Mathematical Statistics
Stockholm University

Relationship between Broadband and Internet Use amongst Swedish Companies 2001–2005

Malin Nilsson

Examensarbete 2008:10

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

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

Mathematical Statistics
Stockholm University
Examensarbete **2008:10**,
http://www.math.su.se/matstat

Relationship between Broadband and Internet Use amongst Swedish Companies 2001–2005

Malin Nilsson*

September 2008

Abstract

In this degree project we have studied the relationship between broadband and internet use amongst Swedish companies 2001–2005. It is a part of an ongoing project by the Organisation for Economic Co-operation and Development (OECD) in which Statistics Sweden is participating.

Most companies have broadband today and it is of interest to see how that has affected their internet use. With this paper we try to bring clarity in what comes first; is it the use of internet that makes the companies acquire faster internet connection (broadband) or is it the access to (fast) broadband that allows companies to use internet more widely.

When trying to explain this relationship we used different models that required different methods. We used the methods 2SLS, that deals with the problem of having more than one endogenous variable in the model, and logistic regression, that is used when the response variable is an indicator variable.

The results of this study show that the effect goes both ways. But one effect appears to be stronger, that is that broadband has a stronger impact on the internet use. All the results in this study are controlled for company size, to which type of industry the companies belong, and if they are multinational or not.

Keywords: Broadband, Internet use, Logistic regression, 2SLS.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: malin237@hotmail.com Supervisor: Rolf Sundberg. rolfs@math.su.se.

Acknowledgement

I would like to thank Hans-Olof Hagén and Carolina Ahlstrand at Statistic Sweden for providing me with data and valuable information. I would also like to thank my supervisor at Stockholm University, Rolf Sundberg for having lots of ideas and for valuable discussions. Last but not least I would like to thank Jennie Glantz for working together with me on some pieces in this study and for all thoughts and ideas.

Stockholm - 2008

Malin Nilsson

Contents

1. Introduction

Broadband and internet are more important today than they were a few years ago. Most companies in Sweden use the internet every day, for business. In this paper we try to see how the use of internet and broadband has affected the companies in Sweden. Many studies of companies have been made concerning innovation and how it affects productivity, but not much concerning different levels of internet use and the frequency of broadband. One example is a study by Hagén et al. (2007), on how internet use affects innovation and therefore also productivity. In this paper however we are not interested in that effect, but we want to measure how the use of internet and the acquisition of broadband affect each other. A parallel study (Glantz,J. 2008) has been made on the effect of broadband on productivity.

With this paper we try to bring clarity in what comes first; is it the use of internet that makes the companies acquire faster internet connection (broadband) or is it the access to (fast) broadband that allows companies to use internet more widely. Probably it is a combination of both, and it would be interesting to get some measure of the level of impact they have on each other.

The questions we try to answer in this paper are:

- What kind of company acquires/has broadband?
- Is it the high level of internet use that makes a company acquire/have broadband?
- Is it the access to broadband that increases the company's internet use?

To answer these questions we have economic data registered from all companies in Sweden and data from a survey made by Statistics Sweden. The survey was made during 2001-2005 amongst companies in Sweden concerning their internet use and internet connection.

This paper will start with an introduction in section 1 and a description of the data in section 2. Models and methods will be explained in section 3, with a short description of the procedures 2SLS and 3SLS. Section 4 gives a presentation of the results and in section 5 we have our discussion. All variables are listed in appendix A, while appendix B contains a more detailed description of the methods 2SLS and 3SLS. Appendix C contains a description of logistic regression, and in appendix D there are some tables and diagrams.

2. Our data

We have two types of data, registered economic data and data from the survey. The data are *panel data* or time-series cross-sectional data (TSCS), it is a combination of multiple subjects and how they change over time. We use this to examine changes in variables over time.

The registered economic data included all kind of information about the companies. We divided the companies in three size categories: small, middle and large, dependent of the number of employees. This resulted in two indicator variables, under10 and over250, that represent companies with < 10 employees or > 250. We also created dummy variables for which industry the company belonged to and if it was a multinational business. A measure for the productivity was calculated as a function of economic variables. For a precise description of the equation, see appendix A.

The variables from economic data:	
anst_JE	Number of employees
Labour quality	A measure of the quality of employees
University level	The share of employees with a university education > 3 years

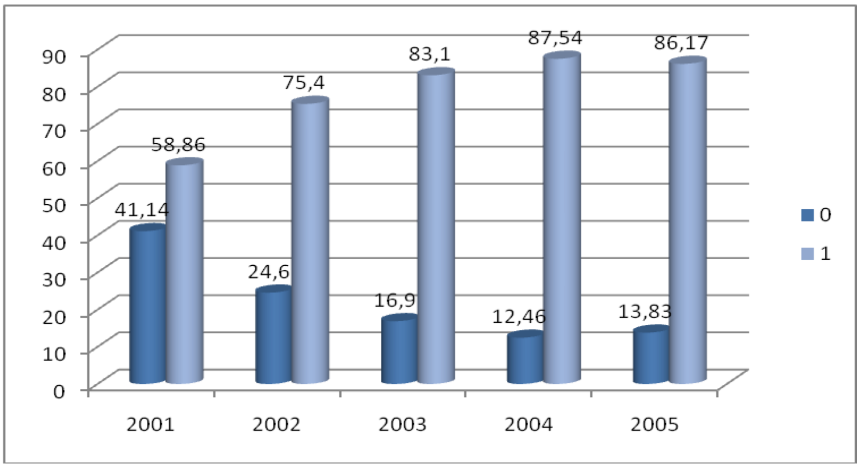
Calculated variables from economic data:	
Industry	The industry, 11 dummies 0/1
Multinational	If the company is part of a concern, 3 dummy 0/1
Under10	< 10 employees, dummy 0/1
Over250	> 250 employees, dummy 0/1
LnGPMFP	A productivity measure (*gross production multifactor productivity*)

The survey from 2001-2005 includes all large companies (over 250 employees) in Sweden and a sample of the smaller ones. Every year one third of the smaller companies were replaced. So over time most of the large companies are represented, but only a small portion of the smaller companies. To obtain a measure of the internet level a new variable was constructed as an arithmetic average of different areas of internet use; business activities, other internet activities, sales and purchase on-line. This definition of the variable IT-Level has been used before (Hagén et al, 2007) and we use it again for consistency. The year 2001 the variables business activities and purchase on-line were missing in the survey and we replaced them with the mean, over all companies, of those variables from 2002. From the survey we had information about the company's level of internet connection, but we concentrated on whether they had access to broadband or not. Therefore we constructed an indicator variable, speed, which is 1 if the company have broadband and 0 if they do not have broadband. The percentage of companies that have broadband for different years is presented in diagram 1.

Variables from the survey:	
Intranet	If the companies have Intranet, 0/1 dummy
Extranet	If the companies have Extranet, 0/1 dummy
LAN	If the companies have Local Area Network, 0/1 dummy
WLAN	If the companies have Wireless Local Area Network, 0/1 dummy
PersInt	Share of employees with internet connection

Calculated variables from the survey:	
Speed	If the company have broadband (1) or not, dummy 0/1
ITLevel	The level of internet use

Diagram 1 The percentage of companies that have broadband (1) and those that do not have broadband (0) for different years



After doing this and removing some outliers we constructed the total data set that consists of data from the survey and from the economic register. The total data set consists of approximately 3000 companies for each year, see diagram 2, of which 595 are represented every year. Over a two year period the data set consists of approximately 1900 companies, see diagram 3.

Diagram 2 The number of companies every year

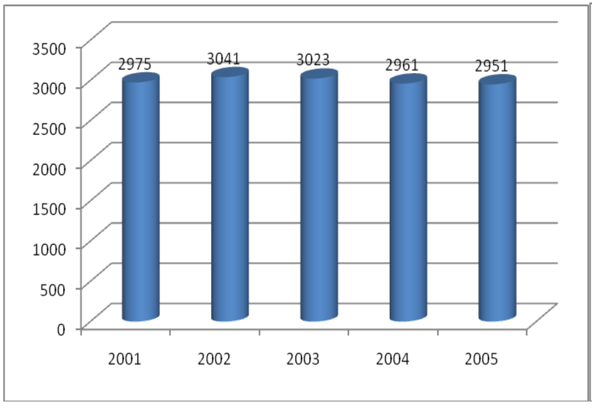
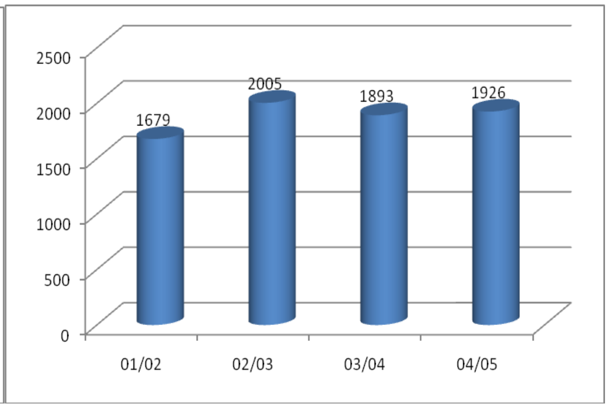
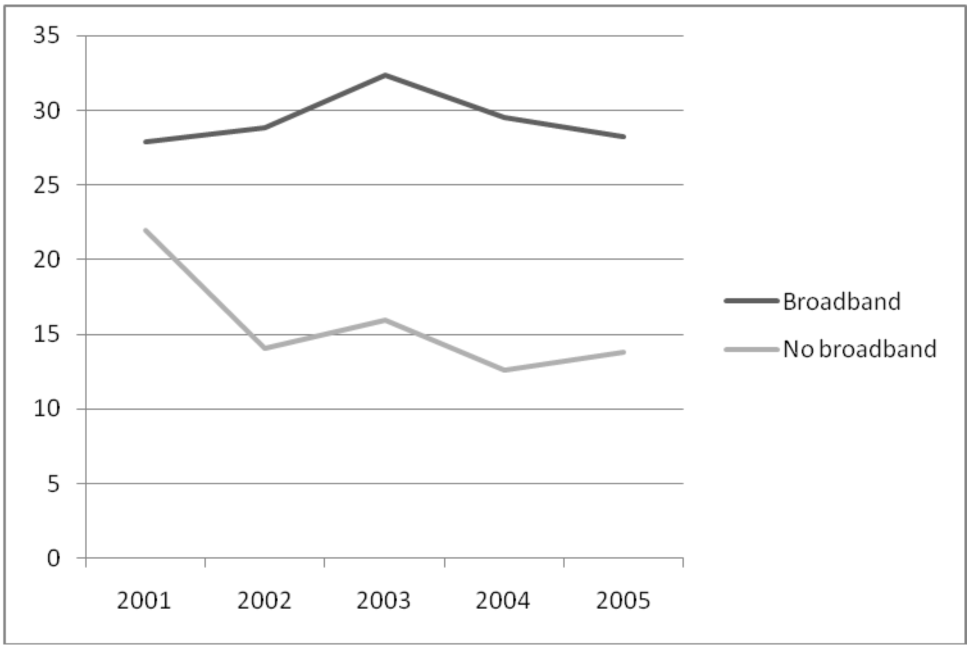


Diagram 3 The number of companies represented over a two year period



As shown in diagram 4 the ITLevel is higher for companies with broadband compared with those that do not have broadband.

Diagram 4 The mean of ITLevel for every year divided in two categories; if the company have broadband or not. ***(Note: Year 2001 is not directly comparable with the other years.)***



If we look at the spread of the ITLevel for the different years we can see that it is smallest for the first year and then ITLevel is less stable, see diagram 5 and 6. It is natural that the ITLevel for the first year (2001) does not vary that much because of the way it is constructed, with the average of some variables from the next year (see description above).

Diagram 5 The spread of ITLevel for companies that do not have broadband

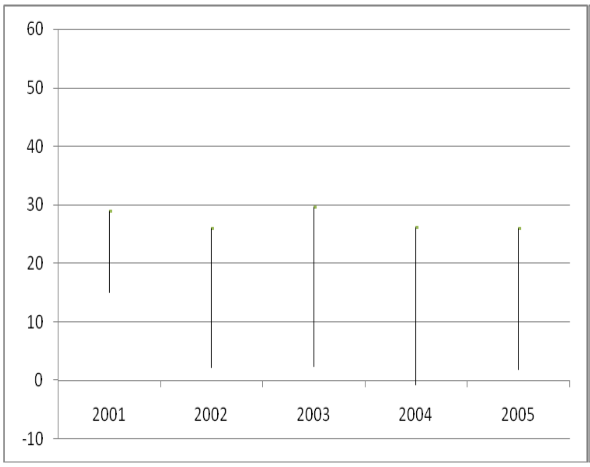
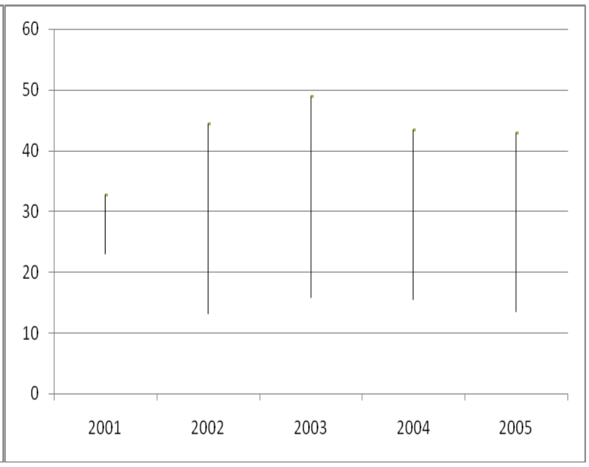


Diagram 6 The spread of ITLevel for companies that have broadband



3. Model and methods

3.1. Equation 1

First we try to determine what makes companies invest in broadband and what variables that affect companies that already have broadband. Because the variable speed is an indicator variable that only can assume the values 0 and 1 whereas some of the explanatory variables are continuous, we use logistic regression. If we use a logistic regression instead of a regular regression the left side of the equation is also continuous. (For more about logistic regression, see appendix C).

The equation is of the type:

1. $\log[p/(1-p)] = \alpha + \boldsymbol{\beta}\,\mathbf{x}$

Here p is the probability that the dependent variable, speed, takes the value 1 and α is the intercept.
x is a matrix of different variables that might affect the dependent variable: dummies for type of industry, if the company is multinational, if it is a small, medium or large company, a measure of the staff efficiency and if the companies have intranet, extranet, LAN or WLAN, the percentage of employees with access to internet, and a measure of the productivity (LnGPMFP, logarithm of gross production multifactor productivity). For more exact description of the variables, see appendix A.

3.2. Equation 2 and 3

Equation 2 has IT-Level as the dependent variable and the internet connection variable (speed) as explanatory. Because the dependent variable can be expected to depend on a number of other variables we have to take that into account. To do this we included in the model if the company was a multinational business and which industry type it belonged to. When including these variables we allow for the possibility that different industries have different conjunctures and different levels of usage of for example the internet. The size of the company probably affects their internet use and therefore we also take that into account.

Equation 3 is constructed in the same way but here speed is the dependent variable, and IT-Level the explanatory one. This equation will be estimated with logistic regression.

The two equations:

2. $Y = \alpha_0 + \beta_0 z + \boldsymbol{\beta}_1 \mathbf{x} + \varepsilon$

   (ITLevel = intercept + β_0Speed + $\boldsymbol{\beta}_1$other variables)

3. $\log[p/(1-p)] = \alpha_1 + \beta_2 y + \boldsymbol{\beta}_3 \mathbf{x}$

   (log odds ratio for speed = intercept + β_2 ITLevel + $\boldsymbol{\beta}_3$other variables)

Here Y is IT-Level, z is speed, p is the probability that the variable speed will be 1 and α is the intercepts.
x is the same matrix as for equation 1 above.

We try to see which variable that has the largest effect on the other one, the broadband connection or the high level of internet use, when doing this we use different methods.

First we wanted to get some measure on the level of impact one of the variable have on the other by using all variables from the same year. This requires a complicated model because more than one variable in the equations is endogenous, which means that the explanatory variable is more or less caused by the response variable. This is because the variables speed and IT-Level correlate and the influence might go in both directions. Because of this the explanatory variable is considered to be endogenous, it is not entirely explained outside the model. If we do not use a method that takes this into account it could cause inconsistent results. The procedure 2SLS makes the results more correct and we use it to avoid for example simultaneous bias. (Read more about 2SLS in section 3.3. below)

When we used lagged values of the explanatory variable to explain the dependent variable we do not have the problem with too many endogenous variables and we can use a simpler method instead. In equation 2, when ITLevel is the dependent variable, we use a multiple regression and in equation 3, when Speed is the dependent variable, we use a logistic regression (for the same reasons as in equation 1).

3.3. 2SLS (Two Stage Least Squares)

If there is more than one endogenous variable in a regression the result will not be consistent, but by using a method that takes this into account we get better results.

2SLS is a variation on multiple regression that gets around the problem of model-implied correlations between disturbances and the cause of endogenous variables. More precisely suppose we have an explanatory variable (x) in the model that is endogenous and correlated with the disturbance term of the endogenous, dependent, variable (y) on which it has a direct effect. This is a violation of the assumptions of recursivity in OLS (Ordinary Least Squares) regression. In 2SLS the problematic variable (x) is replaced by a new variable, this is the first of the two steps. The new variable is constructed by regression on exogenous variables, called instrument variables, that has no direct effect on the dependent variable (y). The new variable (z) is uncorrelated with the disturbance of the endogenous variable and we can use this instead of the problematic one. This is because it is required of the instrumental variables that they are uncorrelated with the disturbance of endogenous variable (y), and therefore this will hold also for the new variable (z). In the second step of the 2sls procedure the endogenous variable (y) is regressed on the new predicted variable (z) and other exogenous variables, if we have a multiple regression. See appendix B for more information about the models.

3.4. 3SLS (Three Stage Least Squares)

This technique to analyze multivariate data combines 2SLS with SUR (Seemingly unrelated regression). SUR is a model developed by Arnold Zellner, (Zellner, A. 1962) for analyzing a system of multiple equations when the error terms are correlated and with cross-equation parameter restrictions. In the present case, however, we do not have multiple equations, so there cannot be any correlated error terms or other cross-equation parameters. We also have dependent variables that are indicator variables and we then use logistic regression model. Hence 3SLS is not the technique we use here. See appendix B for more information about the models.

4. Results

Variable definitions	
Under10	The number of employee under 10, 0/1 dummy
Over250	The number of employee over 250, 0/1 dummy
Labourquality	A measure of the quality of employees
Intranet	If the companies have Intranet, 0/1 dummy
Extranet	If the companies have Extranet, 0/1 dummy
LAN	If the companies have Local Area Network, 0/1 dummy
WLAN	If the companies have Wireless Local Area Network, 0/1 dummy
PersInt	Share of employees with internet connection
LnGPMFP	Logarithm of gross production multifactor productivity
LnGPLP	Logarithm of gross production labour productivity
ITLevel	lTLevel of internet usage
University level	If the employees studied more than 3 years, 0/1 dummy
Speed	If the companies have broadband, 0/1 dummy
dspeed	If the company acquired broadband, 0/1 dummy

For more information about the variables se appendix A

4.1. Equation 1[1]

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.

[1] ”Equation 1” is written together with Jennie Glantz

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 D.

Companies 2001-2002

Table 1 Dependent variable: delta-speed 0102

	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	

(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)

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

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.

Table 2 Dependent variable: speed 02

	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

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

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.

4.2. What comes first?

As was expected it is not just one of the response variables, ITLevel and speed, that affects the other. It is an effect that goes both ways, but one of the effects might be larger. The tables below show some of the results from the two equations. The results are from a dataset that consists of companies that are represented in a time period of two years. This is to make the results over a period of time better, because it is the same companies that are compared. Our data reaches from the year 2001 to 2005, and so we made different results for the different two-year periods; 01/02, 02/03, 03/04 and 04/05. We first used the 2SLS procedure, with speed as an endogenous explanatory variable, but this gave no significant results. It could be expected that it takes time, for example one year, before we can see some results. We then tested to have the explanatory variable lagged, which in this case means that we used speed (or ITLevel) for the previous year. But we have to take into account that some of the exogenous variables are most likely to have effect the same year, so all other explanatory variables are from the same year as the dependent variable, in equation 2 and 3. All equations have been controlled for which industry the companies belong to and if they are a multinational company.

Although some variables have a significant effect on the response variable we are most interested in how ITLevel and Speed affect each other. A hint is to look at the p-value and see if the effect is significant. Equation 2 is a common multiple regression, and the results for the variable Speed is listed in table 3 below. The variable speed (lagged) seems to have a significant effect on the company's ITLevel the next year.

Table 3 Equation 2: Dependent variable ITLevel (one year later)

	Coefficient	Coefficient*	Std Error	
Speed 01	4,57	2,20	0,93	**a**
Speed 02	2,91	1,22	0,90	**a**
Speed 03	2,99	1,05	0,90	**a**
Speed 04	3,39	1,04	1,17	**a**

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

* = Coefficients for variance-standardized explanatory variables

The equations also include the variables **x** from the same year as ITLevel (see section 4.1.).

The result from equation 3, for all years, is presented in table 4 below. It is difficult to compare these results with the results above because we have a logistic regression model here. As we can see in table 4 some years show a significant positive effect on the dependent variable, but the result is not as strong as in table 3 above.

Table 4 Equation 3: Dependent variable Speed (one year later)

	Coefficient	Coefficient*	Std Error	
ITLevel 01	0,02	0,14	0,01	**c**
ITLevel 02	0,03	0,47	0,01	**a**
ITLevel 03	0,01	0,13	0,01	
ITLevel 04	0,05	0,82	0,01	**a**

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

* = Coefficients for variance-standardized explanatory variables

The equations also include the variables **x** from the same year as speed (see section 4.1.).

In these equations the variables industry and multinational have an impact. If a company has broadband and their level of internet use depends on what kind of company it is. However because we are not interested in the exact result of the variables industry or multinational (or the other variables **x**, see section 4.1 for detailed description) we do not present the coefficients here but can say that they appear to have some effect.

We also tried to see what variables that affect the decision to acquire broadband or what affects the increase (or decrease) in ITLevel, and the results will be listed in tables 5 and 6 below. Because we only wanted to compare the companies that acquire broadband with those that do not have broadband at all, we constructed a dataset consisting only of the companies that started the time period without broadband. When doing this we got a much smaller dataset and therefore the results are more difficult to draw conclusions from. The number of companies that acquire broadband is less and less for every year, almost all companies in the end of the survey had broadband.

Table 5 Equation 2: Dependent variable dITLevel

	Coefficient	Coefficient*	Std Error	
dspeed 0102	3,94	1,96	0,87	**a**
dspeed 0203	2,65	1,33	1,66	
dspeed 0304	2,29	1,14	1,44	**b**
dspeed 0405	4,86	2,43	2,07	**b**

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

* = Coefficients for variance-standardized explanatory variables

Example of the equation listed in table 5: dITLevel0405 = dspeed0405 + **x**04 + ε

We try to explain the changes in ITLevel between the years 2004-2005, by using the variable dspeed05 (= 1 if the company acquire broadband between the years 2004-2005) and other variables **x** (see section 4.1.) from the "start" year (2004), the dataset is restricted to companies not having broadband the first of the two years.

As we can see in table 5 we have the strongest result in the first year.

When we tried to see how the ITLevel affects whether a company acquires broadband or not, we got the following result (see table 6). The result is not as strong as in the equation above and the only significant result is for the years 2001/2002.

Table 6 Equation 3: Dependent variable dspeed

	Coefficient	Coefficient*	Std Error	
ITLevel 01	0,03	0,22	0,01	**b**
ITLevel 02	-0,01	-0,12	0,01	
ITLevel 03	0,01	0,14	0,01	
ITLeve 04	0,01	0,09	0,02	

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

* = Coefficients for variance-standardized explanatory variables

Example of the equation listed in table 6: dspeed0405 = ITLevel04 + **x**04 + ε

We try to explain the changes in speed (here if the company acquire broadband) between the years 2004-2005, by using the variable ITLevel04 and other variables **x** (see section 4.1.) from the "start" year, 2004, the dataset is restricted to companies not having broadband the first of the two years.

5. Conclusions

In 2005 almost 90% of the companies in the survey had broadband and probably even more today. Even though most companies in Sweden nowadays have broadband, it is of interest to see what variables affected the company's decision to acquire it. It is also of interest to see what characterizes a company with broadband. As a result of the first equation we can see that large companies have a larger impact on obtaining broadband than medium sized companies. When it comes to small companies we did not find any significant difference from the impact of medium sized companies (that was our control group). This could be because we only had a small selection of the minor companies. We tried to see how the variable Labor quality affected the decision to acquire broadband, but even though it gave some significant results it was difficult to interpret because it is a variable that we think have both increasing and decreasing effect. What also might affect the company's decision to acquire broadband is if they are multinational or if they only appear on the local market. Although we are not interested (in this study) in the level of impact, we can see that it some years have a significant effect. Which industry the companies belong to also have an impact on the decision to get broadband, but again this study is not for determining the level of impact different industries have. We had the variables in the equations because we wanted to control for them. We expect that different types of industries have different usage of the internet and therefore some types of industries are more likely to acquire broadband than others. We could see that some types of industries gave a significant result and some did not.

From the second and the third equation we can see that it is, as expected, an effect that goes both ways. We see from the results in section 4.2 that speed has a significant effect on ITLevel every year (table 3). To have a large ITLevel the companies most likely have a fast internet connection, this is not a surprising result. If we look the other way around we see that the result is not as significant, for all the years, as it was when ITLevel was the dependent variable. This is shown in table 4.

We were also interested in how the ITLevel affected the decision to acquire broadband and if the acquiring of broadband caused a high increase in the ITLevel. But the data material was small, when we tried to test changes over time, and therefore the result cannot be expected to be so significant. We could see that the acquisition of broadband (speed) had a stronger impact on the changes in ITLevel, the result is more significant, than if we look the other way around. This is shown in tables 5 and 6. That is if a company increases its internet connection and gets broadband it is likely to increase its internet use by more than if they did not acquire broadband. The results that we have give us a direction and we can get some idea of which “comes first”, it is more likely that the variable speed is “first” and then the variable ITLevel follows. This is also what could be expected and it feels like a natural way to go. But it should be noted that it is not a one way street.

References

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

Glantz, J, (2008) *Relationship between Broadband and Productivity amongst Swedish Companies 2001-2005,* Bachelor Thesis for Statistics Sweden and Stockholm University.

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

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

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

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 Estimation*, 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.

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

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

Johnston, J. & DiNardo, J, (1997) *Econometric Methods*, McGraw-Hill, pp.

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

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.

Appendix A

Description of the variables

Here is a presentation of the variables that are included in the models.

Economic variables:

Industry	The industry, 11 dummies 0/1
Multinational	If the company have concern in Sweden, USA or other country, 3 dummies 0/1
PeOrgNr	The organization number
anst_JE	Number of employees
Under10	If the number of employees is under 10, 0/1 dummy
Over250	If the number of employees is over 250, 0/1 dummy
Labour quality	A measure of the labor quality
andelunivutb3	The share of employees with a university education > 3 years

Variables only used in calculations:

prod_fp	*The production, inflation secured*
forbruk_fp	*What the companies consume, inflation secured*
va_fp	*Value added = prod-forbruk, inflation secured*
Kapital_FP	*Capital, inflation secured*
Expint_FP	*Export intensity, inflation secured*

Calculated variables:

Intermediateshare	= consumtion share = forbruk/prod
Wageshareprod	= - 0.001*approximate wage/prod
WageshareVA	= - 0.001* approximate wage /value added

GPMFP	*gross production multifactor productivity* =prod_fp - [median (Intermed).*(forbr_fp)] - [median (wageshareprod).*(Labour quality)] - [(1-median (Intermed).-median (wageshareprod.))*(Kapital_fp)]

Variables from the survey:

Speed	If the company have broadband, 0/1 dummy
Intranet	The access to Intranet, 0/1 dummy
Extranet	The access to Extranet, 0/1 dummy
LAN	The access to Local Area Network, 0/1 dummy
WLAN	The access to Wireless Local Area Network, 0/1 dummy
PersInt	The share of employees with access to internet

Variables only used in calculations:

E-sales	*% of total sale over internet*
E-purchase	*% of total purchase over internet*
Business activities	*A measure of how much the company uses internet for business*
Internet activities	*A measure of how much the company uses internet for other activities*

Calculated variables:

ITlevel	A weighted sum of internet activities, business activities, e-purchase and e- sales. (E-sales + E-purchase + Internet activities + Business activities) /4

Appendix B[2]

Statistical description of 2SLS, 3SLS and SUR

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:

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

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.

[2] This part is written together with Jennie Glantz

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.

Appendix C

Statistical description of Logistic Regression[3]

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' =** $(x_1, x_2, \ldots x_h)$ then let p(**x**) = P(Y=1| **x**) be the conditional probability for y. The logistic function that describes the model is given by

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

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' = $(\beta_0\ \beta_1\ \beta_2 \ldots \beta_h)$ 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.

[3] Statistic description of Logistic regression is written together with Jennie Glantz

Appendix D

Results from equation 1 for the years 2001-2005

* = Coefficients for variance-standardized explanatory variables

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

Table 1 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**

Table 2 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**

Table 3 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**

Table 4 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**

Table 5 Dependent variable: delta-speed0102

	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	

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

Table 6 Dependent variable: delta-speed0203

	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**

Table 7 Dependent variable: delta-speed0304

	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**

Table 8 Dependent variable: delta-speed0405

	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	

Boxplot – over companies that started the period without broadband.

Period 2001-2002

In diagram D.1 we can see the IT-level 2001 for companies that will acquire broadband the next year (speed02 =1) and companies that will not acquire broadband (speed02=0)

In diagram D.2 we can see the ITLevel 2002 when some companies has acquired broadband (speed02=1)

Diagram D.1

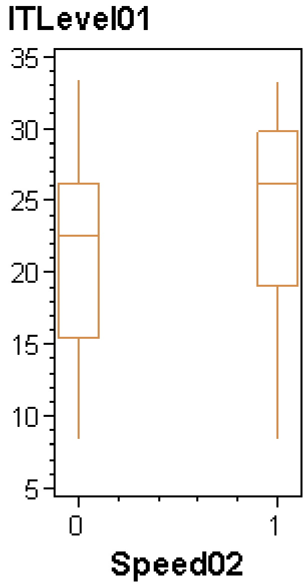
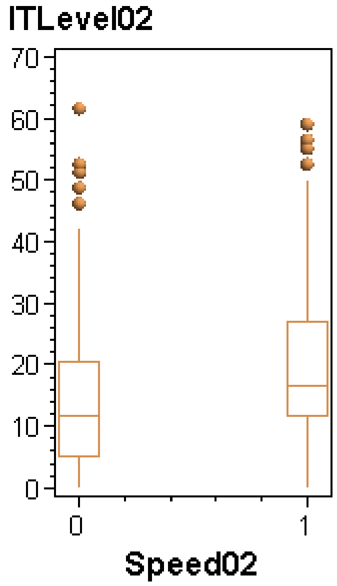


Diagram D.2



Period 2002-2003

Diagram D.3 and D.4 show same as D.1 and D.2 above but for the time period 2002-2003.

Diagram D.3

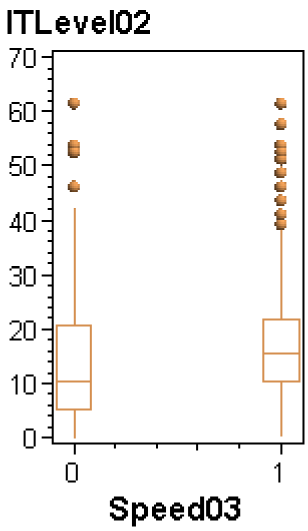


Diagram D.4