Matematisk statistik
Stockholms universitet

Relationship between Broadband and Productivity amongst Swedish Companies 2001–2005

Jennie Glantz

Examensarbete 2008:11

Postal address:

Matematisk statistik
Dept. of Mathematics
Stockholms universitet
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se/matstat

Matematisk statistik
Stockholms universitet
Examensarbete **2008:11**,
http://www.math.su.se/matstat

Relationship between Broadband and Productivity amongst Swedish Companies 2001–2005

Jennie Glantz*

September 2008

Sammanfattning

In this degree project we have studied the relationship between broadband and productivity amongst Swedish companies during 2001–2005. It is a part of an ongoing project at Statistics Sweden.

We used a model with three equations to study the relationship:

1. The broadband equation explains what kind of companies that chooses to acquire/ have broadband.
2. The Internet equation explains the connection between Broadband and IT-level (a variable that explain the internet-activities at a company).
3. The third equation is a productivity equation, where we had a productivity variable as Response variable, and IT-level was here an explanatory variable.

To estimate the equation parameters we used a method called 3SLS (Three stage least squares) that could deal with the problems we had in our dataset, for example correlated error terms. In the study we have found a positive significant relationship between Broadband and Productivity among Swedish companies during 2001–2005. The relationship depends on the branch to which the company belongs, the size of the company, and if the company is a multinational company or not.

*Postal address: Matematisk statistik, Stockholms universitet, SE-106 91, Stockholm, Sweden. E-mail: jennieglantz@hotmail.com. Handledare: Rolf Sundberg

Foreword

In this master thesis we have analysed the relationship between *Broadband* and *Productivity* amongst Swedish companies during 2001-2005. It is a part of an ongoing project by the Organisation for Economic Co-operation and Development (OECD) which Statistics Sweden is participating on.

I would like to thank Hans-Olof Hagén and Caroline Ahlstrand at Statistics Sweden for providing me with data, a lot of necessary information and help with the three main equations in the thesis. I would also like to thank Rolf Sundberg my supervisor at Stockholm University for valuable ideas and discussion.

Stockholm 20/8 - 2008
Jennie Glantz

Table of contents

1. Introduction

The development of computer technology over the last ten years has truly approved a lot among companies in the world. Internet is one of the most important innovations of our time and it brings important profit to economies and societies. Internet enables new possibilities for companies all over the world. In this degree project we have studied the relationship between broadband and productivity among Swedish companies during 2001-2005.

The Analysis is based on IT-data and Economic Data from Swedish companies that cover the period 2001-2005. We want to find out if there is a relationship between broadband and productivity. Our Model includes three equations where the first equation is free-standing from the two others.

1. The *Broadband equation* will explain which kind of companies that chooses to have/acquire broadband and what kind of variables that impact on the selection.

2. The *Internet equation*, will explain the connection between Broadband and IT-level (a variable that explain the internet-activities in a company). IT-level is here a response variable and broadband-speed is one of the explanatory variables.

3. Finally, the *Productivity equation* relates productivity to IT-level. In this equation we have a productivity variable as a Response variable, and IT-level is here an explanatory variable.

We estimate our model using econometric methods that can deal with the problems that belong to this kind of data, for example correlation between the disturbances in the two different equations.

Several studies have been made to analyse the relationship between Innovation activities and Productivity in Sweden and other European countries[1]. In this degree-project we will focus on the connection between broadband and productivity, we want to see if broadband is an innovative process that affects the company's productivity. A parallel study has been made by Malin Nilsson[2], in that degree-project the focus is on the relationship between broadband and internet-use.

Keywords: Productivity, Broadband-speed, IT-level.

---

[1] See for example:

Lööf & Heshmati (2006) '*On the relationship between innovation and performance: a sensitivity analysis'*. Economics of Innovation and New Technology, Vol. 15, No. 4 & 5, pp.317-344

Griffith, R., Huergo, E., Mairesse, J. & Peters, B. (2006) *'Innovation and Productivity across four European Countries'*. NEBR working paper no. 12722, Cambridge, MA

Hagén, H., Ahlstrand, C. Daniels, M., (2007) '*Innovation matters; An empirical analysis of innovation 2002-2004 and its impact on productivity'* Stockholm: Statistics Sweden

[2] Nilsson M. (2008) *Relationship between Broadband and Internet-use amongst Swedish companies 2001-2005* Master thesis, Stockholm University.

2. The data

The data used in this paper are collected from five surveys made by Statistics Sweden during the period 2001-2005. The survey contained questions about the internet activities at Swedish companies. The survey includes all the big companies (with more than 250 employees) in Sweden and a selection of the small companies. Every year one third of the small companies are replaced. The survey was voluntary so unfortunately we did not have answers from all participating companies every year.

We also had access to a database, with economic information from all Swedish companies. We have constructed some own variables from the survey and the database that we needed for the analysis.

A problem we had was that the surveys did not contain the same questions every year. One of our most important variables, IT-level, is different the first year, (2001). The survey from 2001 did not contain two variables, E-purchase and Business-Activities, which we needed to calculate IT-level. To be able to make a comparison between IT-Level different years we calculated the mean-value for E-Purchase and Business-Activities from 2002 and used them when we calculated IT-level for 2001. We also removed some outliers that had obviously unreasonable values. See Appendix A for more information about the variables.

In some of the analyses we made, we wanted to have the opportunity to follow some variables over time. We had the economic information from the database about every company, whether they had participated in the survey or not, but of course, we only had information about their IT-use if they had answered the questions. To be able to make a comparison between the Internet-variables over time we had to restrict the datasets in some tests to companies that had answered the questions at least two years.

To get a "feeling" for the datasets we created some different diagrams that we present on the following pages.

Diagram 1:

Number of participating companies

Diagram 2:

Number of companies participating at least two years.

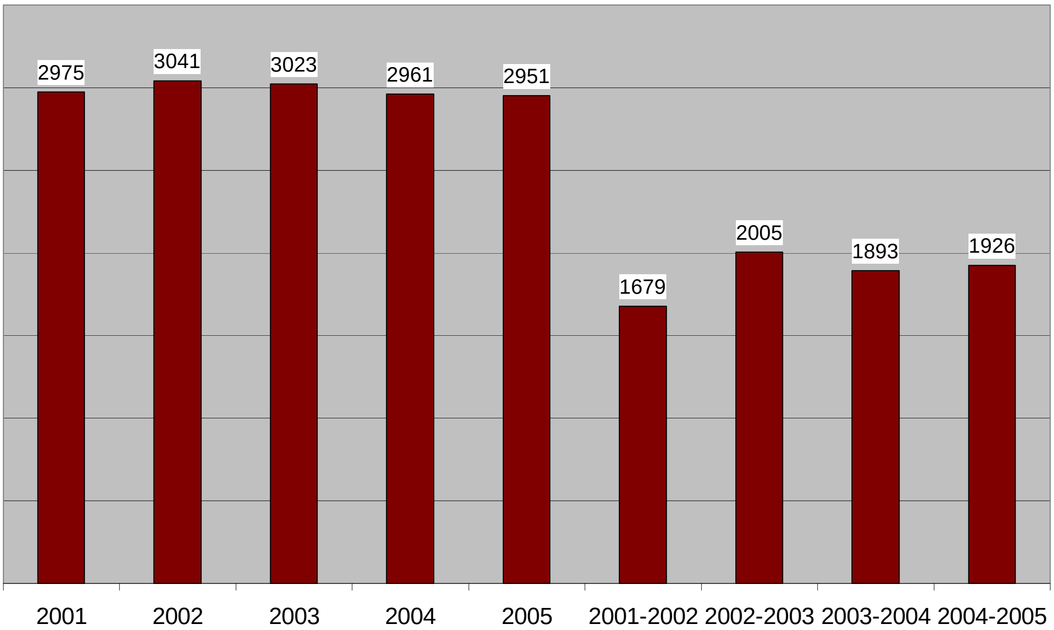
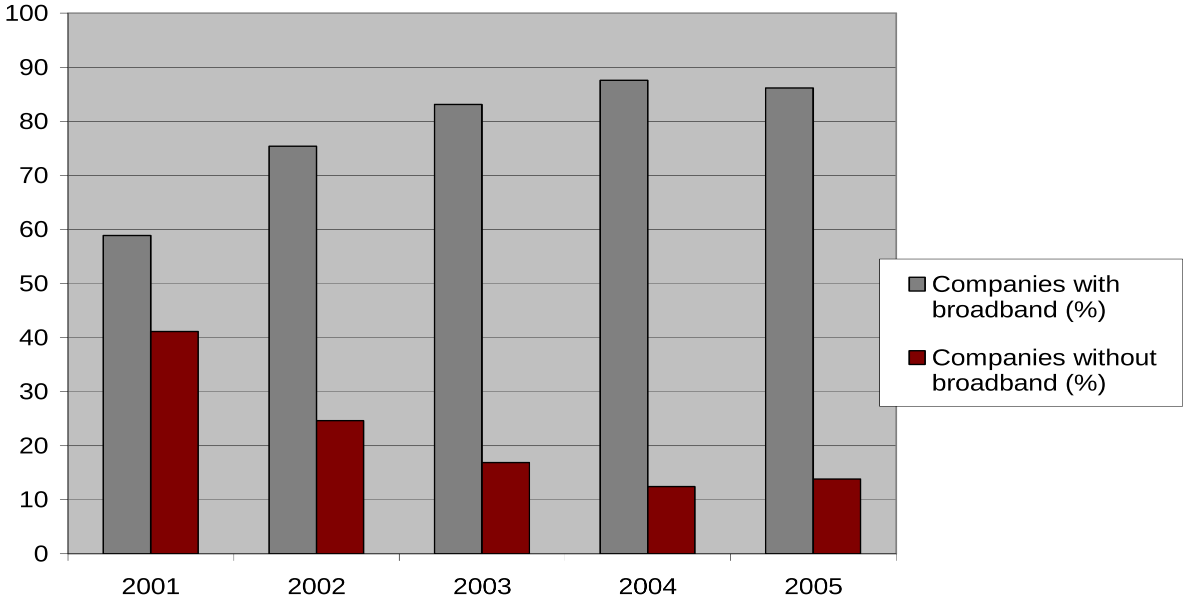


Diagram 3

Number of Companies with broadband / without broadband



It was more common not to have broadband in the beginning of the survey. In the end almost every company had broadband.

Diagram 4:

IT-level with /without broadband

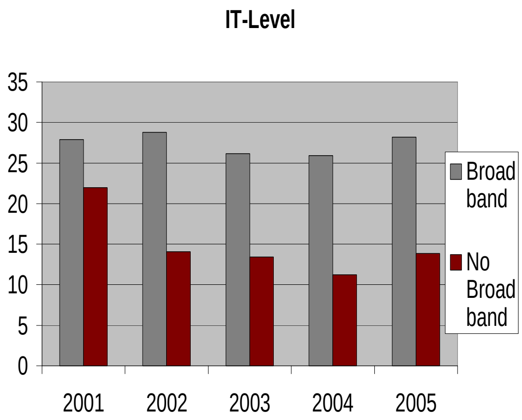


Diagram 5:

Productivity with /without broadband

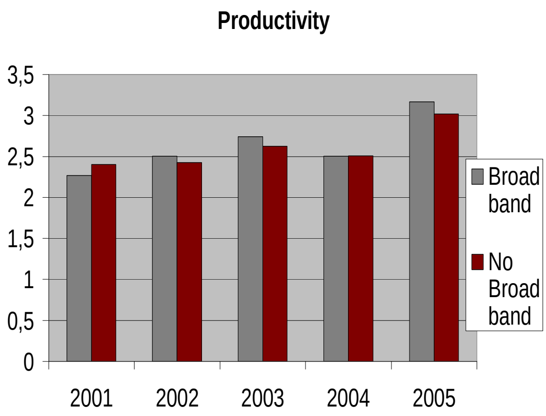
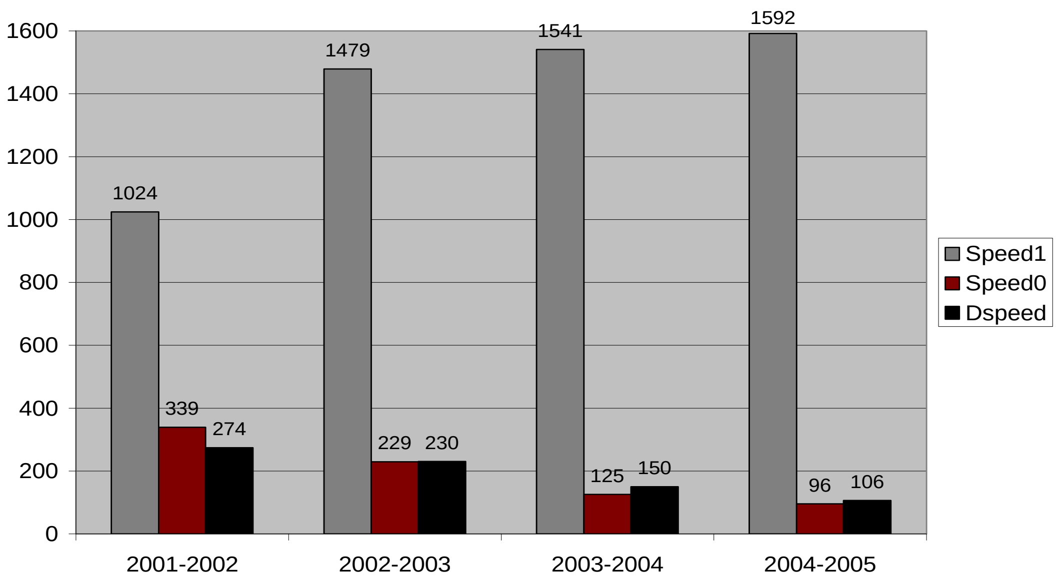


Diagram 6:

Companies that acquire broadband



Speed1 = Companies that had broadband already first year.

Speed0 = Companies that did not have broadband even the second year

Delta-speed = Companies that acquired broadband from the first year to the second.

3. The Model and Methods

3.1 Background:

The relationship between innovation and productivity has been a topic of study for a long time all over the world. Many studies have been made to better understand the connection between productivity and innovation.

In 1979 Grichlies was one of the first trying to find the link between innovation and productivity. In 1998 Crepon, Duguet and Mairesse (CDM)[3] developed Grichlies' model in an attempt to better understand the link between productivity, innovation and research at manufacturing companies in France. They introduced some new features in their innovation analysis. Their model included four equations between productivity, innovation input and innovation output.

1) $g_i = x_{0i}b_0 + u_{0i}$
2) $k_i = x_{1i}b_1 + u_{1i}$
3) $t_i = \alpha_K k_i + x_{2i}b_2 + u_{2i}$
4) $q_i = \alpha_T t_i + x_{3i}b_3 + u_{3i}$

The first equation is a selection equation that determines if the company is engaged in research activities where g_i is a latent (unobservable) dependent variable. The company invests in research if $g_i > c$, where c is a common threshold. The second equation shows the size of the research activities at the company where k_i is the research capital per employee. Equation 3 is an Innovation equation where t_i shows the share of innovative sales. The innovation output is here measured by the number of patents. Their last equation is a Productivity equation and it explains how much the innovation affects the productivity and the variable q_i is the logarithm of labour productivity.

[3] Crépon, B., Duguet, E. & Mairesse, J. (1998) Research, Innovation, and productivity: An econometric analysis at the firm level. *Economics of Innovation and New Technology,* Vol. 7, No.2, pp.115-158

In the present analysis our focus is too see if the process of acquiring broadband is an innovative process that affects the company's productivity. We will use a similar model as Crepon et al. since we have a similar problem and similar types of data. We estimate the relationship between broadband and productivity in an equation-system with three equations described below.

3.2 Equation 1:

The first equation is a part of the model where we try to investigate which variables have impact on the decision to innovate, in this case to acquire/have broadband. We used two different response variables in this equation, delta-speed and speed. Delta-speed is an indicator variable for companies that has acquired broadband from one year to another, in these tests we compared the companies that just acquired broadband to the companies that did not have broadband. The other response variable we used, speed, is an indicator variable with 1 if the company has broadband and 0 if it does not have broadband. With delta-speed as response variable we can study the variables that impacts on the decision to acquire broadband and with speed as a response variable we can study the variables that impacts on the companies that already have broadband.

The explanatory variables were; eleven dummies for type of industry, two dummies for company size, three dummies for the geographic market, a variable that describes labour quality, and four variables that explains how the company uses internet: If they have intranet, extranet, LAN (Local Area Network) or WLAN (Wireless Local Area Network). Since our response variable is a 0-1 variable we used a logistic regression to estimate this model to describe the influence of the explanatory variables.

$$\text{Log}\,(p / (1\text{-}p)) = \alpha + \sum_i \beta_i x_i$$

P is here the probability for the response variable to assume the value 1. x_i are all the explanatory variables, β the coefficient for each x. See Appendix C for a statistical explanation of Logistic regression.

3.3 Equations 2 and 3:

Our main question in this analysis was *"Does broadband-speed has any impact on productivity?"* We did not believe that speed in itself had any influence on the company's productivity: we thought that it was more likely that the level of internet-use in the company had some influence. We therefore used two equations: In equation 2 we had IT-level (a variable that explains the internet-use at a company) as a response variable (h) with a broadband-indicator as an explanatory variable and in equation 3 we had productivity (t) as a response variable with IT-level as an explanatory variable.

Equations 2 and 3:

$$h = \sum_i x_i \beta + \varepsilon_i$$
$$t = \hat{h} + \sum_i z_i \beta + \varepsilon_i$$

The explanatory variables x in equation 2 are one dummy for broadband, eleven dummies for type of industry, two dummies for the company size, three dummies for the geographic market the company sells to, four dummies for different Internet-use variables and labour quality.

t = Productivity, here measured as LnGPMFP, The Logarithm of (Gross Production Multi Factor Productivity). LnGPMFP is calculated as a function of economic variables. This estimator of the productivity takes account of more production factors than just labour quality. We also tried another productivity-variable; it gave similar results so we will only present the results with LnGPMFP. See Appendix A for more information about LnGPMFP.

The explanatory variables z, in the third equation are almost the same as in the second equation. The difference is that we do not have any dummies for internet-use and we have IT-level as an explanatory variable instead.

We also tried a model with a single equation relating productivity directly with broadband status. We wanted to see if there was any difference in productivity between companies that acquired broadband and companies that did not acquire broadband. The explanatory variables we used were the same as in equation 2, (One dummy for broadband, eleven dummies for type of industry, two dummies for the company-size, three dummies for the geographic market the company sells to, four dummies for different Internet-use variables and labour-quality).

3.4 2SLS, 3SLS and SUR

In a conventional regression model all the explanatory variables must be exogenous (the explanatory variables must be explained outside the model; the response variable is not allowed to have any influence on the explanatory variables). However it is reasonable to think that IT-level and Broadband-speed had influence on each other. That is a violation of the OLS (Ordinary Least Square) assumptions. Therefore we had to use a model that could deal with endogenous (when the response variable has influence on the variable) variables. Another problem we had was that equations 1, 2 and 3 came from the same dataset and it is therefore likely that the error terms in the equations are correlated.

Two-stage least squares (2SLS) is useful in regression equations when some of the variables are correlated with the error term (endogenous variable). 2SLS replaces the endogenous variable with a regression on instrument variables. An instrumental variable should be uncorrelated with the error term but highly correlated with the explanatory variables in the equation. The regression on the instrumental variables will replace the variable that is correlated with the error term.

Stage one: Estimate the problematic variable with the instrumental by OLS
Stage two: Replace the right-hand side endogenous variables with the new estimated variable and estimate the equation by OLS, (Ordinary Least Square).

Seemingly unrelated regression (SUR) methodology can deal with the problem of correlated error terms between two equations. If the equations use the same data the errors may be correlated across the equations. In our case it would be unrealistic not to allow correlation between the error terms in equation 2 and 3. 3SLS is a combination of 2SLS and SUR. See Appendix D for more information about SUR, 2SLS and 3SLS.

4. Results

4.1 Variabels Definition

Under 10	Dummy for small companies
Over 250	Dummy for large companies
Labour Quality (LQ)	A measure of the quality of employees
Intranet	Dummy for intranet-availability
Extranet	Dummy for extranet-availability
Speed	Dummy for broadband-availability
Delta-Speed	Dummy being 1 if the company has acquired broadband from one year to next
LAN	Dummy for Local Area Network
WLAN	Dummy for Wireless Local Area Network
Persint	Share of the employees that has access to internet
LnGPMFP	A productivity variable

The equations are controlled for other factors in the model, which industry the company belongs to and which geographic market the company sells to.

4.2 Equation 1[4]

In this equation we study what kind of variables that have influence on the decision to acquire broadband. The dependent variable, delta-speed, is an indicator variable being 1 if the company has changed to broadband and being 0 if the company did not change. In this test we restricted the dataset to only consisting of the companies that did not have broadband at the beginning of the two-year-period. Thus we can compare the companies that have just acquired broadband with the companies that still do not have broadband.

We also used this equation to characterize what kind of variables that are typical for companies that already have broadband, independent of how long they have had it. The dependent variable is the indicator variable, speed. We compare the companies that have broadband one year with those that did not have broadband the same year.

[4] ”Equation 1” is written together with Malin Nilsson

We did this test with different explanatory variables. We had two different measures on the employees, university level and labour quality. Labour quality gave more significant results and therefore we decided to use that variable. One reason that Labour quality gave a better result can be that it takes more factors into account. The reason why we also tried to test the variable university level is that labour quality can be expected to have both an increasing and decreasing effect on the response variable. One variable that we think have a decreasing effect is age, it is more likely that younger people use internet more and have more knowledge about it.

In these studies we used logistic regression to model the impact of variables on broadband. We obtained the following results for the first time-period of two years. For results other years see Appendix B.

Companies 2001-2002

Dependent variable: delta-speed 01-02

	Coefficient	**Coefficient***	**Std-error**	
Under 10 01	0.36	0.08	0.38	
Over 250 01	1.89	0.40	0.53	a
Labour Quality 01	0.00	0.20	0.06	c
Intranet 01	0.79	0.35	0.23	a
Extranet 01	0.03	0.01	0.37	

a Significant at 1%
b Sicnificant at 5%
c Significant at 10 %

(Here all companies started the time-period without broadband)

* =Coefficients for variance-standardized explanatory variables

(The survey from 2001 did not contain LAN, WLAN and PersInt)

The result shows that, among companies that not already having broadband, large companies where more inclined than small and medium sized companies to acquire broadband. We can also see that intranet has a strong significant effect and labour quality has some indication of effect.

Dependent variable: speed02

	Coefficient	Coefficient*	Std Error	
Under 10 02	-0.24	-0.05	0.31	
Over 250 02	1.85	0.84	0.33	a
Labour Quality 02	0.00	0.18	0.00	
Intranet 02	0.65	0.31	0.20	a
Extranet 02	0.19	0.09	0.31	
LAN 02	0.89	0.35	0.19	a
WLAN 02	1.05	0.45	0.30	a
Pers Int 02	0.02	0.83	0.00	a

* = Coefficients for variance-standardized explanatory variables

We can see that it is more common that large companies have broadband than medium sized or small companies. The reason why we did not get any significant results from the small companies could be that we do not have so many observations in that group. It could also be that they do not differ from the medium size companies when it comes to having broadband.

4.3 Equations 2 and 3:

In equation 2 speed was an endogenous variable and therefore we had to estimate the speed-variable with instrument variables. We first used some variables from the survey as instrumental, but the results we received from these tests with many variables as instrument were strange. Probably because the instrument we used to explain our endogenous variable was a linear combination of the other variables we had in our model. To get around that problem we used a lagged variable of speed, (the lagged variable is not endogenous). We then received results that were easier to interpret. It is actually likely that it takes some time before speed impacts on IT-level, so maybe it is even better to use speed from previous year. This conveys that we had a three-year period in these tests and therefore we were only able to do it three times.

We also restricted our dataset to consisting only of companies that started the period without broadband and looked at the difference in productivity between two years at the companies that acquired broadband against the companies that did not. In these tests we only used one equation with the difference in productivity as a dependent variable and speed as an explanatory variable.

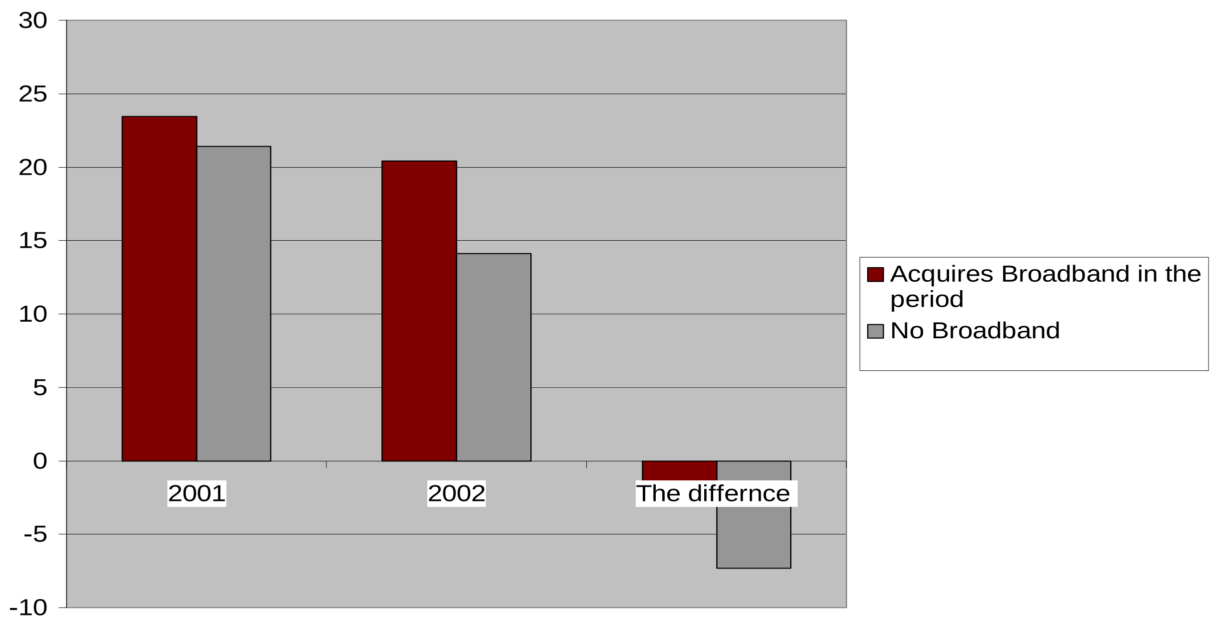
We will only present our results from one dataset here, for more results from other years see Appendix B.

Dataset companies participating 2001-2002

Results from equation 2 (the Internet-equation)

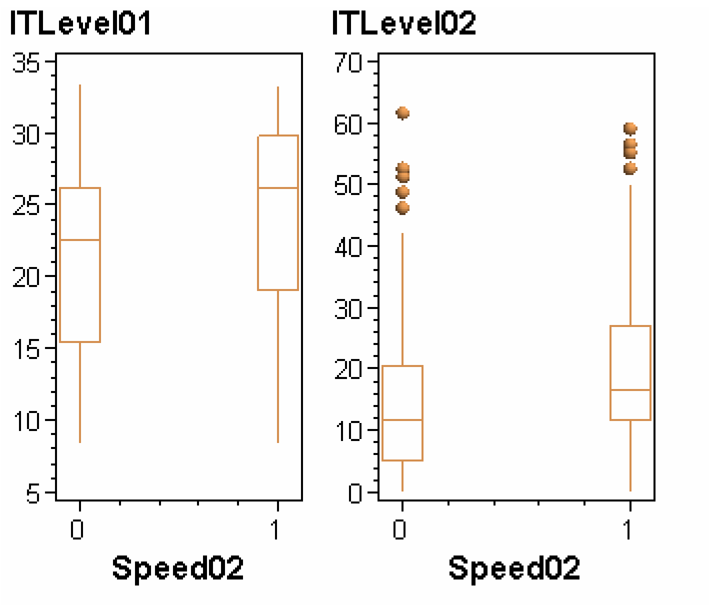
In the diagram (7) below we can see the difference in mean-value for IT-Level for companies that acquire broadband between 2001 and 2002 and companies that did not acquire broadband. Note that IT-level from 2001 is different (see p.5) so this explains that the average of IT-level is higher 2001. We can see that it looks like IT-level is higher for the companies that acquired broadband both before they acquired broadband and after. It is a smaller decrease in IT-level for the companies that acquired broadband compared to the companies that did not acquire broadband.

Diagram 7



In diagram 8 and 9 we have box plots over IT-level 2001 and 2002 for companies that acquired broadband (speed02 = 1) and companies that did not acquire (Speed02 = 0) broadband.

Diagram 8 **Diagram 9**



We can see in these box-plots that the companies that acquired broadband between 2001 and 2002 have on average a higher IT-Level already before acquiring broadband compared to the companies that did not acquire broadband.

Equation 2 - Dependent variable IT-level 02.

	Coefficient	Coefficient*	Std Error	
Under10 02	-3.00	-0.66	1.69	c
Over250 02	1.30	0.59	0.94	
Labour Quality 02	-0.01	-0.54	0.01	a
Intranet 02	2.85	1.38	0.97	a
Extranet 02	6.56	3.02	0.95	a
LAN 02	4.03	1.58	1.06	a
WLAN 02	5.55	2.35	0.88	a
PersInt 02	0.06	2.23	0.01	a
Speed 01	4.51	2.17	0.95	a

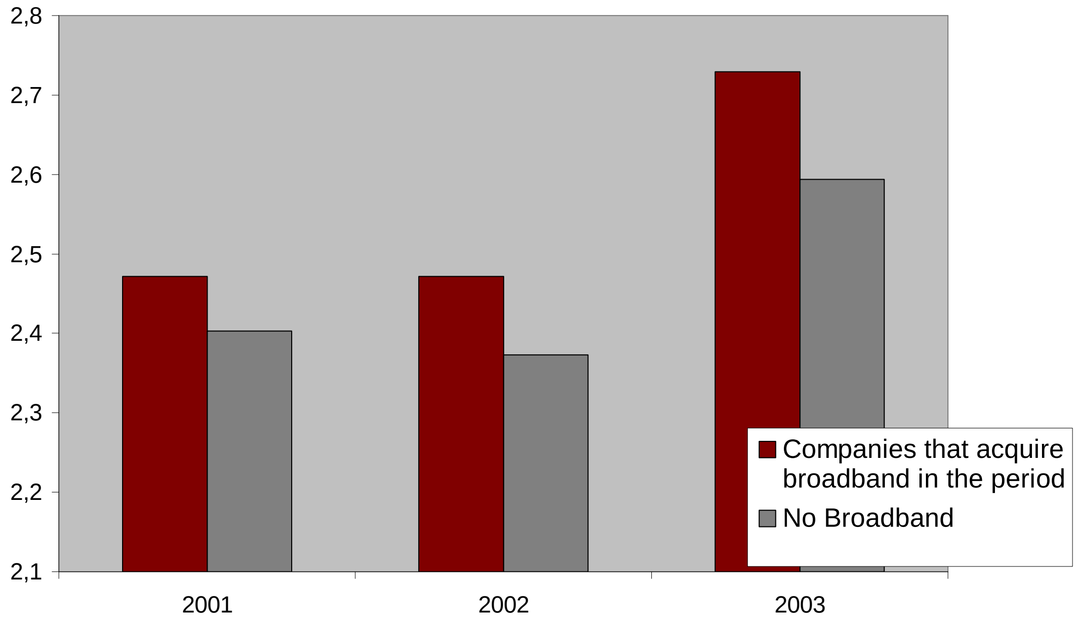
* =Coefficients for variance-standardized explanatory variables

In the table above we can see that we received a lot of significant results from our regression. Speed has a strong significant impact on IT-level. One surprising result is that Labour Quality has a negative impact on IT-level. Labour Quality contains one variable, age, which can have a decreasing effect on IT-level.

Results from equation 3 (the productivity equation)

In the diagram (10) below we can see the difference in productivity (mean-value) between companies that acquired broadband between 2001 and 2002 and the companies that did not acquire broadband.

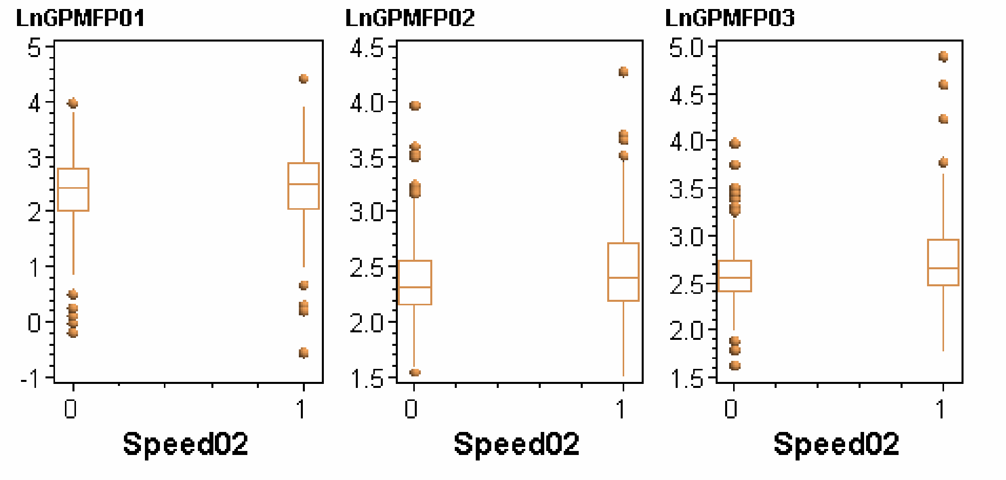
Diagram 10: Productivity different years



Here it looks like the difference in productivity between the companies that acquired broadband between 2001 and 2002, and the companies that did not acquire broadband is highest 2003. It is likely to think that it takes sometime before broadband gives any effect on the productivity.

In Diagram 11 below we can follow the productivity for the companies that acquired broadband between 2001 and 2002 (speed02 = 1) and for the companies that did not acquire broadband (speed02 = 0) in a box-plot instead. In the box-plots it looks like the median for the companies that acquire broadband is a little bit higher every year compared to the companies that did not acquire broadband.

Diagram 11



Equation 3 – Dependent variable LnGPMFP03

	Coefficient	Coefficient*	Std-Error	
Under10 03	0.04	0.01	0.06	b
Over250 03	-0.01	0.00	0.04	
Labour Quality 03	0.00	0.04	0.00	a
ITLevel 02	0.00	0.04	0.00	b

* =Coefficients for variance-standardized explanatory variables

In equation 2 with IT-level as a dependent variable we received many significant variables in our test. It is not a surprising result that we got some internet-variables that have impact on IT-Level. More interesting is that it looks like speed has a strong impact on IT-Level. We can see a high coefficient and it is a significant result. In our next equation with Productivity as a dependent variable we received three significant variables. Most interesting is that IT-Level has a significant positive impact on productivity.

In our other testes we made, we had the difference between productivity (different years) as a dependent variable and speed as an explanatory variable. We restricted the test to only consisting companies that did not have broadband from the beginning. We received following result from the first period:

Dependent variable: “Delta-productivity” 01-03

	Coefficient	Coefficient*	Std-Error	
Speed 02	0.09	0.04	0.04	b
Under 10 01	0.11	0.02	0.08	
Over 250 01	-0.05	-0.01	0.09	
LQ 01	0.00	0.01	0.00	
Intranet 01	0.05	0.02	0.05	
Extranet 01	0.07	0.02	0.08	

* =Coefficients for variance-standardized explanatory variables

In this test we had the difference in productivity between 2003 and 2001 (for the companies that acquired broadband and the companies that did not acquire broadband), we did not receive any significant results when we had the difference between 2002 and 2001. It is more likely that it takes some time before we can see any results. As we can see speed is the only significant variable that has some impact on the difference in productivity.

5. Discussion

In this project we have tried to find a positive relationship between broadband-speed and productivity. Our results show that in the most of the datasets we can see a positive relationship between speed and productivity.

What kind of company chooses to acquire/have broadband? Nowadays most of the companies in Sweden have broadband, but only five to ten years ago it was not like that. Has an earlier obtaining of broadband resulted in high productivity? Our first thought was that big companies with employee that had education decided early to innovate and obtain broadband earlier than other companies. The results we received in our first equation shows that our thoughts were not totally wrong. It seems like the size of the company has impact on the obtaining of broadband.

Our model was also controlled for which type of industry the company belonged to. We decided not to present any results from the industry-dummies since we are not interested in which industry we can find a connection between productivity and broadband. Our results show that the industry the company belongs to had impact on the link between productivity and broadband. That is not a surprising result, of course some industries brings more benefit from broadband than others. Another thing our model was controlled for was if the company had a concern. If they had, in which country, these dummies gave us very different results, some years with a significant result and some years without. We decided not to present the results but still have them in the model.

Speed gave us a significant result in every dataset in equation 2. Since our main-question was, if broadband-speed has any impact on the productivity this variable was the most important in this equation. But we can also see that some different IT-variables got positive impact on IT-Level. That is not so strange, if the company for example have intranet or extranet it is more likely that they use internet for more things and therefore they will receive a higher IT-level. It is different from year to year if the size of the company has any impact on IT-level, it is hard to have any explanation why it is like that, but maybe the difference between the participating companies that answered made the results like that. Also this question was controlled for which industry the company belonged to and if the company had a concern and for the same reason as earlier we decided not to present those results, but we could see that which industry the company belonged to had an impact on IT-level.

Maybe broadband sometimes can make a lower productivity at a company. A guy that worked at a company were they changed back to dialup internet says that this made him more productive. *"..it has forced me to become more focused and to prioritize my tasks every day"* [5]. We think this is an exception; it is hard to believe that it is common that broadband actually makes a lower productivity, but it is anyway an interesting thought.

In our last equation, "*The productivity equation*", we can observe that in the most of our datasets, IT-level has a positive significant impact on productivity. We received one strange value from dataset 2002-2003. It-level had a negative impact on the productivity. We can see in diagram 5 (p.7) that productivity decreases 2004. It looks like a general decrease in productivity and maybe IT-level is not an important variable in "decreasing-times".

In the tests we made with the difference in productivity as a dependent variable and speed as an explanatory variable we received speed as an significant variable in the first dataset from 2001-2002. The datasets we used in these tests gets smaller and smaller every year so that can be an explanation why we did not receive any

[5] http://www.lifehacker.com.au/tips/2007/12/14/can_dial_up_internet_make_you.html

significant results later. Interesting is that in the first test speed is the only variable that is significant.

These tests show that speed has a positive impact on IT-level in all the tests and IT-level has a positive impact on the productivity in most of the tests we have made. The relationship depends on which industry the company belongs to, the size of the company, if the company is a multinational company.

6. References

Bollen, K.A (1996) 'An Alternative Two Stage Least Squares (2SLS) Estimator for Latent Variable Equations.' Psychometrika, 61: 109-121.

Crépon, B., Duguet, E. & Mairesse, J. (1998) Research, Innovation, and productivity: An econometric analysis at the firm level. *Economics of Innovation and New Technology,* Vol. 7, No.2, pp.115-158

Griffith, R., Huergo, E., Mairesse, J. & Peters, B. (2006) *Innovation and Productivity across four European Countries*. NEBR working paper no. 12722, Cambridge, MA

Hagén, H., Ahlstrand, C. Daniels, M., (2007) '*Innovation matters; An empirical analysis of innovation 2002-2004 and its impact on productivity'* Stockholm: Statistics Sweden

Heckman, J.J (1979) Sample Selection Bias as a Specification Error, *Econometrica,* Vol. 47, No.1, pp.153-161

Kline, Rex B. (1998). *Principles and practice of structural equation modeling*. NY: Guilford Press, pp.175-180

Lööf & Heshmati (2006) On the relationship between innovation and performance: a sensitivity analysis. *Economics of Innovation and New Technology,* Vol. 15, No. 4 & 5, pp.317-344

Maddala, G.S.(2001) '*Introduction to Econometrics'* Third edition pp. 318-326 343-366

Nilsson M. (2008) *Relationship between Broadband and Internet-use amongst Swedish companies 2001-2005* Master thesis, Stockholm University

Pindyck, R., Rubinfeld, D. (1998) ‘*Econometric Models and Economic Forecasts’* Fourth edition, pp. 349-363, 371-376

Zellner, A. & Theil, H. (1962) Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Estimations. *Econometrica,* Vol.30, No.1, pp. 54-78

Literature-study

Adermon, A., Nilsson, E. (2007) ‘*Innovation and Productivity amongst Swedish Firms 2002-2004; An Empirical Analysis’* Bachelor Thesis for Statistics Sweden and Uppsala University.

Berry, W. D. (1984). *Nonrecursive causal models*. Beverly Hills, CA: Sage Publications, Ch. 5. Covers 2SLS with emphasis on the identification problem.

Powell, J. L Zellner´s *Seemingly Unrelated Regressions Model*. Department of Economics University of California, Berkeley

Theil, H. (1971), *Principles of Econometrics*, New York: John Wiley & Sons, Inc.

7. Appendix A

Our data

IT-variables from the survey: (Dummy-variables if nothing else is written)

PersInt – How many employees has internet (%)

Internet-use-variables:

Financial Transaction	*Education*
Download	*Homepage*

Marketwatch (information only from some years)

Internetactivities - % from Internet-use (If the company use internet for everything above the internet-activities is 100 %)

What can the company do with their IT-system?

Intern systems for orders in the company	*Payment system*
Logistic-system	*Customer business system*
Supplier business system	

Businessactivities - % from IT-system (If the company can use their IT-system to everything above their business-activities is 100 %)

Other information about their IT-system

E-sales (%)	*E-Purchase (%)*
Service support	*Homepage*
Intranet	*Extranet*
LAN (Local Area Network)	*WLAN (Wireless Local Area Network)*

IT-Level:

(E-sales + E-Purchase + Internetactivities + Businessactivities) / 4

(The survey from 2001 did not contain E-purchase and Businessactivities)

Broadband-Level:

Broadband over 8 MB	*ISDN*
Broadband under 8 over 2 MB	*Mobile*
Broadband under 2 MB	*Modem*

Economic variables from the database:

NR-branch – 11 dummies for different industries
MultiSwe – If the company has a concern in Sweden
MultiUSA – If the company has a concern in USA
MultiOvr - If the company has a concern somewhere else
lnL – How many employees
Labour Quality – Labour Quality
Universitylevel – How many employees has a 3-years education (%)
Prod_FP – Production (Inflation-secured)
Forbruk_FP – What the firm consume (Inflation-secured)
Va_FP – (Prod – forbruk) (Inflation-secured)
Kapital_FP – How much capital the firm has
Expint_FP – Exportintensity (Inflation secured)
Loneapprox – Wages-approximation

Created Variables
Intermediateshare – (Forbruk/prod)
WageshareVA – (-0.001*loneapprox/va)

GPMFP *(Gross Production Multi Function Productivity)* – (prod_fp)-(Median (Intermed)*(forbruk_fp))-(Median(wageshareprod)*(LQ))-(1-Median (Intermed.)- Median(wageshareproduction))*(kapital_fp))

This variable will measure the productivity and will handle Labour, capital and the consume-share.

We used the median in many variables instead of the mean-value this is because we had some extreme values and the median gave us therefore a better value.

8. Appendix B

Results from other years

Dataset companies-0203

Equation 1: Dependent variable speed03

	Coefficient	Coefficient*	Std Error	
Under 10 03	-0.19	-0.04	0.29	
Over 250 03	1.45	0.62	0.45	a
Labour Quality 03	0.00	0.45	0.05	a
Intranet 03	0.45	0.22	0.20	b
Extranet 03	0.64	0.29	0.35	c
LAN 03	1.65	0.58	0.18	a
WLAN 03	0.96	0.42	0.34	a
Pers Int 03	0.14	5.36	0.03	a

a Significant at 1%
b Significant at 5%
c Significant at 10 %

Equation 2: Dependent variable IT-level03

	Coefficient	Coefficient*	Std-Error	
Under10 03	0.20	0.05	1.50	
Over250 03	2.25	5.28	0.91	b
Labour Quality 03	0.00	0.17	0.00	
Intranet 03	4.81	2.35	0.88	
Extranet 03	4.47	2.03	0.88	a
LAN 03	7.05	2.47	1.07	a
WLAN 03	4.64	2.03	0.81	a
PersInt 03	0.05	2.00	0.01	a
Speed 02	2.71	1.14	0.91	a

Equation 3: Dependent variable Productivity04 (LnGPMFP04)

	Coefficient	Coefficient*	Std-Error	
Under10 04	0.11	0.03	0.07	
Over250 04	-0.01	0.00	0.05	
Labour Quality 04	0.00	0.16	0.00	a
ITLevel 03	-0.01	-0.10	0.00	b

In this test we received a strange result. IT-level has a negative significant impact on productivity. We can also see some unsurprising results, Some Internet-use-variables has a positive impact on IT-level.

Dataset Companies-0304

Equation 1: Dependent variable speed04

	Coefficient	Coefficient*	Std Error	
Under 10 04	-0.44	-0.10	0.32	
Over 250 04	-0.02	-0.01	0.37	
Labour Quality 04	0.00	-0.20	0.00	
Intranet 04	0.69	0.33	0.23	a
Extranet 04	0.95	0.44	0.40	b
LAN 04	1.47	0.49	0.21	a
WLAN 04	0.88	0.42	0.30	a
Pers Int 04	0.03	1.19	0.00	a

Equation 2: Dependent variable IT-level04

	Coefficient	Coefficient*	Std-Error	
Under10 04	-3.57	-0.79	2.50	b
Over250 04	2.74	1.16	0.91	a
Labour Quality 04	0.00	-0.77	0.00	
Intranet 04	3.85	1.85	0.91	a
Extranet 04	4.52	2.08	0.86	a
LAN 04	6.31	2.12	1.26	a
WLAN 04	4.06	1.92	0.79	a
PersInt 04	0.08	3.08	0.01	a
Speed 03	3.00	1.06	2.43	a

Equation 3: Dependent variable Productivity05 (LnGPMFP05)

	Coefficient	Coefficient*	Std-Error	
Under10 05	0.66	0.30	0.06	a
Over250 05	-0.67	-0.10	0.17	a
ITLevel 04	0.01	0.13	0.00	c

In this test we can see that speed has a positive significant influence on IT-level and that IT-level has a positive significant influence on productivity. We can also see that the size on the company have positive impact on the productivity.

Dataset companies – 0405

Equation 1: Dependent variable speed05

	Coefficient	Coefficient*	Std Error	
Under 10 05	-0.21	-0.09	0.22	
Over 250 05	0.35	0.06	0.75	
Labour Quality 05	Non excisting	Non excisting	Non excisting	
Intranet 05	0.68	0.33	0.21	a
Extranet 05	0.26	0.12	0.27	
LAN 05	Non excisting	Non excisting	Non excisting	
WLAN 05	0.94	0.46	0.23	a
Pers Int 05	0.15	5.21	0.03	a

Dataset from the companies that acquired broadband between 2002-2003

Equation 1 with dependent variable: Delta-speed 0203

	Coefficient	Coefficient*	Std-error	
Under 10 02	0.23	0.06	0.40	
Over 250 02	0.01	0.00	0.79	
Labour Quality 02	0.00	0.14	0.83	
Intranet 02	0.38	0.17	0.31	
Extranet 02	-0.64	-0.15	0.61	
LAN 02	0.70	0.35	0.24	a
WLAN 02	0.43	0.10	0.55	
Pers Int 02	0.02	0.53	0.00	a

Dataset from the companies that acquired broadband between 2003-2004

Equation 1 with dependent variable: Delta-speed 0304

	Coefficient	Coefficient*	Std-error	
Under 10 03	-0.75	-0.21	0.55	
Over 250 03	12.26	1.82	44.90	
Labour Quality 03	0.00	0.31	0.00	
Intranet 03	1.06	0.46	0.44	b
Extranet 03	1.63	0.33	1.22	
LAN 03	0.45	0.22	0.33	
WLAN 03	0.24	0.05	0.91	
Pers Int 03	0.12	3.87	0.07	c

Dataset from the companies that acquired broadband between 2004-2005

Equation 1 with dependent variable: Delta-speed 0405

	Coefficient	Coefficient*	Std-error	
Under 10 04	0.10	0.03	0.63	
Over 250 04	-0.53	-0.14	1.04	
Labour Quality 04	0.00	0.11	0.11	
Intranet 04	-0.61	-0.27	0.60	
Extranet 04	-0.15	-0.04	0.94	
LAN 04	1.97	0.98	0.43	a
WLAN 04	2.79	0.77	1.23	b
Pers Int 04	0.61	19.24	0.78	

Results from the equation with delta productivity as a dependent variable and speed as an explanatory variable (instead of IT-level). Delta productivity is the difference in productivity (between the companies that acquired broadband and the companies that did not acquire broadband) as a dependent variable and speed as an explanatory variable.

Dataset from companies that acquired broadband between 2002-2003

Dependent variable: Delta-productivity 0204

	Coefficient	Coefficient*	Std-Error	
Speed 03	-0.06	-0.12	0.04	
Under 10 02	0.02	0.06	0.08	
Over 250 02	0.32	1.92	0.09	c
Labour Quality 02	0.00	0.00	0.00	
Intranet 02	0.07	0.17	0.05	
Extranet 02	0.08	0.33	0.08	
Lan 02	0.18	0.35	0.06	a
WLAN 02	0.04	0.17	0.13	
Pers int 02	0.00	0.00	0.00	

Dataset from companies that acquired broadband between 2003-2004

Dependent variable Delta-productivity 0305

	Coefficient	Coefficient*	Std-Error	
Speed 04	0.11	0.23	0.10	
Under 10 03	-0.21	-0.76	0.16	b
Over 250 03	0.77	5.16	0.36	
Labour Quality 03	0.00	0.00	0.00	
Intranet 03	0.06	0.15	0.13	
Extranet 03	0.48	2.33	0.23	b
LAN 03	0.34	0.69	0.10	a
WLAN 03	0.16	0.80	0.25	
Pers int 03	0.00	0.00	0.00	

9. Appendix C

Statistical description of Logistic Regression[6]

Logistic regression (Maddala, 2001 and Pindyck, 1998) is a form of regression that analyzes binomially distributed data.

$Y_j \sim Bin\ (n_j, p_j),\ \ for\ j = 1,\ldots,k$

where n in our case is 1 and k is the number of the companies. The independent variables can be of any type.

When the response variable is an indicator variable it is not appropriate to use a linear regression. This is because the right side in a regression equation represents the real line, whereas Y is a 0-1 variable. Assume that we have **x' =** (x_1, x_2,…x_h) then let p(**x**) = P(Y=1| **x**) be the conditional probability for y. The logistic function that describes the model is given by

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i}.$$

Then the conditional probability is

$$p_i = \frac{1}{1+e^{-(\beta_0+\beta_1 x_{1,i}+\cdots+\beta_k x_{k,i})}}.$$

In a multiple logistic regression model we estimate the coefficients
β' = (β_0 β_1 β_2… β_h) by using the ML(Maximum-Likelihood)-method. The likelihood equation looks like

$$l(\beta) = \prod_j p(x_j)^{yj}\,[1- p(x_j)]^{1-yj}$$

Take the logarithm of the function and derivate the function for the estimation of β.

We estimated the models with the procedure "proc Logistic" in SAS.

[6] Statistic description of Logistic regression is written together with Malin Nilsson

10. Appendix D

Statistical description of 2 sls, 3 sls and SUR [7]

2SLS (Two Stages Least Squares) (Bollen, 1996, Kline 1998, Maddala, 2001)

Our problem is that we have an explanatory variable that is endogenous in the model and this is a violation of the OLS assumptions. Therefore we have to use a method that deals with this problem. In 2SLS the problematic, endogenous, variable is replaced with a new estimated variable. The new variable is estimated with an ordinary least square (OLS) regression on some exogenous variables that are correlated with the problematic variable and uncorrelated with the error term. These new variables are called instrument variables.

From the model of the type:

$$y_i = \beta x_i + \varepsilon_i$$

standard estimation of β by OLS yields:

$$\widehat{\beta}_{\text{OLS}} = \frac{\sum_i x_i y_i}{\sum_i x_i^2} = \frac{\sum_i x_i(x_i\beta + \varepsilon_i)}{\sum_i x_i^2} = \beta + \frac{\sum_i x_i \varepsilon_i}{\sum_i x_i^2}.$$

An underlying assumption is that x_i is uncorrelated with the error term and then the estimation is unbiased for any set of x-values (not all zero). On the other hand when x is correlated with the error term we get biased result.

Estimate x_i by regression on z_i then estimate Y as a regression on the new estimated variable.

The two steps in 2SLS:

Step 1: Estimate the problematic variable with regression on the instrumental variables.

Step 2: Replace the endogenous variable with the new uncorrelated one and estimate the original equation with OLS.

Lagged values can be used as instruments.

[7] ”Statistic description of 2SLS 3SLS and SUR” are written together with Malin Nilsson

SUR (Seemingly Unrelated Regression)

This method, developed by Arnold Zellner, analyzes a system of multiple equations when there are both cross-equation parameter restrictions, correlated error terms and different explanatory variables.

Each equation satisfies the CLRM (classical linear regression model) assumptions, and therefore OLS gives an unbiased and consistent estimation. Since we have a system of equations with correlated error terms, the OLS-estimations may not always be efficient.

The system has the following form:

$y_i = x_i\beta_i + \varepsilon_i$ i = 1...m (In our case m = 2)

Each equation has N observations. From the second step in 2SLS a correlation matrix ($\sum$) is estimated from the residuals.

SUR uses GLS (Generalized least squares) to estimate β.

$$\hat{\beta}_{SUR} = \left(X'V^{-1}X\right)^{-1} X'V^{-1}Y$$

(where $Y' = (y_1 \ldots y_i)$)

Where

$$V(Y) = \Sigma \otimes I_N$$

where $\otimes$ is the Kronecker Product and *V*(*Y*) is an $M \times N$ *by* $M \times N$ matrix.

This matrix will also include non-diagonal values (since we have cross-equation correlation).

This matrix shows how the Kronecker Product works.

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}.$$

If *A* is an *m*-by-*n* matrix and *B* is a *p*-by-*q* matrix, then the Kronecker product $A \otimes B$ is the *mp*-by-*nq* block matrix

3SLS (Three Stage Least Squares)

This is a statistical technique to analyze multiple equations. It is a combination of 2SLS and SUR and it is used when we have endogenous explanatory variables and cross-equation parameter restrictions and correlated error terms.

The three steps in 3SLS

Step 1: Estimate the problematic variable with regression on the instrumental variables.

Step 2: Replace the endogenous variable with the new uncorrelated (with the response variable) one and estimate the equation for y with OLS, (These two steps are the same as 2SLS). Then use the residuals from these equations to estimate the cross-equation correlation matrix.

Step 3: Estimate the equations with help of the cross-equation correlation matrix.

We used the procedure “proc syslin” with 3SLS in SAS.