of relevant split variables

Predictors in Tree	Variable description
Share_Docu_ Online	Company uses online technologies to share documents with colleagues or to perform collaborative work in an online environment.
HRM	Company uses online technologies to support human resources management.
Emarket	Company trades goods or services through a B2B e-marketplace.
Purch_Dir_G	Company uses the Internet to purchase direct goods.
Use_CMS	Company uses a content management system for its webpage.
Inhouse_IT_ Training	Company provides in-house computer- or IT-training for its employees.
Negotiate_ Contracts	Company uses online technologies other than email to negotiate contracts.

Table 4 – Average number of installed complementary Internet technologies other than e-learning per firm (k_i)

	Total	E-Learning Adopters	E-Learning Non-Adopters
Tree 1	14.22 (2.92)	14.57 (2.88)	13.46 (2.93)
Tree 2	12.29 (3.0)	12.37 (2.89)	12.19 (3.16)
Tree 3	10.71 (2.61)	10.71 (2.57)	10.7 (2.68)
Tree 4	9.26 (2.73)	9.59 (2.86)	9.03 (2.62)
Tree 5	8.69 (2.54)	9.27 (2.74)	8.44 (2.42)
Tree 6	7.7 (2.41)	7.85 (1.68)	7.66 (2.57)
Tree 7	6.65 (2.71)	7.47 (2.76)	6.46 (2.67)
Tree 8	4.37 (2.7)	5.98 (3.25)	4.21 (2.59)
Total	6.38 (3.65)	8.9 (3.79)	5.77 (3.34)
Table displays means, standard deviations in ( ), N=4,098			

The final tree consists of 8 terminal nodes. CART uses 7 different predictor variables to construct the tree. Each of the terminal nodes exhibits different fractions of e-learning users. The most e-learning affine segment (number 1) contains almost 70% of adopters, whereas in the least e-learning affine segment (number 8) a fraction of only 9% uses e-learning. The terminal nodes each contain a different number of firms. Some of the nodes are rather small and describe rare, but statistically relevant

sub-groups (like number 1, which contains only 77 firms or 1.9% of the population), whereas others are very large (like number 8, which contains 2,007 firms or 49% of the population). Note that the impact of each predictor variable on the ratio of e-learning users can be followed along the tree branches. For example, the fraction of e-learning users increases from 19.5% (root node) to 29.7% for firms that share documents online. It again increases sharply if these firms also use an Internet-based Human Resource Management system. It is interesting to observe that all co-variables in the tree are good predictors only for a specific sub-set of the population, in interaction with specific predictors, and do not turn out to be relevant in other parts of the tree. This is one of the unique insights into the data structures revealed by CART.

Table 5 summarizes inference statistics for each split in the tree, listing the entropy impurity measure (see appendix), the relative resubstitution risk, and the according 95% confidence interval. The relative resubstitution risk is the probability of being an e-learning adopter if a subject is a member of one node divided by the probability of being an e-learning adopter if the subject is a member of the other node. For example, the two daughter nodes of the first split (Share_Docu_Online) have a resubstitution risk of 3.31. This means that the ratio of e-learning adopters is 3.31 times higher for those subjects that share documents online than for those that do not.

A split is significant if we can be sure that the ratio of e-learning adopters is not equal in both daughter nodes. Thus, the α-confidence level should not include 1. According to this criterion, all splits in the final tree are significant at the 95% confidence level or above.

Table 5: Inference statistical measures for splits in the tree

Split	Impurity of split	Relative resubstitution relative risk	95% confidence interval
Share_Docu_Online	.46	3.31	2.77 ; 3.96
HRM	.59	1.90	1.56 ; 2.31
Emarket	.67	1.72	1.03 ; 2.85
Purch_Dir_G	.52	1.78	1.37 ; 2.32
Use_CMS	.66	1.56	1.09 ; 2.24
Inhouse_IT_Training	.61	1.94	1.22 ; 3.09
Negotiate_Contracts	.63	1.78	1.16 ; 2.76

We complete the evaluation of the tree by analyzing its overall performance in terms of loglikelihood, significance of terminal nodes, and predictive performance (Table 6). For this purpose, we define dummy variables for all terminal nodes of the tree. For example, the dummy for segment 1 is set to 1 for the 77 firms in this segment, and zero otherwise. We run a logistic regression using only these tree dummies as predictors, with tree cluster 7 as reference category.

The clusters that exhibit either a very high or a very low ratio of e-learning adopters turn out to be excellent and highly significant predictors. For example, the odds of a segment 1 member being an e-learning adopter are 9 times higher than on average. On the other extreme, the odds of a segment 8 member being an e-learning user are 57% lower than on average. Overall, the tests demonstrate that CART returns several significant results and clusters that deserve closer interpretation.

Table 6 – Evaluation of terminal nodes

Variables in the equation			
Variable	*Odds Ratio*	*Coefficient*	*Significance*
Tree1	9.65	2.267	.000
Tree2	5.26	1.659	.000
Tree3	4.86	1.582	.000
Tree4	3.05	1.115	.000
Tree5	1.83	.605	.000
Tree6	1.17	.159	.454
Tree8	.43	- .843	.000
Constant		-1.475	.000
Model Diagnostics			
Significance	.000		
Nagelkerke R2	.182		
-2 Loglikelihood	3552.8		
df	7		

Table 7 shows how the tree segments correspond to the control variables (sector membership, size class, country of origin, and number of establishments). It can be seen that some significant correlations between the control variables and the tree segments prevail, however, they are by no means equivalent or trivial.

CART reveals a more complex relationship between the control variables and adoption behavior. E.g., large companies are over-proportionately represented in the highly e-learning affine clusters 1, 2, and 5, but cluster 4 remains entirely independent of size-class effects.

Table 7 – Significant correlations of tree segments with sectors, countries, and size classes

	Tree							
	1	2	3	4	5	6	7	8
Food								.0377
Publishing								
Chemicals								
Metal				- .0419				.0569
Machinery							-.0340	.0317
Electronics						.0332		- .0509
Transport Eq.								
Retail							- .0312	.0456
Tourism								.0513
Banks					.0393		.0635	- .0651
Insurances				- .0396		- .0355		
Real Estate	- .0388	- .0378					.0487	
Telcos & IT	.0540	.0849		.1149		.0804	- .0461	- .0958
Business Services							- .0341	
Health				- .0448				
France	- .357	- .0413	- .0312			- .0312	.0598	
Germany			- .0559		- .0485	.0755	- .0365	.0355
Italy				- .0512				.0344
UK	.0349		.0904		.0341			- .0495
1-49 empl	- .0428	- .1186	- .0690		- .1347	.0439	- .0386	.1741
50-249			.0518		.0643		.0466	- .0884
>250 empl	.0699	.1615			.1067	- .0351		- .1307
One establishment	- .0647	- .0990	- .0742		- .0896	.0545		.1445
> 1 establishments	.0652	.0998	.0717		.0908	-.0538		- .1442
All entries significant at 95%								

The results of the tree illustrate the importance of technological complementarities and their consequences. In fact, six of the seven relevant predictor variables in the tree directly relate to the usage of other e-business technologies. Other indicators in the dataset that reflected firm heterogeneity, such as size class, sector membership, or turnover development, are not used in the tree. The four variables with the highest predictor power with respect to e-learning (the variables in layers one and two) exclusively indicate the usage of other e-business technologies. This should not be mistaken to indicate an irrelevance of other factors leading to rank effects, such as firm size, sector, or country of origin. Indeed, several of these variables exhibited a

significant impact on e-learning adoption in the logit regression (see table 2). The reason why they do not show up in the tree lies in the CART method, which only uses the predictor that minimizes node impurity. The second best predictor that might even be closely related does not show up in the tree. Firm size, sectors, or country of origin are candidates for such factors, based on previous analysis. However, such relationships may not be as simple as already pointed out.

It has to be kept in mind that the usage of other e-business technologies as explanatory variables in the tree does not imply a simple causal relationship. From this cross-sectional dataset we cannot tell in which order a company has adopted various e-business technologies. For example, we do not know whether firms in cluster 1 have first adopted e-marketplaces or e-learning. Because of this we cannot say that e-marketplaces “explain” e-learning or vice versa. Thus, all variables in the model that relate to the usage of some other e-business technology have to be interpreted as a proxy for the general Internet competence of a firm, i.e. its position on the e-business trajectory.

The results of the tree contribute towards the perception of a growing “digital divide”. Moreover, we see that there are different paths to high adoption probability and that significant differences still prevail between the adopter segments.

Segment 1, which exhibits almost 70 per cent of e-learning users, can be referred to as fully Internet-enabled enterprises. The average number of other Internet technologies installed (k_i) in this segment is 14.22, the highest among all terminal nodes in the tree. Segment 1 is sufficiently characterized by just three predictor variables: It includes firms that share documents online, use Internet technologies to support human resource management functions (HRM), and use B2B online market

places to sell or purchase goods and services. At least HRM and B2B market places can be seen as rather advanced e-business applications that are not yet used by many companies. In other words, firms in this segment are already very advanced in the usage of Internet technologies. The complementarities between the technologies and the collected experience with these technologies seem to imply that these firms indeed expect lower implementation costs and higher benefits from e-learning. Large British firms from the telecommunications and computer services sector are over-proportionately represented in this cluster.

Firms in segments 2 and 3, which include more than 50% of e-learning users each, are comparable to segment 1. They are also characterized by an advanced degree of e-business technology usage, which makes e-learning attractive to them. These clusters include an over-proportionate number of medium-sized and large companies.

An interesting constellation appears in segment 4. The odds of being an e-learning adopter in this cluster are still 3 times higher than on average. This segment also contains firms that are familiar with basic Internet applications, but they are not as advanced in usage as segment 1, 2, and 3. The average k_i in this segment is just 9.26, less than what we observe in segment 1 (14.22), 2 (12.29) or 3 (10.71) respectively. Firms in segment 4 do not use HRM tools and most of them also do not use B2B online market places. They partially compensate for that by using the Internet to purchase direct goods. However, firms in this segment offer in-house computer training to their employees. Apparently, firms in this segment make a notable effort to invest into the IT competence of their employees. The adoption of e-learning corresponds with this objective. Firms from the telecommunication and

computer services sector are heavily represented in this group, regardless of their size.

The segment with the lowest rate of e-learning users (number 8) captures a major part of the sample population. 2,007 firms fall into this class, which is almost half of the sample. These companies have in common that they do not share documents online. This appears to be a very powerful proxy for the basic "e-readiness" of a company. Firms that do not use this rather simple form of Internet technology do not seem to be ready yet to adopt more complex solutions, such as e-learning. Consequently, they are more likely to adopt e-learning either later or never. Cluster 8 also features the lowest average k_i (6.38). This cluster is very typical for small firms from Germany and Italy. Classes 6 and 7 share a mixture of attributes from the characteristics of the more noticeable segments described above. Firms in these remaining classes exhibit e-learning adoption rates that are close to the average of the entire population.

Conclusion

We find strong empirical evidence for the existence of increasing returns to adoption due to technological complementarities. Our empirical results suggest that the positive externalities of related technologies on one another retain some cumulative features: the probability of adopting one particular kind of e-business technology generally increases with the number of e-business technologies that a company has already implemented. Thus, if a company is relatively close to the technological frontier, its probability of adoption increases and vice versa. This result raises the question whether we observe a growing "digital divide" among firms, regions, and

sectors. If so, this could have important consequences for market structures and economic development, if the introduction of e-business applications actually leads to lower variable costs and higher productivity. Also, our results reinforce the suggestion that the diffusion of innovations should be studied as a path dependent, evolutionary phenomenon, where firm heterogeneity is both a cause and an effect of technology adoption.

References

Arthur, W.B. (1989). “Competing technologies, increasing returns, and lock-in by historical events”, ECONOMIC JOURNAL, vol. 99, no. 394, pp. 116-131.

Breiman, L. et. al. (1984). *Classification and Regression Trees*, Belmont (CA-USA): Wadsworth International Group, Statistics/Probability Series.

Brynjolfsson, E. and Hitt, L.M. (2003). “Computing productivity: firm-level evidence”, MIT Sloan School of Management, Center for eBusiness@MIT, Paper 139, forthcoming in *Review of Economics and Statistics*.

Christensen, R. (1990). *Log-linear models*, New York et. al.: Springer.

Church, J. and Gandal, N. (1993). “Complementary network externalities and technological adoption”, *International Journal of Industrial Organization*, vol. 11, pp. 239-260.

Colombo, M.G. and Mosconi, R. (1995). “Complementarity and cumulative learning effects in the early diffusion of multiple technologies”, *Journal of Industrial Economics*, vol. 43, pp. 13-48.

David, P. (1969). “A contribution to the theory of diffusion”, *Center for Research in Economic Growth Research Memorandum*, no. 71, Stanford University.

David, P. (1991). *Behind the Diffusion Curve*, Oxford: Westview Press.

Davies, P.A. (1979). *The Diffusion of Process Innovations*, Cambridge: Cambridge University Press.

Dosi, G. (1982). “Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change”, *Research Policy*, vol. 11, pp. 147-162.

Economides, N. (1996). “The economics of networks”, *International Journal of Industrial Organization*, vol. 14, no. 6, pp. 673-700.

Hall, B.H. and Khan, B. (2003). “Adoption of new technology”, NBER working paper 9730.

Jorgenson, D.W. (2001). “Information technology and the U.S. economy”, *American Economic Review*, vol. 91, no. 1, pp. 1-32.

Karshenas, M. and Stoneman, P.L. (1993). “Rank, stock, order, and epidemic effects in the diffusion of new process technologies: an empirical model”, *RAND Journal of Economics*, vol. 24, no. 4, pp. 503-528.

Katz, M.L. and Shapiro, C. (1985). “Network externalities, competition, and compatibility”, *American Economic Review*, vol. 75, no. 3, pp. 424-440.

Kuhn, T. (1962). *The Structure of Scientific Revolutions*”, pp. 23-41, Chicago: Chicago University Press.

Mansfield, E. (1961). “Technical change and the rate of imitation”, *Econometrica*, vol. 29, no. 4, pp. 741-766.

Nordhaus, W. D. (2002). “Productivity growth and the new economy”, *Brooking Papers on Economic Activity*, no. 2 2002, pp. 211-245.

OECD (2000). Information Technology Outlook – ICT’s, E-Commerce and the Information Economy, Paris: OECD Publications Service.

Oliner, S.D. and Sichel, D.E. (2000). “The resurgence of growth in the late 1990’s: is information technology the story?”, *Journal of Economic Perspectives*, vol. 14, no. 4, pp. 3-22.

Pagano, M. and Gauvreau, K. (1993). *Principles of Biostatistics*, Belmont (CA): Duxbury Press.

Ruttan, V. (1997). “Induced innovation, evolutionary theory and path dependence: sources of technical change”, ECONOMIC JOURNAL, vol. 107, no. 444, pp. 1520-1529.

Schumpeter, J.A. (1934). *The Theory of Economic Development*, Cambridge: Harvard Press.

Sheskin, D. J. (2000). *Handbook of Parametric and Nonparametric Statistical Procedures*, 2nd edition, Boca Raton et. al.: Chapman & Hall.

Shy, O. (1996). “Technology revolutions in the presence of network externalities”, *International Journal of Industrial Organization*, vol. 14, no. 6, pp. 785-800.

Stoneman, P.L. (1983). *The Economic Analysis of Technological Change*, Oxford: Oxford University Press.

Stoneman, P.L. (1986). “Technological diffusion: the viewpoint of economic theory”, *Richerche Economiche*, XL, 4, pp. 585-606.

Stoneman, P.L. et. al. (1995). *Handbook of the Economics of Innovation and Technological Change*, Oxford and Cambridge: Blackwell Publishers..

Stoneman, P.L. and Kwon, M.J. (1994). „The diffusion of multiple process technologies“, ECONOMIC JOURNAL, vol. 104, no. 423, pp. 420-431.

Zhang, H. and Bracken, M. (1995). „Tree-based risk factor analysis of preterm delivery and small-for-gestational-age birth“, *American Journal of Epidemiology*, vol. 141, pp. 70-78.

Zhang, H. and Singer, B. (1999). *Recursive Partitioning in the Health Sciences*,New York et. al.: Springer.

Appendix: Classification and regression trees (CART)

Splitting nodes

A number of methods have been proposed to define the best split (Breiman, 1984, chapter 4). We have decided to use entropy impurity. The entropy criterion is related to the likelihood function. It tends to look for splits where as many levels as possible are divided perfectly or near perfectly. As a result, entropy puts more emphasis on getting rare characteristics right than e.g. Gini or Twoing.

Consider the following split, where *a, b, c,* and *d* are the number of subjects in the two daughter nodes:

Table 8 – Cross table for two daughter nodes

	Predictor	Adopter	Non-Adopter	
Left node (t_L)	$s_i = 1$	a	b	a+b
Right node (t_R)	$s_i = 0$	c	d	c+d
		a+c	b+d	n = a+b+c+d

Following Breiman et. al. (1984, pp. 94-102), the entropy impurity in the left daughter node is

(1) $$i(t_L) = -\frac{a}{a+b}\log\left(\frac{a}{a+b}\right) - \frac{b}{a+b}\log\left(\frac{b}{a+b}\right).$$

Likewise, the entropy impurity in the right daughter node is

(2) $i(t_R) = -\frac{c}{c+d}\log\left(\frac{c}{c+d}\right) - \frac{d}{c+d}\log\left(\frac{d}{c+d}\right).$

The impurity of the parent node consequently is

(3) $i(t) = -\frac{a+c}{n}\log\left(\frac{a+c}{n}\right) - \frac{b+d}{n}\log\left(\frac{b+d}{n}\right).$

The goodness of a split, *s*, is then measured by

(4) $\Delta I(s,t) = i(t) - P\{t_L\}i(t_L) - P\{t_R\}i\{t_R\}$

The goodness of a split is calculated for all available predictor variables, and the best predictor, which is the one with the highest $\Delta I(s,t)$, is selected.

This recursive partitioning process continues until the tree is saturated in the sense that the offspring nodes subject to further division cannot be split any further (e.g. when there is perfect homogeneity in the node). The resulting saturated tree is called T_0.

Pruning

The purpose of pruning is to find the right-sized tree, which should be a nested sub-tree of T_0. The right-sized tree should not be subject to over-fitting and insignificant splits, but detailed enough to exhibit a good classification performance. To begin, we need to define a concept to measure classification performance. Recall that CART predicts the outcome (e.g. adoption or non-adoption) based on the group membership of a subject. In the tree, each subject falls into exactly one terminal node. We choose a class assignment rule that assigns a class to every terminal node

$t \in \widetilde{T}$. In our application, node *t* is assigned "adopter {Y=1}" if $P\{Y=1|t\} \geq 0.5$ and vice versa. In this simple case, the expected cost resulting from any subject within a node is given by

(5) $r(t) = 1 - P(i \mid t)$,

where $P(i \mid t)$ is the percentage of misclassified subjects in a node.

Note that *r(t)* becomes smaller for any additional split. The formal proof is given by Breiman et. al. (1984, p. 95-96). Thus, *r(t)* is minimal for the saturated tree.

The classification performance of the entire tree is given by the quality of its terminal nodes

(6) $R(\mathrm{T}) = \sum_{t \in \widetilde{T}} P(t) r(t)$,

where $R(\mathrm{T})$ is the misclassification cost of all terminal nodes in the tree, $\widetilde{\mathrm{T}}$ the set of terminal nodes, and $P(t)$ the probability of a subject to fall into the terminal node t.

We are now ready to turn to the main idea of cost-complexity pruning (Breiman et. al., 1984, pp. 66-71): For any subtree $\mathrm{T} \leq \mathrm{T}_0$, define its complexity as $\left|\widetilde{\mathrm{T}}\right|$, the number of terminal nodes in T. Let $\alpha(\geq 0)$ be a real number called the complexity parameter and define the cost complexity of the entire tree as

(7) $R_\alpha(\mathrm{T}) = R(\mathrm{T}) + \alpha\left|\widetilde{\mathrm{T}}\right|$.

For any value of $\alpha(\geq 0)$, there is a unique smallest subtree of T_0 that minimizes $R_\alpha(\mathrm{T})$. The formal proof is in Breiman et. al. (1985, chapter 10). Thus, by gradually increasing α , a sequence of nested essential subtrees of T_0 can be constructed by

pruning off the weakest branches at each threshold level of α. Note that T_0 minimizes $R_\alpha(\mathrm{T})$ if $\alpha = 0$. If α becomes large enough, the root node becomes the optimal solution.

Selection of the best pruned tree using cross-validation

The classification performance $R(\mathrm{T})$ as specified in (6) is obviously biased and results in severe over-fitting. To select the best pruned tree, we need a more honest estimate of the true misclassification cost of the tree. Using cross-validation (Breiman et. al., 1984, pp. 75-78), we estimate $\hat{R}(\mathrm{T})$ by growing a series of *V* auxiliary trees together with the main tree grown on the learning sample Λ. The *V* auxiliary trees are grown on randomly divided, same sized subsets, $\Lambda_V, v = 1,...,V$, with the *v*-th learning sample being $\Lambda^{(v)} = \Lambda - \Lambda_v$ so that $\Lambda^{(v)}$ contains the fraction $(V-1)/V$ of the total data cases. For each *v,* the trees and their pruning sequence are constructed without ever seeing the cases in Λ_V. Thus, they can serve as an independent test sample for the tree $T^{(V)}(\alpha)$. The idea now is that for *V* large, $T^{(V)}(\alpha)$ should have about the same classification accuracy as $T(\alpha)$. If unit misclassification costs are used, and priors are data estimated as in our application, the estimated misclassification costs $\hat{R}(\mathrm{T})$ equal the proportion of misclassified test set cases in the *V* auxiliary trees. The best pruned tree is the one with the smallest $\hat{R}(\mathrm{T})$.

Significance of splits

Finally, the significance of each individual split in the selected tree can be tested following Sheskin (2000; section 16.6): Recall the notation from table 8. We calculate the resubstitution risk as

(8) $$r = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

The calculation of the confidence interval of *r* requires to compute the standard error of the two daughter nodes, which is given by

(9) $$SE_r = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}\,.$$

Since the sampling distribution of the resubstitution risk is positively skewed, a logarithmic scale transformation is employed in computing the confidence interval (Christensen, 1990; Pagano and Gauvreau, 1993). The α-confidence level is obtained by

(10) $$\left\{ e^{[\ln(r) - SE \bullet z_\alpha]}; e^{[\ln(r) + SE \bullet z_\alpha]} \right\},$$

where z_α is the tabled two-tailed *z* value for the $(1-\alpha)$ confidence level. For the 95% confidence level, the relevant .05 value is $z_{.05} = 1.96$. This test is computed for all splits in the tree that was selected from the pruning sequence after the cross-validation procedure.