Discussion Paper No. 06-085

Regional Versus Individual Aspects of the Digital Divide in Germany

Katrin Schleife

ZEW
Zentrum für Europäische Wirtschaftsforschung GmbH
Centre for European Economic Research

Discussion Paper No. 06-085

Regional Versus Individual Aspects of the Digital Divide in Germany

Katrin Schleife

first version: December 2006
substantially revised version: June 2008

Download this ZEW Discussion Paper from our ftp server:

ftp://ftp.zew.de/pub/zew-docs/dp/dp06085.pdf

Non-technical summary

The Internet has diffused rapidly in households throughout Germany in recent years. However, not all population groups participate similarly in the diffusion process. Individual characteristics, such as education, age, and income can create large Internet access barriers. In addition, regional characteristics, such as the existing ICT infrastructure and price structure determine the individual access probability. Differences in accessing the Internet due to these determinants are facets of the digital divide.

This paper focuses on regional aspects of the digital divide in Germany. For the period 2000/2001 it shows how regional characteristics are related to differences in rates of Internet use between counties as well as to differences in the Internet access probability between individuals. The empirical work is based on two large data sets: the SOEP data and the INKAR data. This provides the opportunity of merging detailed individual and regional information.

The results of the multivariate analyses at regional level do not support the hypothesis that a higher local proportion of people living in rural communities is accompanied with a lower county-wide Internet use rate. Other regional characteristics, such as the proportion of foreigners, the proportion of highly qualified employees, and the regional rate of unemployment, turn out to be more important. Thus, it is not rurality per se that explains differences in Internet use. The results rather indicate that it is the different composition of individual characteristics between rural and urban population that accounts for the regional digital divide.

At individual level, the estimation results show that the decision to become a new Internet user is strongly influenced by individual characteristics. In line with previous research results, individuals who are more highly educated, younger, and wealthier are more likely to access the Internet. Moreover, a strong and positive impact of the local proportion of experienced Internet users who live around hitherto non-users is observable, which underlines the importance of network effects for the individual adoption decision. This could be a result of learning

spillovers. The population density turns out to play a minor role. However, if East and West Germany are studied separately, the results show that living in a rural area in East Germany strongly reduces the probability of accessing the Internet compared to individuals living in East German city regions. This effect cannot be found for West Germany. Differences in the Internet infrastructure between rural and urban areas in East Germany are likely to be a reason for this result.

It can be concluded that policies aimed at decreasing the digital divide should provide programs which encourage the Internet literacy of less qualified, unemployed, and older individuals. Furthermore, due to the existence of network effects, experienced users should be involved in public programs in order to motivate non-users by teaching them how to use the Internet and by showing them its advantages.

Zusammenfassung

Das Internet hat sich in den vergangenen Jahren in Deutschland stark verbreitet. Jedoch partizipieren nicht alle Bevölkerungsgruppen in gleichem Maße an dieser Entwicklung. Individuelle Charakteristika, wie z.B. Bildungsstand, Alter und Einkommen, können Hemmnisse für die Nutzung des Internets darstellen. Zudem beeinflussen regionale Faktoren, z.B. die bestehende IKT-Infrastruktur und -Preisstruktur, die individuelle Zugangswahrscheinlichkeit. Die aus diesen Determinanten resultierenden Unterschiede in der Internetnutzung sind Facetten der so genannten digitalen Kluft (digital divide).

Im Mittelpunkt dieses Papiers stehen regionale Aspekte der digitalen Kluft. Für den Zeitraum 2000/2001 wird untersucht, inwiefern regionale Faktoren Unterschiede in den Internetnutzungsraten zwischen Kreisen einerseits sowie Unterschiede in der Wahrscheinlichkeit des Internetzugangs zwischen Individuen andererseits erklären können. Hierfür werden zwei umfangreiche Datensätze miteinander verbunden: das Sozio-ökonomische Panel (SOEP) und der INKAR-Datensatz. Dadurch lassen sich regionale Informationen den Individuen zuspielen.

Auf der Ebene von Kreisen können die Ergebnisse der multivariaten Analysen die Hypothese, dass eine geringere Einwohnerdichte mit einer geringeren Internetnutzungsrate einhergeht, nicht bestätigen. Vielmehr sind andere regionale Charakteristika, wie der Anteil an Ausländern und Hochqualifizierten sowie die regionale Arbeitslosenquote, von Bedeutung. Somit lässt sich schließen, dass nicht die Ländlichkeit einer Region Unterschiede in der Höhe der Internetnutzung erklären kann, sondern eher die unterschiedliche Verteilung individueller Charakteristika in der Bevölkerung von Stadt und Land.

Auf der Individualebene zeigt sich, dass die individuelle Zugangswahrscheinlichkeit insbesondere durch persönliche Charakteristika beeinflusst wird. Die Ergebnisse der ökonometrischen Schätzungen bestätigen die Resultate früherer Studien: Jüngere, besser ausgebildete und wohlhabendere Individuen haben eine höhere Wahrscheinlichkeit, das Internet zu nutzen. Zudem lassen sich deutlich positive Netzwerkeffekte beobachten. Das bedeutet, je mehr erfahrene Nutzer in der

Region eines bisherigen Nichtnutzers leben, umso höher ist seine Nutzungswahrscheinlichkeit. Dies gilt vor allem für den Westteil Deutschlands. Im Osten bleibt jedoch auch unter Berücksichtigung der Netzwerkeffekte der Einfluss der Bevölkerungsdichte signifikant. Dies wird vermutlich durch Unterschiede in der Internetinfrastruktur zwischen städtischen und ländlichen Regionen Ostdeutschlands hervorgerufen.

Politikmaßnahmen, die darauf abzielen, die digitale Kluft zu verringern, sollten daher vor allem Programme beinhalten, die die IT-Fähigkeiten von gering qualifizierten, arbeitslosen und älteren Personen verbessern. Da die Ergebnisse dieses Papiers zudem deutlich zeigen, wie wichtig Netzwerkeffekte für die Internet-Nutzungswahrscheinlichkeit sind, sollten erfahrene Internetnutzer in entsprechende Programme involviert werden, um ihr Wissen an bisherige Nichtnutzer weitergeben zu können.

Regional Versus Individual Aspects of the Digital Divide in Germany*†

Katrin Schleife‡

first version: December 2006
substantially revised version: June 2008

Abstract

This paper analyzes the regional dimension of the German digital divide. It studies the determinants of home Internet use in Germany on the level of counties as well as on the level of individuals. Based on two large data sets, the analyses show that population density itself cannot explain regional differences in Internet use rates. The results rather indicate that it is the different composition of individual characteristics between rural and urban populations that accounts for the regional digital divide. At individual level, the findings underline the importance of network effects.

Keywords: digital divide, Internet use, network effects
JEL-Classification: O33, O18, R20

---

*__Data availability:__ The analyses of this paper are based on the Socio-Economic Panel (SOEP) data provided by the German Institute for Economic Research (DIW) as well as on the INKAR data provided by the Federal Office for Building and Regional Planning (BBR).

†__Acknowledgments:__ Helpful comments by Irene Bertschek, Daniel Cerquera, Bettina Müller, and Marianne Saam are gratefully acknowledged.

‡__Address:__ ZEW Mannheim, Centre for European Economic Research, Research Group Information and Communication Technologies, P.O.Box 103443, D-68034 Mannheim, Germany. Email: schleife@zew.de

1 Introduction

The Internet is an information and communication technology which has diffused rapidly in households throughout Germany in recent years. It has led to considerable changes in the living and working conditions of an increasing part of the population. However, not all population groups participate similarly in the diffusion process. In the year 2001, which will be one of the years of observation within this paper's empirical analyses, 37 percent of the population in Germany (aged 14 years and above) was online (at home or at work) (TNS Infratest, 2007). Although the proportion of Internet users increased to 60 percent by 2007, the diffusion of this technology has by no means reached all parts of the population (ibid.). Individual characteristics, such as education, age, and income can create large Internet access barriers. In addition, regional characteristics, such as the existing ICT infrastructure and price structure can determine the individual access probability. These differences in accessing and using the Internet are facets of the digital divide.[1]

Rural regions tend to be economically lagging behind urban areas as industrial and labor markets are concentrated in densely populated regions (Malecki, 2003). People living in rural areas have to overcome long distances to most markets and face limited access to consumer goods, labor, information and other resources. Against the background of this *rural penalty* the use of information and communication technologies, especially the Internet, provides various possibilities to reduce the associated disadvantages (Hudson and Parker, 1990). The Internet can encourage rural development by reducing or even eliminating the difficulties of distance. Besides providing various opportunities for firms that are located in rural areas, the Internet offers, for instance, convenient shopping opportunities and a broader product mix to consumers. It provides the possibility of distance learning and it can facilitate job search activities as described by McQuaid, Lindsay, and Greig (2004). Communication with family members and friends is an additional important motive of using the Internet. However, in spite of these

[1]The OECD defines the digital divide as "the gap between individuals, households, business and geographic areas at different socio-economic levels with regard both to their opportunities to access information and communication technologies (ICTs) and to their use of the Internet for a wide variety of activities." (OECD/DSTI, 2001, p. 5).

possibilities and advantages the diffusion of the Internet is much slower in rural regions than in city areas, which potentially hampers economic development in rural areas and increases the gap in economic well-being between rural and urban regions.

In order to reduce the existing divides it is crucial to identify the driving forces behind that development. The analyses of this paper contribute to the empirical research on the various dimensions of the digital divide. They study the determinants of home Internet use in Germany on the level of counties as well as on the level of individuals by merging two large data sets. At individual level the study focuses on *network effects*, that is the impact of the local proportion of experienced Internet users on the access probability of individuals, as many empirical studies underline the importance of such effects.

The results at regional level show that regions with higher rates of highly educated employees and students exhibit a higher proportion of Internet users, which provides evidence for the important role of education with regard to the use of new technologies. A higher unemployment rate and a higher proportion of foreigners is accompanied by a lower proportion of local Internet use. At individual level, the decision to become a new Internet user is strongly influenced by individual characteristics. In line with previous research results, individuals who are more highly educated, younger, and wealthier are more likely to access the Internet. Moreover, a positive impact of the local proportion of experienced Internet users who live around hitherto non-users is to be found, which underlines the importance of network effects for the individual adoption decision. This could be a result of learning spillovers, for example. The population density turns out to play a minor role. However, if East and West Germany are studied separately, the results show that living in a rural area in East Germany strongly reduces the probability of accessing the Internet compared to individuals living in East German city regions. This effect cannot be found for West Germany. Large differences in the Internet infrastructure between rural and urban areas in East Germany are likely to be a reason for this result.

This paper proceeds as follows: the next section provides a short description of the Internet diffusion process and gives an overview of some empirical literature regarding the individual and regional differences in ICT adoption as well as the role of network effects. In addition, studies that deal with the consequences of

the digital divide are presented. Section 3 describes the two data sets used for the analyses and specifies how the sample is created. The derived hypotheses, the estimation strategy as well as the results of the empirical analyses at regional level are provided in section 4, while the focus of section 5 is on the individual level. Section 6 gives a short summary of the findings.

2 Background Discussion

2.1 The Process of Internet Adoption

The diffusion process of the Internet technology can be analyzed by applying general diffusion models of innovations. The adoption rate of an innovation can be defined as "the relative speed with which an innovation is adopted by members of a social system" (Rogers, 2003, p. 221). With regard to the Internet it can be measured as the number of individuals who become new Internet users within a specific period. Rogers (2003) describes the typical distribution pattern for the general process of innovation adoption. When plotted over time the adoption rate can be expected to follow a normal, bell-shaped curve, which is accompanied by an S-shaped curve of the cumulative number of adopters. Rogers (2003) elaborates on five important perceived attributes of innovations that determine the speed of technology adoption: relative advantage, compatibility, complexity, trialability, and observability.

The *relative advantage* measures the extent to which the new technology is of greater value to adopters than the previous technology. The aspect of *compatibility* primarily considers the experiences of a potential adopter with a previous technology, which helps to reduce uncertainty as it facilitates the evaluation of the new technology. A high *complexity* can work as a barrier to technology adoption. *Trialability* measures the degree to which new users have the ability of experimenting with (parts of) the new technology. *Observability* describes the degree to which the use of technology and its results are observable to other individuals.

Greenstein and Prince (2006) state that the Internet diffusion process follows similar patterns to those which can be observed for the adoption rate of other innovations. The speed of the process of Internet adoption is thereby positively

related to the relative advantage, compatibility, trialability, and observability of the Internet technology. It is negatively related to its complexity.

Individuals differ in the importance of these attributes and thus in the speed of their Internet adoption. The following sections describe several socioeconomic and region-specific factors that may have an impact on that relation.

2.2 Heterogeneity of Potential Internet Adopters

The demand for Internet connectivity depends on the individual utility of using this new technology. Thus, individual preferences as well as perceptions regarding the possibilities of the Internet play an important role in the decision to become an Internet user. In addition, technical and financial resources determine the individual start-up costs and price sensitivity. All these factors interact with several demographic and social characteristics, such as age, education, skill level, occupation, and income.

Research in various countries confirms that individual characteristics can generate differences in the timing of adoption. An international comparative analysis of the OECD in 2001 based on national statistics of OECD countries shows that Internet access and use is more frequent for individuals and households with a higher income and a higher level of education (OECD/DSTI, 2001). Analyzing the online population in the U.S., Lenhart (2003) reaches similar conclusions with survey data from 2002.

Research results obtained for Germany, for example by Korupp and Szydlik (2005) and Korupp, Künemund, and Schupp (2006), resemble those mentioned above. Jäckel, Lenz, and Zillien (2005) confirm the crucial role of education analyzing German data from 2003. An important reason for this finding may be that as computers and the Internet were introduced first in academic and research institutions, highly educated people grew accustomed to these new technologies earlier than others. The important role of universities in the process of Internet diffusion is analyzed by Goldfarb (2006). He provides evidence that in the mid-1990s universities in the U.S. intensely taught students how to use the Internet and that "these students then brought the technology into their homes" (Goldfarb, 2006, p. 203).

As connecting to the Internet depends heavily on using a computer, factors influencing computer adoption also affect Internet adoption to a large extent. Several studies analyze the determinants of computer use. Most of them confirm that a higher education and a higher income are positively related to computer adoption.[2]

2.3 Regional Differences in the Use of New Technologies

The geographical diffusion of the Internet is primarily determined by both the decision of individuals to adopt the Internet and the decision of firms to supply Internet connectivity in a specific region. On the demand side, the individual adoption decision largely depends on the individual willingness to pay for the new technology. On the supply side, firms will only enter a market if it is profitable.

The innovation attributes, which influence the demand for the new technology to a great extent, can be assumed not to be equally distributed geographically. In particular, differences between rural and urban areas are observable. For example, the Internet can provide a higher *relative advantage* for people living in rural areas than for those in cities, as better and faster communication channels or easier consumption possibilities are of higher value for those living in remote areas (Greenstein and Prince, 2006). Thus, the demand for Internet access is large in remote areas.

Providing related considerations, Sinai and Waldfogel (2004) state that the Internet may act as both a substitute and a complement for cities. The substitute function originates from individuals using the Internet to overcome local isolation regarding communication and product availability. Local product variety is expected to be higher in larger markets. If the substitute function prevails, Internet use should be higher in rural areas. On the other hand, the Internet complements cities as it offers local websites containing local news and information. The amount of these sites grows with size and population density of the region. Besides, local sellers may provide special services or may offer additional products via the Internet. This is also part of the complement function as the density

[2]See, for example, Prince (2008), Borghans and ter Weel (2002), and Haisken-DeNew, Pischner, and Wagner (2000).

of sellers is much higher in city areas. Using data for the U.S., Sinai and Waldfogel (2004) conclude that both the substitute and the complement function are observable. They additionally explain that the strength of these functions varies between population groups.

However, Sinai and Waldfogel (2004) provide evidence that the tendency to connect to the Internet is not affected by the size of the market. It seems that the complement and the substitute function offset each other. However, the authors do not account for regional variations in the supply of Internet connectivity.

Mills and Whitacre (2003) test the relative importance of household attributes versus region-based characteristics when explaining the large differences in Internet use between metropolitan and non-metropolitan areas in the U.S. Their results suggest that nearly two thirds of the divide can be explained by differences in household characteristics like education (of household head) and income. One third stems from place-based characteristics, especially from network externalities.[3] Hindman (2000) also provides evidence that individual characteristics determine Internet use rather than the place of residence.

However, although it could provide a substantial return on investment to the economy as a whole, as stated by Parker (2000), there is little supply of Internet infrastructure in rural regions because the return on investment for each potential supplying firm is often too small to justify the investment. In addition to the barriers created by large distances and the low density of markets, *complexity* and *observability* may be big problems for people in rural regions, as these people do not have the variety of learning and observation possibilities which exist in cities. Moreover, higher unemployment rates in rural areas lead to lower income levels and less financial resources compared to cities.[4] This further reduces the incentives of firms to invest in Internet infrastructure and services in rural areas and thus further decreases the possibilities of adopting the Internet for the people living in these regions.

[3]See the next section for a discussion of network externalities.

[4]In East Germany, rural regions exhibit the highest unemployment rates. In contrast, in West Germany rural areas have lower unemployment rates compared with city regions (OECD, 2007).

In her analyses of Internet connectivity in rural regions of the U.S., Strover (2001) confirms that rural citizens are much less likely to connect. Competing telecommunication service providers are disproportionally clustered in urban areas. As a result, the Internet is provided at higher costs in rural regions. In addition, Strover (2001) argues that Internet connectivity is offered with fewer services and lower quality in these areas. Additional differences in infrastructure can emerge by a lower density of retailers supplying telecommunication goods and services in rural regions.

2.4 The Importance of Network Effects

The importance of *network effects* in the technology diffusion process attracts more and more attention in the literature. Positive network effects arise if an individual's benefits of participating in a network increase with the size of the network (Goolsbee and Klenow, 2002). Regarding the Internet these effects are obvious as a larger network increases the individual's communication possibilities (especially if family members or friends join the network) as well as the content that is available online. A survey of individuals carried out in 2005 shows that in Germany the majority of Internet users (75 percent) used the help of relatives and friends when learning how to use the Internet (Statistisches Bundesamt, 2006).

If one-time costs of joining a network or switching to another one exist, network effects are likely to cause inefficient outcomes (Goolsbee and Klenow, 2002). As a result, technology adoption can be too fast or too slow. For example, technology adoption may be too fast in the case of an inferior technology if too many people join this network (because it was available earlier, for example) instead of a superior one and they cannot move to the latter without any costs (ibid.). Thereby, the adoption of the superior technology may be too slow. As stated by Goolsbee and Klenow (2002), users and suppliers should take these network externalities and the resulting dynamics into account when making their decision to join and provide a network, respectively.

The term *network effects* also comprises social influence exerted by the user network that surrounds current non-users. Agarwal, Animesh, and Prasad (2005) state that the existence of such social networks can further increase the Internet adoption probability, for example due to learning from others or just due

to pressure to conform. Learning from others is important as experienced users can teach a hitherto non-user how to use the Internet and what its benefits are. Rogers (2003) concludes that the diffusion of an interactive communication technology, such as the Internet, is characterized by a reciprocal interdependence: Early adopters influence late adopters by communicating their experience and knowledge. At the same time, late adopters have an impact on earlier adopters by directly increasing the network.

Agarwal et al. (2005) find evidence of the existence of such network effects. Thus, the authors conclude that an individual's decision to access the Internet is indeed influenced by the local number of users. The results of Mills and Whitacre (2003) also show that positive externalities exist because a higher regional density of Internet use is positively correlated with a household's probability of using the Internet.

The importance of network effects in the diffusion of ICT is studied in more detail by Goolsbee and Klenow (2002) focusing on the diffusion of home computers in the U.S. They find that households living in regions with a higher proportion of people that already own a computer are more likely to buy a first computer even if various individual characteristics are considered. By conducting a multitude of econometric tests, Goolsbee and Klenow (2002) show that the high network effect is robust as it cannot be explained by common unobserved traits or by the local economic environments of those living in the same region.

As explained by Goolsbee and Klenow (2002), models with network externalities and learning spillovers predict a steady increase in the adoption rate of new technologies when the level of cumulative adoption increases. The authors find evidence that the adoption rate is increasing with the size of the network across all analyzed ranges. This is in contrast to the basic theory of technology adoption provided above which predicts an S-shaped pattern of the diffusion curve.

2.5 The Digital Divide

Many economic research studies find evidence that the economies of counties, states, and countries benefit from investments in local telecommunication infrastructure (Parker, 2000). In particular for rural areas, the Internet provides many

advantages as it can neutralize two major barriers to rural economic growth: the large distances and the lack of economies of scale due to smaller market size (ibid.). However, there are large differences in the geographic distribution of Internet use. Due to a lack in supply of fast, efficient, and inexpensive Internet infrastructure inhabitants of remote areas do not benefit from the possibilities the Internet offers. If Internet connectivity cannot be ensured in rural regions in the near future, rural areas will lose part of their attractiveness to (highly qualified and wealthy) individuals and businesses.

Besides this regional digital divide due to a gap in infrastructure, many studies show that there are large differences in the Internet use between members of different status groups. Earlier adopters of new technologies tend to be younger, more highly educated, and wealthier than those who adopt later. According to Hindman (2000) the differences in Internet use between status groups are very likely to grow as most of the Internet content is designed for higher status groups. As a result, differences do not only exist with regard to the probability of accessing the Internet but also concerning the way it is used. Due to this digital divide the use of information technologies is expected to reinforce or increase existing social and economic inequalities between population groups. In order to find appropriate policy programs which aim at reducing these inequalities, the facettes of the digital divide need to be analyzed.

3 Data

The empirical analyses of this paper are performed on the basis of two data sets: the SOEP, which provides detailed information on individuals, and INKAR[5], which comprises a wide range of official regional figures for Germany. With the combined data set various individual socio-economic as well as region-based characteristics can be considered when analyzing the inequalities in home Internet access between regions and individuals.

[5]INKAR – Indikatoren und Karten zur Raumentwicklung (indicators and maps on land development).

The SOEP is an annually representative longitudinal survey of private households.[6] For the analyses of this paper the 2001 SOEP wave is considered. Besides other socio-economic variables it provides information on the individuals' computer and Internet use.[7] The 2001 SOEP wave covers more than 22,000 individuals aged 16 years or older. However, the SOEP data contains hardly any regional information which makes a second data set necessary to fill this gap.

The INKAR data set is provided by the German Federal Office for Building and Regional Planning[8] and contains a wide range of regional figures, for instance regarding the structure of population, employment, and industry, or levels of education, production, and wages.[9] Thus, INKAR does not only allow regions to be classified as rural or non-rural, but provides a much more detailed description of regions. The INKAR data is given for several regional levels. The county level which is used in the analyses of this paper is the lowest aggregation level.[10] As most variables are given with a time lag, INKAR data sets of 2002 to 2005 are used. This provides information for the year 2001 used at regional level and for the year 2000 additionally used at individual level.[11] By merging the two data sets, regional information can be assigned to individuals. At the individual level of the analyses, regional information can therefore be treated as a person-specific determinant.

Germany consists of 440 counties. Seven of these counties cannot be considered in the analyses since they do not match the SOEP data. Moreover, only counties that contain 20 or more observed SOEP individuals are considered. In addition, the sample only contains individuals aged between 16 and 64 and individuals who

[6] See Haisken-DeNew and Frick (2005) for a detailed description of the SOEP.

[7] The related questions were: 'Do you use a computer and the Internet for activities not related to work? If yes, since when?' and 'Do you use a computer and the Internet at work or in your education? If yes, since when?'. Answers can be given separately for computer use and Internet use. Questions regarding ICT use are not included in every SOEP wave.

[8] Bundesamt für Bauwesen und Raumordnung – BBR.

[9] INKAR does not include information on ICT use and access.

[10] It is the level of the German "Kreise".

[11] As INKAR does not contain the required age structure figures for the year 2001, they are taken from an additional data set: 'Statistik regional 2003' provided by the Federal Statistical Office Germany.

provided information regarding their Internet use. Thus the data set comprises 312 counties and 16,662 individuals.

4 The Regional Level

4.1 Hypotheses and Estimation Strategy

Initially, at regional level, differences in the proportions of private Internet use between German counties are analyzed. The main research questions are: what are the regional characteristics that determine the degree of home Internet use in German counties? Can population density explain differences in Internet use between regions?

The information on individual Internet use given by the SOEP data has been aggregated by county. This leads to the respective regional proportion of Internet users in 2001, which represents the outcome variable of the econometric model. One of the main regional explanatory variables is a rural-urban indicator: the proportion of the population that lives in communities with less than 150 inhabitants per square kilometer, the so called *rurality*. It serves to assess the impact of population density on the proportion of Internet users. In recent papers it is often argued that technological differences, such as the availability and the quality of Internet connectivity, are one of the main reasons for differences in Internet use rates between rural areas and cities.[12] This is also likely to hold for Germany, as remote areas faced higher prices and a lower network capacity even before the diffusion of broadband connectivity.[13]

Sinai and Waldfogel (2004) state that due to the attainable benefits of using the Internet it may be both a substitute and a complement for cities.[14] Depending on

[12]See, for example, Strover (2001) and Greenstein and Prince (2006).

[13]At the beginning of the year 2001, broadband Internet connectivity was scarcely spread in Germany. Only a very small proportion of 4 percent of private user households in Germany used a broadband Internet connection (TNS Infratest, 2002). Thus, when explaining the underlying causes of the digital divide, differences in the availability and capacity of broadband Internet connections did not seem to be among the most important factors a few years ago.

[14]See the discussion provided above.

which function outweighs the other, the relationship between population density and Internet use rates is positive or negative. Part of the rural-urban differences in Internet use caused by the Internet's complement or substitute properties could be explained by various region-specific characteristics, such as the regional size of minority groups like foreigners. Thus, the hypothesis is that the population density itself is not the crucial factor, but the regional size of specific population groups that have specific preferences regarding communication and consumption. Therefore, the significance of rurality is expected to decrease if additional regional factors are considered in the estimation approach. Further variables may be correlated with the proportions of Internet users and are therefore added in a further specification: the proportion of the population aged between 15 and 29, the size of the foreign population, the proportion of one-person households within all households, the proportion of employees in the county who are highly qualified, the proportion of students, the unemployment rate, and the mean disposable household income.[15]

The local proportion of young people is expected to be positively correlated with regional Internet use rates, as adolescents nowadays become familiar with ICT very early by using it in school, during their apprenticeship, or for communication as well as leisure activities among friends. They involve the use of a computer and the Internet within their families, thereby increasing the regional proportion of home Internet users.

Many studies find complementarities between skills and the use of new technologies, showing that educational level increases the probability of using a computer or the Internet at work.[16] By accumulating IT skills at work, highly qualified workers also become more likely to use computers and the Internet at home. Assuming that people work and live in the same county, a larger regional proportion of highly qualified employees is therefore expected to be accompanied by a higher rate of home Internet use.

---

[15]Unfortunately, no data is available regarding the activities the Internet is used for or regarding the amount of region-specific websites. Sinai and Waldfogel (2004) use such data regarding local online content in order to evaluate the relationship between population density and Internet use more precisely.

[16]This result is obtained, for example, by Borghans and ter Weel (2002) using data of Germany, Great Britain and the United States, and by Entorf, Gollac, and Kramarz (1999) analyzing French data.

As the availability of time and money – two important requirements for using the Internet – is often high in one-person households, a positive relationship of this factor with the county-wide proportion of Internet users should be observable.[17] Moreover, a higher mean disposable income should be positively correlated with the Internet use rate.

Unemployed people could highly benefit from using the Internet for their job search activities. But many studies ascertain that those people are less likely to own a computer at home and to have access to the Internet, presumably because of higher financial restrictions (McQuaid et al., 2004). Thus, the proportion of home Internet users is expected to be negatively correlated with the local proportion of unemployed persons.

Following the arguments of Sinai and Waldfogel (2004), foreigners tend to use the Internet to overcome local isolation. Thus, the larger the proportion of foreigners, the less the Internet is needed for their online communication or shopping activities. As a result, this Internet substitution function may intensify the anticipated negative impact of the proportion of the foreign population on the local proportion of Internet users, which can probably be explained by linguistic problems or shortcomings in education.

Part of the regional variation between German counties may be due to differences between East and West Germany with regard to population structure and economic conditions. This possibility will be considered in the estimation approach by including a dummy variable which takes the value one if a county belongs to West Germany and the value zero for East German counties.

Summarizing all these considerations, the hypotheses to be analyzed are: i) a greater rurality leads to a smaller regional proportion of home Internet users. ii) The impact of rurality declines when additional regional characteristics are considered. iii) The proportions of highly qualified employees, of young people, and of one-person households are positively correlated with the regional proportion of Internet users. iv) A higher unemployment rate and a larger proportion of foreigners in a county lead to a smaller proportion of Internet users. v) Differences between East and West German counties can explain part of the correlations.

[17] One-person households consist of those households where one person lives alone. However, this does not imply that the person is unmarried or has no partner (single).

The dependent variable of the econometric model, the county-wide proportion of home Internet users, is measured as a percentage. As the boundary values of zero and one hundred percent can possibly be observed, the hypotheses are tested by using the *fractional response model*. This quasi-maximum likelihood estimation (QMLE) method was developed by Papke and Wooldridge (1996) for analyzing such fractional response variables.[18] Papke and Wooldridge (1996) use a non-linear function $G(.)$ for estimating the expected values of the dependent variable y_r conditional on a vector of covariates x_r. The model can be written as

$$E(y_r|x_r) = G(x_r\beta), \tag{1}$$

where y_r is the proportion of home Internet users in region r in 2001. The chosen $G(.)$ is a cumulative distribution function satisfying $0 < G(.) < 1$. This ensures that the predicted values of y_r lie in the interval between zero and one.

Following Papke and Wooldridge (1996) the multinomial logistic function

$$E(y_r|x_r) = \frac{e^{x_r\beta}}{1 + e^{x_r\beta}} \tag{2}$$

is applied for $G(.)$. As suggested by the authors a Bernoulli distribution is assumed for y_r and the following binary choice log-likelihood function is maximized using QMLE:

$$l_r(\beta) = y_r log[G(x_r\beta)] + (1 - y_r)log[1 - G(x_r\beta)]. \tag{3}$$

Papke and Wooldridge (1996) show that the obtained quasi-maximum likelihood estimator, $\hat{\beta}$, is consistent and $\sqrt{N}$-asymptotically normal distributed regardless of the distribution of y_r conditional on x_r. The parameter vector β measures the impact of the considered region-specific covariates x_r on the proportion of Internet users.

[18] See Papke and Wooldridge (1996) and Wagner (2001) for a description of the shortcomings of several alternative estimation approaches when analyzing such variables, especially for data where there is the possibility of observing the boundary values.

4.2 Empirical Results

Tables 6 and 7 in the appendix provide an overview of the characteristics of the German counties in the sample. On average, the county-wide proportion of home Internet users is 33 percent in West Germany and 27 percent in East Germany in 2001. Large differences between the two parts of the country can be observed for the unemployment rate, the disposable household income, and the proportion of foreigners.

A comparison of rural, suburban, and urban counties shows that Internet use is significantly more prevalent in urban counties (34 percent) than in rural regions (26 percent).[19] The Internet use rate of suburban counties lies in between (31 percent). Moreover, compared with rural regions, urban counties have significantly larger proportions of highly qualified employees, one-person households, and foreigners, as well as a higher mean disposable household income per capita.

The results of the *fractional response model* estimations of the proportion of home Internet use in German counties are shown in Table 1.[20] The first specification includes rurality as the only explanatory variable. It shows the expected negative bivariate correlation with Internet use proportions, indicating that counties with a larger proportion of persons living in rural areas show a lower Internet penetration compared to more densely populated regions.

As expected, the impact of rurality declines and even becomes insignificant when further regional characteristics are included in the estimation approach (specification (2)). The results show significantly positive correlations between the proportion of highly qualified employees as well as the proportion of students and the regional penetration rates of home Internet use. These findings support the hypothesis that human capital is an important factor for the technological diffusion process. The regional unemployment rate as well as the size of the foreign population show a significantly negative correlation with Internet penetration rates, which also supports the expectations.

[19] See section 5.1 for the definition of rural, suburban, and urban.

[20] The OLS approach was applied for comparison. It provides very similar results.

Table 1: Diffusion of home Internet use at county level in 2001 – FRM results

dependent variable: proportion of population with home Internet use			
regional characteristics	**(1)**	**(2)**	**(3)**
rurality[i]	-.658 (.134)***	-.264 (.201)	-.252 (.203)
prop. of population aged betw. 15 and 29		-7.355 (2.694)***	-6.709 (3.022)**
prop. of highly qualified employees		3.125 (1.353)**	3.622 (1.882)*
prop. of one-person households		.719 (.604)	.646 (.623)
prop. of students		2.832 (1.316)**	2.624 (1.419)*
unemployment rate		-3.295 (1.018)***	-2.807 (1.389)**
prop. of foreign population		-2.118 (.946)**	-2.379 (1.104)**
household income per capita (log)		-.217 (.371)	-.212 (.371)
west			.109 (.237)
constant	-.620 (.042)***	2.086 (2.894)	1.816 (2.940)
log pseudolikelihood	-132.794	-131.580	-131.572
number of observations	312	312	312

Notes: Fractional response model (FRM). ***, **, * indicate significance at the 1%, 5% and 10% level. Standard errors in parentheses. [i]) Proportion of population in communities with a population density of less than 150 inhabitants per square kilometer.
Source: Author's calculations based on SOEP 2001, INKAR 2002 to 2005, Statistik regional 2003.

The proportion of young people has a highly significantly negative effect. Thus, in contrast to the expectations the data on German counties does not provide evidence that young people encourage their families to use new technologies, although adolescents are very likely to use ICT. However, the analyzed year 2001 may be too early to observe the results of this transfer process already.

Taking into account whether a county belongs to East or West Germany reduces some of the significant effects (specification (3)). However, those variables which were significant in specification (2) remain of significant size in specification (3), at least at the 10 percent level. The effect of the 'west' dummy variable itself turns out to be not significantly different from zero. Thus, differences in the population structure between East and West Germany cannot explain regional differences in the diffusion of home Internet use.

5 The Individual Level

5.1 Hypotheses and Estimation Strategy

At an individual level, differences in the probability of becoming a new Internet user are analyzed. Besides several individual demographic and employment-related characteristics, the rurality of the individual's home county and the local proportion of experienced Internet users are taken into account. The main questions are: which individual factors determine the probability of starting to use the Internet? Is the rurality of the home county an influencing factor? And what is the role of network effects when analyzing individual differences in Internet access?

Contrary to the regional level, where differences in the proportions of Internet users between counties are explored, the focus is now on the individual decision to *start* using the Internet. By exploiting the information on the duration of Internet use provided by the SOEP data from 2001, 'beginners' are defined as those individuals who declared that they had used the Internet at home since 2000 or 2001.[21]

The data set comprises 2,346 individuals who are new users, compared to 11,280 individuals who have not started to use the Internet so far. The remaining 3,036 individuals are those with a usage experience of more than one year. Thus, 17 percent of those individuals who had not used the Internet by 2000, connected to the Internet for the first time within the following two years. The proportion of new users differed only slightly between East and West Germany (16.9 compared to 17.3 percent).

As the rurality variable used for the analysis at county level is not provided for the year 2000, a *county type* variable is used which categorizes the counties into 'urban', 'suburban', and 'rural'. The definition follows the basic county

[21]The question concerning the duration of usage is addressed to users only. In order to increase the number of new users in the data set, not only one but two years are taken into account. Thus, new users of 2000 *and* 2001 are defined as 'beginners'.

type classification of the BBR which distinguishes between 'urban agglomeration', 'urbanized areas', and 'rural areas'.[22]

The place of residence may have an additional impact on the individual Internet access decision induced by spillovers from experienced users, as described above. To account for such network effects, the regional proportion of experienced Internet users is considered in the estimation approach.

The hypotheses analyzed at individual level are: i) young and highly qualified individuals have a higher probability of starting home Internet use. ii) Individuals living in rural areas are less likely to become a home Internet user than individuals in urban areas. iii) There is a positive network effect: in counties with large proportions of experienced Internet users non-using individuals have a higher probability of accessing the Internet for the first time.

The impact of individual and regional factors on the individual's decision to become a new Internet user is examined by including these variables in a PROBIT model of the form:

$$\begin{aligned} Pr(y_i = 1|X_i) &= \Phi(\alpha + \beta \cdot X_i + \gamma \cdot countytype_r + \delta \cdot west_r \\ &\quad + \lambda \cdot userrate_r + \varepsilon_{ri}) \end{aligned} \tag{4}$$

$$\text{with} \quad r = 1 \ldots k \quad \text{and} \quad i = 1 \ldots n,$$

where y_i is the dependent variable indicating whether an individual i connects to the Internet at home for the first time in the years 2000 or 2001 ($y_i = 1$) or not ($y_i = 0$). The coefficient vector β shows the effects of various individual observables X_i.[23] The coefficient γ describes the impact of the *county type* of the region r the individual lives in. Whether living in West Germany is correlated with the decision to access the Internet is measured by the dummy variable *west* and its coefficient δ. The variable *userrate* indicates the regional proportion of

[22]These basic types of regional population structures are generated by taking into account a region's population density as well as the importance and function of the region's core. See Bundesamt für Bauwesen und Raumordnung BBR (2002) for details.

[23]The demographic characteristics are taken from the year 2001. For simplicity it is assumed that they are constant within the observation period 2000/2001.

experienced Internet users, that is individuals who stated that they have been using the Internet since 1999 or earlier. The size of the *userrate* effect on the probability of becoming a new user is measured by λ. The error term ε_{ri} covers unobservable individual and regional characteristics. Φ is the cumulative normal distribution function.

5.2 Empirical Results

Table 8 in the appendix shows the average individual characteristics of new Internet users, non users, and experienced users. It can be seen that compared to non-users, new users are significantly younger, better educated, and richer. In addition, new users are significantly more likely to work in a full-time job and to be male, single, and German. Similarly, compared to new users, those individuals who have already been using the Internet for more than one year are better educated, richer, and more likely to work in a full-time job and to be male and single.

As can be seen in Table 9 in the appendix, in rural regions there is a larger proportion of individuals not using the Internet and a lower proportion of experienced users compared to urban areas. The proportion of individuals accessing the Internet for the first time is also lower in rural areas than in suburban and urban regions. Taking only the proportion of non-users into consideration, it turns out that the proportion of new users is significantly larger in urban areas (Table 10). On average, 15 percent of non-users in rural counties start using the Internet in the years 2000 or 2001. In urban counties the proportion is 18 percent. Thus, although in suburban and urban regions a higher proportion already uses the Internet compared to rural areas, the access rates of first time users are still higher. This indicates an increasing gap in the rate of Internet use between densely populated areas and rural regions in Germany.

The results of the PROBIT model are shown in Table 2.[24] The results of the first specification reveal that compared to city regions, individuals living in rural

[24]All of the standard errors of regional determinants are corrected for the fact that they do not vary between individuals living in the same county. For later comparison with the results of the PROBIT-IV approach the estimated coefficients instead of marginal effects are shown in the table.

areas have a significantly lower probability of becoming new users. In line with the results at regional level, this denotes a predominant complementarity of the Internet to cities, caused for example by the number of websites offering local information. It could also be the result of differences in Internet infrastructure or price structure. The coefficients of the individual characteristics largely show the expected signs. The results indicate that individual characteristics are crucial when explaining the individual decision to become a new Internet user.

In order to examine the regional network effect, specification (2) additionally includes the local proportion of experienced Internet users (*userrate*). The results support the hypothesis that the probability of becoming an Internet user increases with the size of the regional network even after considering the county type as well as various individual characteristics. The correlation is highly significant. On average, an increase in the proportion of experienced users by 10 percentage points increases the probability of accessing the Internet by 1.8 percent (not shown in the table). Moreover, considering the local proportion of experienced Internet users results in an insignificant county type effect.

Specification (3) additionally controls for differences due to living in a West German county instead of an East German one, which turns out to be an insignificant factor. However, controlling for this regional difference slightly increases the absolute value of living in a rural county, which becomes significant at the 10 percent level. Thus, besides the positive impact of the local proportion of experienced Internet users, the individual decision to become a new Internet user is negatively affected by the population density, although the relationship is not strong.

The results of the previous estimations reveal that differences between East and West Germany cannot explain differences in the probability of becoming a new Internet user. However, there may be differences in the determinants if they are analyzed *within* both parts of the country. Therefore, the analyses are repeated separately for East and West Germany. The results are shown in Table 3.

Table 2: Determinants of starting home Internet use in 2000 or 2001 – Probit results

dependent variable: probability of starting home Internet use			
variable (reference group)	**(1)**	**(2)**	**(3)**
county type (ref.: urban)			
rural	-.155 (.053)***	-.093 (.057)	-.097 (.057)*
suburban	-.071 (.044)	-.036 (.044)	-.037 (.044)
userrate[i]		.786 (.241)***	.807 (.243)***
west			-.035 (.039)
age in years (ref.: age less than 25)			
25-34	.239 (.059)***	.241 (.059)***	.236 (.060)***
35-44	-.186 (.046)***	-.186 (.046)***	-.187 (.046)***
45-54	-.443 (.053)***	-.446 (.054)***	-.449 (.054)***
55-64	-.792 (.061)***	-.800 (.061)***	-.802 (.061)***
male	.184 (.029)***	.190 (.029)***	.190 (.029)***
single	-.090 (.046)*	-.092 (.046)**	-.093 (.046)**
one-person household	.190 (.056)***	.164 (.056)***	.170 (.056)***
German nationality (ref: foreigner)	.500 (.071)***	.507 (.070)***	.498 (.070)***
education (ref.: university degree)			
lower secondary education or less	-.646 (.068)***	-.640 (.068)***	-.633 (.068)***
other vocational education	-.522 (.122)***	-.522 (.123)***	-.518 (.123)***
apprenticeship	-.560 (.062)***	-.557 (.062)***	-.551 (.061)***
specialized vocational school	-.436 (.067)***	-.429 (.067)***	-.425 (.067)***
technical/commercial college	-.365 (.090)***	-.357 (.090)***	-.353 (.090)***
civil servant college	-.314 (.108)***	-.326 (.108)***	-.317 (.108)***
polytechnic or college abroad[ii]	-.255 (.063)***	-.247 (.063)***	-.248 (.063)***
occup. status (ref.: employed full-time)			
employed part-time	.142 (.044)***	.137 (.044)***	.141 (.044)***
apprentice	-.055 (.077)	-.047 (.077)	-.047 (.077)
not employed	.009 (.040)	.008 (.040)	.009 (.040)
retired	-.228 (.071)***	-.227 (.071)***	-.226 (.071)***
log net income of household	.476 (.048)***	.461 (.048)***	.469 (.046)***
pseudo-R^2	.093	.095	.095
number of observations	12,480	12,480	12,480

Notes: The table shows the estimated coefficients. Standard errors (clustered at regional level) in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level.

[i]) Proportion of Internet users with more than one year usage experience by county.

[ii]) College abroad: In the data it is not clear what kind of degree is meant.

Source: Author's calculations based on SOEP 2001, INKAR 2002 to 2005, Statistik regional 2002.

Table 3: Determinants of starting home Internet use in 2000 or 2001 – East and West Germany, Probit results

dependent variable: probability of starting home Internet use		
variable (reference group)	**East**	**West**
county type (ref.: urban)		
rural	-.342 (.082)***	.037 (.067)
suburban	-.108 (.068)	-.016 (.053)
userrate[i]	.486 (.303)	.862 (.303)***
further covariates[ii]	demographic and job-related characteristics	
pseudo-R^2	.116	.094
number of observations	3,590	8,890

Notes: The table shows the estimated coefficients. Standard errors (clustered at regional level) in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level.
[i]) Proportion of Internet users with more than one year usage experience by county.
[ii]) Further covariates are: male, single, one-person household, German nationality, education, log net income of household, occupational status.
Complete tables are available from the author on request.
Source: Author's calculations based on SOEP 2001, INKAR 2002 to 2005, Statistik regional 2002.

While the effects of the individual characteristics are quite similar for East and West Germany (not shown in the table), the results clearly indicate differences in the impact of population density and network effects within the two parts of the country. In East Germany, individuals in rural regions are significantly less likely to become new Internet users compared with individuals in urban areas, even when individual characteristics are considered. Contrary to the results for Germany as a whole and for West Germany, the network effect is not significant. This result could be explained by large differences in the Internet infrastructure between East German cities and remote areas. As a result, differences in infrastructure can explain a good deal of the regional differences in the proportion of experienced users. In West Germany, however, network effects turn out to be a crucial factor. The county type effects are not significantly different from zero. The differences in the county type effects between East and West Germany could have been expected from the differences in the proportions of new users between the county types shown in Table 10 in the appendix.

5.3 Controlling for Selectivity Bias

When empirically analyzing local spillovers, two important difficulties need to be considered. Firstly, there may be a problem of unobserved individual variables, as for example people who have a greater affinity for technology may self-select into regions where many experienced users live (Goolsbee and Klenow, 2002). Secondly, people living in the same region face similar infrastructure conditions and price structures (Agarwal et al., 2005). These unobserved regional factors may influence both the individual and the group choice of accessing the Internet. The resulting correlation between individual choice and group choice could lead to a bias in the measured network effect.

Following the reasoning of Agarwal et al. (2005), one might be less worried about regional unobservables. Many regional variables, for example unobservable access barriers due to the infrastructure, can be assumed to be correlated with the county type of the region. As this factor is included in the estimation approach it will correct for a large part of the potential bias.

However, a potential simultaneity bias remains. A possible approach to consistently estimating the model with a binary dependent variable and a potential simultaneity bias is the *instrumental variables PROBIT* (IV-PROBIT) approach. The idea is to find one or more observable variables (the instruments) that have an impact on the local proportion of experienced Internet users (the *userrate*) but are otherwise uncorrelated with the individual decision to access the Internet for the first time.

On average, people who have reached a higher level of education are more likely to use the Internet than less educated individuals. Thus, individuals living in areas with a larger proportion of highly educated individuals are more likely to be surrounded by Internet users. The local proportion of highly qualified employees is therefore included as an instrument in the IV estimation approach. Further instruments are the regional unemployment rate and the proportion of foreigners, which are assumed to be negatively correlated with the proportion of experienced users. Moreover, the proportion of one-person households is used as an instrumental variable, which is assumed to be positively correlated with the local proportion of experienced Internet users. The mentioned regional charac-

teristics are not expected to be correlated with the individual decision to start using the Internet.

The instruments are assumed to be plausible due to these considerations, but they also need to be empirically valid. A first assessment of the validity is provided by the instruments' coefficients at the first stage of the instrumental variables approach. As mentioned above, valid instruments should be correlated with the local proportion of experienced Internet users. A Wald test of the null hypothesis that the estimated coefficients associated with the instruments are jointly equal to zero can clearly be rejected for all specifications, as the test statistics far exceed their critical values. In addition, nearly all instrumental variables have themselves a significant coefficient. Thus, the instruments are correlated with the first-stage outcome variable. This holds for the German sample as a whole, but also for the subsamples for East and West Germany. The results of the first-stage regressions are shown in the Tables 11 and 12 in the appendix.[25]

A second criterion that should be fulfilled for credible instruments is that they should not have a direct impact on the dependent variable at the second stage of the estimation approach. One way to test this criterion is the inclusion of the instrumental variables themselves in the baseline (or ordinary Probit) equation. In each specification the Wald test of the null hypothesis that the estimated coefficients are jointly equal to zero cannot be rejected, based on a 99% confidence level. In addition, the coefficient of every single instrumental variable is not significantly different from zero either. Thus, the set of instruments does not show a direct impact on the second-stage outcome variable.

The regional variables used as instruments will not be valid instruments if they are correlated with individual unobservables (Goolsbee and Klenow, 2002). But Goolsbee and Klenow (2002) argue that since individual observables are included in the regressions, any correlation between observables and unobservables should not bias the coefficient on *userrate* (at most the β coefficients of the observables).

[25]The unemployment rate is not used as an instrument in the subsamples' regressions, because here it does not fulfill the criterions of being a valid instrument.

Table 4: Determinants of starting home Internet use in 2000 or 2001 – IV-Probit results, second-stage regressions

dependent variable: probability of starting home Internet use		
variable (reference group)	**(1)**	**(2)**
userrate^{i}	1.730 (.679)**	2.021 (.682)***
county type (ref.: urban)		
rural	-.018 (.078)	-.005 (.080)
suburban	.005 (.052)	.015 (.053)
west		-.063 (.042)
further covariates^{ii}	demographic and job-related characteristics	
constant	-1.766 (.158)***	-1.786 (.158)***
log pseudolikelihood	8672.861	8682.445
Wald test of exogeneity $chi^2(Prob < chi^2)$	2.20 (.138)	3.39 (.066)*
number of observations	12,480	12,480

Notes: Standard errors (clustered at regional level) in parentheses. ***, ** indicate significance at the 1% and 5% level. Instruments for *userrate*: proportion of highly qualified employees, regional unemployment rate, proportion of foreigners, and proportion of one-person households.
^{i}) Proportion of Internet users with more than one year usage experience by county.
^{ii}) Further covariates are: male, single, one-person household, German nationality, education, log net income of household, occupational status.
Source: Author's calculations based on SOEP 2001, INKAR 2002 to 2005, Statistik regional 2002.

Following their illustration of the problem, one might argue that there is a positive correlation because, for example, people with a special interest in technology may predominantly live in cities with a larger proportion of highly educated people. However, the coefficient on individual educational level should absorb any correlation between technology affinity and the proportion of well-educated inhabitants. A correlation between the proportion of the highly educated population and technological affinity would mean that, given that the individual education is controlled for, an increase in the local proportion of well-educated people leads to a higher individual technology affinity. Such a relationship is not very plausible.

Table 5: Determinants of starting home Internet use in 2000 or 2001 – East and West Germany, IV-Probit results, second-stage regressions

dependent variable: probability of starting home Internet use		
variable (reference group)	**East**	**West**
userrate^{i}	1.406 (.728)*	1.814 (.802)**
county type (ref.: urban)		
rural	-.248 (.119)**	.095 (.083)
suburban	-.052 (.093)	.018 (.059)
further covariates^{ii}	demographic and job-related characteristics	
constant	-.973 (.265)***	-1.923 (.185)***
log pseudolikelihood	2916.646	6201.775
Wald test of exogeneity $chi^2(Prob < chi^2)$	1.53 (.217)	1.62 (.203)
number of observations	3,590	8,890

Notes: Standard errors (clustered at regional level) in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level. Instruments for *userrate*: proportion of highly qualified employees, proportion of foreigners, and proportion of one-person households.
^{i}) Proportion of Internet users with more than one year usage experience by county.
^{ii}) Further covariates are: male, single, one-person household, German nationality, education, log net income of household, occupational status.
Source: Author's calculations based on SOEP 2001, INKAR 2002 to 2005, Statistik regional 2002.

The results of the IV-PROBIT estimations are shown in the Tables 4 and 5. The results are very similar to those of the PROBIT model without instruments and I therefore reach the same conclusions. Given the validity of the chosen instruments, endogeneity seems to be a minor problem. This assumption is supported by the results of the Wald test of exogeneity shown for every IV-PROBIT specification. The null hypothesis of this test is that there is no endogeneity in the baseline equation. As the test is insignificant in all specifications (except for one), there is not sufficient information in the sample to reject the null hypothesis. Thus, the regular PROBIT approach should be appropriate.

6 Concluding Remarks

Although Internet use has spread rapidly in recent years in Germany, large discrepancies still exist between population groups regarding their Internet access. This so-called digital divide has many dimensions. Besides differences due to individual characteristics such as age, educational background, and income, there is also a regional gap in ICT use: rural regions show smaller Internet use rates than cities. In order to reduce the existing divides, it is crucial to understand the driving forces behind that development. The aim of this paper is to analyze econometrically the relationship between individual as well as regional characteristics and home Internet access.

At regional level, the results of the bivariate analyses support the hypothesis that a greater rurality is related to a lower Internet use rate. However, the correlation becomes insignificant if further regional characteristics are considered. Several county characteristics are identified that exhibit a high correlation with home Internet use rates. As expected, the proportion of highly qualified employees as well as the proportion of students have a positive impact on regional Internet use penetration. Regions with a higher unemployment rate and those with a larger size of foreign population exhibit a smaller proportion of Internet users.

At individual level, the results underline the importance of individual factors, such as education, age, and income for the decision to start Internet use. Furthermore, a positive network effect is observable for the German and the West German data set. Thus, the probability of becoming a new Internet user is higher for individuals that are surrounded by experienced users. Besides, living in a rural region remains important, particularly in East Germany. This effect is assumed to capture differences in the Internet infrastructure as well as in retail and price structures. These differences seem to be more pronounced in East than in West Germany.

It can be concluded that especially in East Germany differences in the Internet infrastructure between rural and urban regions need to be reduced in order to increase the diffusion of Internet access. Moreover, policies aimed at decreasing the digital divide should provide programs which encourage the Internet literacy of less qualified, unemployed, and older individuals. Furthermore, due to the

existence of network effects, experienced users should be involved in public programs in order to motivate non-users by teaching them how to use the Internet and by showing them its advantages.

Appendix

Table 6: Regional characteristics of East and West German counties, 2001

regional variable	Germany	East	West
proportion of Internet users	0.32	0.27***	0.33
rurality[i]	0.24	0.37***	0.20
proportion of population aged between 15 and 29	0.17	0.19***	0.17
proportion of highly qualified employees	0.07	0.09***	0.07
proportion of one-person households	0.34	0.33	0.34
proportion of students	0.02	0.02	0.02
unemployment rate	0.11	0.18***	0.08
proportion of foreign population	0.07	0.02***	0.09
disposable household income per capita (in Euro)	1,336	1,149***	1,405
number of counties	312	85	227

Notes: Mean values of regional figures for the year 2001. *** indicate that East German means significantly differ from West German values at the 1% level (measured by a t-test on the equality of means). [i]) Proportion of population in communities with a population density of less than 150 inhabitants per square kilometer.
Source: Author's calculations based on SOEP 2001, INKAR 2002 to 2005, Statistik regional 2003.

Table 7: Comparison of rural, suburban, and urban German counties, 2001

regional variable	rural	suburban	urban
proportion of Internet users	0.26***	0.31	0.34
rurality[i]	0.55***	0.25***	0.12
proportion of population aged between 15 and 29	0.18***	0.18***	0.17
proportion of highly qualified employees	0.05***	0.07***	0.09
proportion of one-person households	0.30***	0.33*	0.35
proportion of students	0.01**	0.02	0.02
unemployment rate	0.11*	0.11*	0.10
proportion of foreign population	0.04***	0.06***	0.10
disposable household income per capita (in Euro)	1,211***	1,305***	1,418
number of counties	49	137	126

Notes: Mean values of regional figures for the year 2001. ***, **, * indicate that means are significantly different from urban means at the 1%, 5% and 10% level (measured by a t-test on the equality of means). [i]) Proportion of population in communities with a population density of less than 150 inhabitants per square kilometer.
Source: Author's calculations based on SOEP 2001, INKAR 2002 and 2003, Statistik regional 2002.

Table 8: Characteristics of new Internet users, non-users, and experienced users[i], 2001

individual variable	non-users	new users	experienced users
number of individuals	11,280	2,346	3,036
age in years	42.6	36.8***	36.7
male	0.45	0.51***	0.62***
single	0.20	0.27***	0.34***
one-person-household[ii]	0.10	0.08***	0.11***
German nationality	0.87	0.94***	0.95
education			
number of individuals[iii]	11,028	2,293	2,973
lower secondary education or less	0.26	0.22***	0.20
other vocational education	0.02	0.01	0.01
apprenticeship	0.42	0.35***	0.29***
specialized vocational school	0.11	0.11	0.08***
technical/commercial college	0.05	0.06**	0.07
civil servant college	0.02	0.03***	0.03
polytechnic or college abroad[iv]	0.08	0.11***	0.12**
university	0.05	0.13***	0.20***
occupational status			
number of individuals[ii]	11,159	2,299	2,984
employed full-time	0.46	0.54***	0.61***
employed part-time	0.15	0.17***	0.14***
apprentice	0.04	0.05***	0.04***
not employed	0.25	0.21***	0.20
retired	0.11	0.03***	0.02***
income of household (in 1,000 Euro)	4.44	5.44***	5.87***

Notes: Mean values of individual characteristics for 2001.
***, ** indicate that means are significantly different at the 1% and 5% level: means of new users are compared with those of non-users, means of experienced users are compared with those of new users (measured by a t-test on the equality of means).
[i]) Experienced Internet users are those with more than one year usage experience.
[ii]) In the data, two non-using individuals have missing values for that characteristic.
[iii]) Differences in the number of observations originate from missing values in the data set.
[iv]) College abroad: in the data it is not clear what kind of degree is meant.
Example: In 2001, the proportion of singles among the new users is 27 percent, among the non-users the single proportion is 20 percent.
Source: Author's calculations based on SOEP 2001.

Table 9: Proportion of new users[i], experienced users[ii], and non-users by county type

regional variable	total	rural	suburban	urban
Germany				
new users	0.14	0.13	0.15	0.14
experienced users	0.18	0.13	0.16	0.21
non-users	0.68	0.74	0.69	0.65
number of individuals	16,662	1,839	5,787	9,036
East Germany				
new users	0.14	0.11	0.14	0.16
experienced users	0.16	0.10	0.13	0.20
non-users	0.70	0.80	0.73	0.64
number of individuals	4,533	781	1,691	2,061
West Germany				
new users	0.14	0.15	0.15	0.13
experienced users	0.19	0.15	0.17	0.21
non-users	0.67	0.70	0.68	0.66
number of individuals	12,129	1,058	4,096	6,975

Notes: Mean values for the year 2000.
[i]) Proportion of total population of those who became new users in 2000 or 2001.
[ii]) Proportion of total population of those with more than one year Internet use experience.
Source: Author's calculations based on SOEP 2001, INKAR 2002.

Table 10: Proportion of new users among hitherto non-users by county type

new users	total	rural	suburban	urban
Germany	0.17	0.15**	0.17	0.18
number of individuals	2,346	243	841	1,262
East Germany	0.17	0.12***	0.16***	0.20
number of individuals	645	83	235	327
West Germany	0.17	0.18	0.18	0.17
number of individuals	1,701	160	606	935

Notes: Mean values for the year 2000.
*** and ** indicate that means are significantly different from urban means at the 1% and 5% level (measured by a t-test on the equality of means).
Source: Author's calculations based on SOEP 2001, INKAR 2002.

Table 11: Determinants of starting home Internet use in 2000 or 2001 – IV-Probit results, first-stage regressions

dependent variable: proportion of experienced Internet users		
variable (reference group)	**(1)**	**(2)**
county type (ref.: urban)		
rural	-.047 (.015)***	-.047 (.015)***
suburban	-.030 (.011)***	-.030 (.011)***
west		.018 (.046)
prop. of highly qualified employees	.512 (.166)***	.605 (.305)**
unemployment rate	-.465 (.124)***	-.375 (.207)*
prop. of foreigners	-.249 (.187)	-.284 (.223)
prop. of one-person households	.250 (.096)***	.229 (.121)*
further covariates[i]	demographic and job-related characteristics	
constant	.114 (.031)***	.094 (.048)*
number of observations	12,480	12,480

Notes: Standard errors (clustered at regional level) in parentheses.
***, **, * indicate significance at the 1%, 5% and 10% level.
[i]) Further covariates are: male, single, one-person household, German nationality, education, log net income of household, occupational status.
Source: Author's calculations based on SOEP 2001, INKAR 2002 to 2005, Statistik regional 2002.

Table 12: Determinants of starting home Internet use in 2000 or 2001 – East and West Germany, IV-Probit results, first-stage regressions

dependent variable: proportion of experienced Internet users		
variable (reference group)	**East**	**West**
county type (ref.: urban)		
rural	-.028 (.025)	-.049 (.019)***
suburban	-.014 (.020)	-.028 (.013)**
prop. of highly qualified employees	.539 (.303)*	1.240 (.290)***
prop. of foreigners	.811 (.267)***	-.569 (.231)**
prop. of one-person households	.149 (.179)	.036 (.107)
further covariates[i]	demographic and job-related characteristics	
constant	.021 (.046)	.135 (.029)***
number of observations	3,590	8,890

Notes: Standard errors (clustered at regional level) in parentheses.
***, **, * indicate significance at the 1%, 5% and 10% level.
[i]) Further covariates are: male, single, one-person household, German nationality, education, log net income of household, occupational status.
Source: Author's calculations based on SOEP 2001, INKAR 2002 to 2005, Statistik regional 2002.

References

Agarwal, R., Animesh, A., and Prasad, K. (2005). *Social Interactions and the 'Digital Divide': Explaining Regional Variations in Internet Use* (working paper No. 06-024). RHS School of Business.

Borghans, L., and ter Weel, B. (2002). Do Older Workers Have More Trouble Using a Computer Than Younger Workers? In A. de Grip, J. van Loo, and K. Mayhew (Eds.), *The Economics of Skills Obsolescence: Theoretical Innovations and Empirical Applications, Research in Labor Economics* (Vol. 21, pp. 1–26). JAI Press, Amsterdam, The Netherlands.

Bundesamt für Bauwesen und Raumordnung BBR. (2002). *Aktuelle Daten zur Entwicklung der Städte, Kreise und Gemeinden.* Berichte, Band 14.

Entorf, H., Gollac, M., and Kramarz, F. (1999). New Technologies, Wages, and Worker Selection. *Journal of Labor Economics*, *17*(3), 464–491.

Goldfarb, A. (2006). The (Teaching) Role of Universities in the Diffusion of the Internet. *International Journal of Industrial Organization*, *24*(2), 203–225.

Goolsbee, A., and Klenow, P. (2002). Evidence on Learning and Network Externalities in the Diffusion of Home Computers. *The Journal of Law and Economics*, *45*, 317–344.

Greenstein, S., and Prince, J. (2006). *The Diffusion of the Internet and the Geography of the Digital Divide in the United States* (working paper No. 12182). NBER.

Haisken-DeNew, J., and Frick, J. (2005). *Desktop Companion to the German Socio-Economic Panel Study (DTC).* http://www.diw.de/english/compendium_(dtc)/33921.html. (accessed: March 03, 2008)

Haisken-DeNew, J., Pischner, R., and Wagner, G. (2000). Computer- und Internetnutzung hängen stark von Einkommen und Bildung ab – Geschlechtsspezifische Nutzungsunterschiede in der Freizeit besonders ausgeprägt. *DIW-Wochenbericht*(41/00).

Hindman, D. B. (2000). The Rural-Urban Digital Divide. *Journalism & Mass Communication Quarterly*, *77*(3), 549–560.

Hudson, H., and Parker, E. (1990). Information Gaps in Rural America. *Telecommunications Policy*, *14*(3), 193–205.

Jäckel, M., Lenz, T., and Zillien, N. (2005). *Die regionale digitale Spaltung. Eine empirische Studie zu Unterschieden in der Internetnutzung in Stadt und Land.* Schriftenreihe zum E-Business (Band 4).

Korupp, S., Künemund, H., and Schupp, J. (2006). Digitale Spaltung in Deutschland: Geringere Bildung – seltener am PC. *DIW Wochenbericht*, *73*(19), 289–294.

Korupp, S., and Szydlik, M. (2005). Causes and Trends of the Digital Divide. *European Sociological Review*, *21*(4), 409–422.

Lenhart, A. (2003). *The Ever-Shifting Internet Population – A New Look at Internet Access and the Digital Divide.* Pew Internet & American Life Project Report, http://www.pewinternet.org/reports.asp. (accessed: February 15, 2008)

Malecki, E. (2003). Digital Development in Rural Areas: Potentials and Pitfalls. *Journal of Rural Studies*, *19*(2), 201–214.

McQuaid, R., Lindsay, C., and Greig, M. (2004). Re-Connecting the Unemployed: ICT and Services for Job Seekers in Rural Areas. *Information, Communication and Society*, *7*(3), 364–388.

Mills, B., and Whitacre, B. (2003). Understanding the Non-Metropolitan–Metropolitan Digital Divide. *Growth and Change*, *34*(2), 219–243.

OECD. (2007). *OECD Rural Policy Reviews: Germany* (Vol. 2007). Paris, France.

OECD/DSTI. (2001). *Understanding the Digital Divide.* http://www.oecd.org/dataoecd/38/57/1888451.pdf. (accessed: March 03, 2008)

Papke, L., and Wooldridge, J. (1996). Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates. *Journal of Applied Econometrics*, *11*(6), 619–632.

Parker, E. (2000). Closing the Digital Divide in Rural America. *Telecommunications Policy*, *24*(4), 281–290.

Prince, J. (2008). Repeat Purchase amid Rapid Quality Improvement: Structural Estimation of Demand for Personal Computers. *Journal of Economics & Management Strategy*, *17*(1), 1–33.

Rogers, E. (2003). *Diffusion of Innovations* (5th ed.). Free Press, New York, USA.

Sinai, T., and Waldfogel, J. (2004). Geography and the Internet: Is the Internet a Substitute or a Complement for Cities? *Journal of Urban Economics*, *56*(1), 1–24.

Statistisches Bundesamt. (2006). Internetnutzung und die Aneignung von E-Skills. *Wirtschaft und Statistik*(7), 722–733.

Strover, S. (2001). Rural Internet Connectivity. *Telecommunications Policy*, *25*(5), 331–347.

TNS Infratest. (2002). *(N)Onliner-Atlas 2002 – Eine Topographie des digitalen Grabens durch Deutschland.*

TNS Infratest. (2007). *(N)Onliner-Atlas 2007 – Eine Topographie des digitalen Grabens durch Deutschland.*

Wagner, J. (2001). A Note on the Firm Size – Export Relationship. *Small Business Economics*, *17*(4), 229–237.