LUND UNIVERSITY
School of Economics and Management

Business or Pleasure:

Broadband and Employment in Swedish Municipalities 2007-2011

Author: Erik Grenestam

Supervisor: Martin Nordin, Ph.D

NEKN01 Master Essay I, 15 ECTS

Seminar: 2013-09-03

Abstract

This thesis examines the relationship between expanding coverage of fiber optic broadband and employment in Swedish municipalities during 2007-2011 using a fixed effects panel and an instrumental variables model. Increased fiber coverage among households is estimated to have a negative effect on municipal employment whereas increased coverage among workplaces is weakly positively related to employment.

The estimation strategy relies on the differences in broadband deployment both within and between municipalities. Data on fiber coverage is taken from the yearly surveys conducted by the Swedish telecommunications agency, PTS. To mitigate issues of endogenity, topographical variation is used as an instrument for fiber coverage. The instrument is obtained by drawing a custom sample of public data on elevation and calculation sample standard deviation and variance. Common tests suggest that it is too weak to provide robust results, a shortcoming which may be related to low correlation with the endogenous variable as well as a small sample.

Contents

1 Introduction

High speed internet access is becoming as ubiquitous a utility as heating or running water in many developed countries. But so far there have been few studies examining the effects of this information highway connecting every apartment and office to a global community. The economic effects, though often cited as immensely positive by policy makers (Näringsdepartementet[1] 2009, 2011), can in many cases not be substantiated. While there is consensus regarding the positive impact of IT in general on growth and productivity (Jorgensen, Ho & Stiroh, 2008) there are few studies on individual aspects of IT development, such as widely available broadband (Crandall, Lehr & Litan, 2007). As internet infrastructure is still very much under development, there is a need for frequent updates in research methodology and data gathering to identify economic effects. This thesis will use a panel data regression analysis complemented by an instrumental variable model to determine the effects on local employment rates associated with developing fiber coverage. The main question that we will attempt to answer is: *How are municipal employment rates affected by increased availability of internet access via optic fiber?* Our findings suggest that there are multiple effects at work, increased fiber coverage among workplaces might have a positive effect whereas increased coverage among households appears to be negatively associated with employment. Such a negative effect has not been found in previous research, but its existence is not unlikely. We argue that the consumer applications facilitated by fiber are geared towards entertainment, giving rise an increased demand for leisure over labor.

Regarding IT, in 1987 renowned growth economist Robert Solow famously stated that *"you can see the computer age everywhere but in the productivity statistics".* In the late 90's, this dismal Solow paradox was replaced with an almost euphoric optimism surrounding information technology. Some said we had entered a new economy characterized by very high productivity growth. Growth accounting models told of a significant increase in American labor productivity, a large part of which was attributed to investments in IT (Jorgenson & Stiroh, 2000). However, as IT saturated even the less information intensive sectors, information technology diminished as a driving factor behind productivity increases (Jorgenson et al, 2008).

There is a small but growing body of research concerning the effects of internet diffusion. As with any new infrastructure, there are many aspects which call for attention.

[1] Swedish Ministry of Enterprise, Energy and Communications

Greenstein & McDevitt (2011) estimate the additional consumer surplus generated by broadband compared to a benchmark case of continued use of dial-up modems. They find a significant amount of added surplus for American consumers. Noh & Yoo (2007) examines how internet relates to growth and equality using a panel of 60 countries. Their findings suggest that among countries with a high income inequality, the digital divide between rich and poor seem to hinder the potentially positive effects of internet adoption. The digital divide is also highlighted by Forman, Goldfarb and Greenstein (2009) who examines how the diffusion of internet access during 1995-2000 affected regional wage distribution in the U.S. While they found that internet investment is related to increased wages, this benefit is mostly limited to highly skilled workers in urban areas.

The study conducted by Crandall et al (2007) forms a prototype to our regression analysis. Their cross-sectional study of broadband, output and employment in 48 American states conclude that there is a positive relationship between employment and broadband deployment in several sectors of the labor market. However, they do not make any attempt to disentangle issues of endogeneity. The comprehensive approach by Kolko (2012) provides the methodological foundation for this thesis. Armed with a substantial American panel (sourced from the Federal Communications Commission, FCC) on the number of broadband providers in an area, Kolko (2012) estimates a fixed effects model as well as an IV model with the slope of the local terrain as an instrument for changes in the number of providers. In line with Crandall et al (2007), his findings point to a positive causal link between broadband and employment.

Kim and Orazem (2012) acknowledges Kolko's (2012) approach, but claim that his results are highly sensitive to changes in specification as well as unobservable variables. Their own approach is based on the positive effects of broadband on firm productivity, making areas with broadband more desirable when a new firm decides on where to locate. Controlling for other location specific characteristics, they find that broadband has a positive effect on the number of new establishments. They do not find evidence that the broadband effect differs across industries.

Sweden is at the forefront when it comes to the optical fiber-based infrastructure required for the latest generation of broadband technology (OECD, 2012). Since 2007, The Swedish telecommunications authority (Post- och telestyrelsen, PTS) have conducted yearly municipality-level surveys of the availability of various forms of internet access. These surveys form the core of our data set. Many studies, such as the aforementioned by Kolko (2012) use so called “Form 77”-data supplied by the FCC. In comparison, the Swedish PTS surveys provide an incredible amount of detail. The most interesting feature of our dataset is that coverage among workplaces is separated from coverage among households, allowing us to jointly estimate two separate – and potentially very different – effects of increased fiber coverage on employment.

Today, a number of access technologies are available in the Swedish marketplace. Rather than looking at archaic copper-based technology, we have chosen to focus on the kind of high-speed broadband made possible only by fiber optic cable. The relative novelty of this technology means that there is a healthy amount of heterogeneity exhibited in the annual surveys, with significant coverage growth in most municipalities across our narrow time frame (2007-2011). By exploiting the differences in broadband availability within municipalities over time, we can identify the change in employment rate associated with increased broadband availability. As a robustness check, we identify two subsamples based on municipal population and re-estimate our fixed effects panel data model.

To further qualify our results, we build on Kolko’s (2012) methodology, using topographical variation as an instrument to examine any causal link running from increased broadband availability to changes in employment rates. Topographical variation is an intuitively appealing instrument as the increased costs associated with extending coverage in areas with mountainous terrain is likely to result in lower levels of coverage in these areas. The municipal terrain is obviously unaffected by short term changes in the employment rate, but to qualify as an ideal instrument, terrain should not have a direct influence on employment. While we can’t argue that this is the case, specifying a model in differences and introducing the right set of controls can mitigate these issues to some extent. We measure topographical variation as the standard deviation and variance of a sample elevation profile constructed for each municipality using open source data. A detailed description is given in section 3.

The paper is laid out as follows. Section 2 provides a brief introduction to broadband technology with a focus on the Swedish market for internet access. Section 3 describes the data used in the study. Section 4 provides brief explanations of the fixed effects panel data model, the instrumental variables model and the two-stage least squares (TSLS) estimator as well as the empirical considerations pertaining to our study. Section 5 details the modelling approach and presents results. Sections 6 and 7 provides a brief discussion and our concluding remarks, respectively.

2 Background

2.1 Benefits of a broader band

The word broadband has, in a way, become redundant. As Moore's law[2] keeps fulfilling its promise of steep technological progress, internet applications demand more bandwidth and lower latency times by the month. What is today considered a bare minimum speed was viewed as viewed as blazingly fast only a few years ago. As more of essential services, entertainment and education move to an online platform, not having a speedy and reliant internet connection is slowly becoming equivalent to being excluded from aspects of society.

After the era of dial-up modems came the first generation of publicly available broadband technology. Access often relied on existing telephone lines to make the connection between a service provider's node and the customer. A signal travelling by copper is subject to severe degradation even short distances from origin, so available bandwidth was limited by the distance to the closest node. The signal is also sensitive to interference between data and telephone traffic and less reliable than modern technology based on fiber optic cable (PTS, 2007). Today, in most developed countries, the commonly available way to obtain a high speed[3] connection is by optic fiber. Transferring data as pulses of light through optically pure glass enable bandwidths between five to 200 times greater than those of copper-based technology. This modern access technology, for here on referred to simply as 'fiber', is the focus of this study.

With increasing bandwidth comes a new set of internet applications aimed at consumers, corporations and the public sector. High-traffic server applications, high-definition video streaming, video conferencing, telemedicine and real-time backups are some examples of applications made viable by optic fiber. In addition, the increased reliability of fiber is in itself essential many businesses. Organizations or large households where multiple persons simultaneously use the same connection are likely to experience an across-the-board quality increase using most common internet applications.

---

[2] In 1965, Gordon E. Moore proclaimed that the number of components fitted unto an integrated circuit would double each year for the foreseeable future. This law of rapid advancement has been successful in predicting advances in many areas of IT, such as internet access.

[3] Popularly defined as a downlink bandwidth of at least 100 megabits per second.

Does the increased number of internet applications have an effect on economic fundamentals? That is what this study aims to find out. The fast-paced IT development over the last couple of decades is often described as a series of sudden leaps ahead. However, while a switch from copper to fiber has an over-night effect on bandwidth, the applications and user experience changes only gradually. Therefore, changes in broadband availability can be viewed as a proxy to an increased supply of and demand for internet applications. These application may be new innovations (e.g. services such as YouTube and LinkedIn) or close substitutes for existing services (e-mail instead of regular mail, Skype instead of phone calls). No matter the nature of the applications, it is their utilization that we expect to have an aggregate effect on economic fundamentals, not the mere existence of fiber-based internet access. Since we cannot credibly observe the use of all available applications, the supply and demand of internet access will have to suffice.

2.2 The Swedish setting

In Sweden, staying competitive in field of the IT realm is on the agenda of policy makers (Näringsdepartementet, 2009). From an infrastructure point-of-view, achieving high broadband coverage using fiber comes at a higher cost than working with existing copper-wiring or radio technology. During 2001-2007, the government spent roughly five billion SEK in subsidies for municipal broadband development, most of which was spent on putting fiber in the ground (PTS, 2007). As of today, rural broadband development is still publicly funded, but to a lesser extent (PTS, 2012).

Municipalities played a major role in the distribution of these grants, and were ultimately in charge of how the funds were spent. Therefore, during 2001-2007, it was common for municipal governments to start broadband enterprises of their own, putting optic cable in existing utility tunnels and selling “dark fiber” (optical fiber cable without any active telecom equipment attached) to operators who in turn supplied broadband services to end-users. As of 2008, local government-owned enterprises and Skanova, a subsidiary of the former telecommunications monopolist TeliaSonera, together supplied over 90 percent of all “dark fiber” (PTS, 2008). The high market share enjoyed by TeliaSonera is likely a result of its history as a government sanctioned monopolist, controlling an overwhelming majority of the copper-based infrastructure came with the advantage of being able to use existing tunnels to replace copper cable with optic fiber at a low cost. In recent years, PTS (2010) has enforced price regulations to facilitate a fair marketplace for consumer broadband services. In spite of regulation, the market for fiber from an end-users perspective does not seem to have changed significantly during our timeframe (2007-2011).

With centralized government funding, a difference-in-differences approach to estimating the effects of increased fiber coverage might seem suitable. If we could identify a subsample of municipalities differing only in the amount of exogenous support they received, we could be able to identify causality. Unfortunately, this does not seem feasible. The government subsidies were designed so that they would not interfere with market forces. Before extending financing, the government required municipalities to credibly identify local areas where the market would likely fail to provide coverage. In practice, this meant that only townships and villages with less than 3000 inhabitants were eligible for subsidized broadband development. In addition, sub-municipality level data on how the grants were spent seems to be unavailable. This mean that even if we could identify subsamples of rural areas, we would still be missing essential data on broadband development in these areas.

3 Data

Most studies of low-level effects of broadband are restricted by data availability. In the American setting, Kolko (2012) as well as Crandall et al (2007) argue that there are no American alternatives to the FCC's so called "Form 77"-data, providing annual data on the number of broadband suppliers within any given U.S. zip code. While rich in observations, the number of suppliers is only a rough proxy to actual availability. Furthermore, there is no way to separate availability among households from availability among workplaces. There is also no data on the quality of the services provided or any indication of actual coverage within a given zip code.

The Swedish telecommunications authority (PTS) provides excellent data on municipal broadband availability. However, detailed surveys only date back to 2007. This provides a bit of a caveat for our purposes. By 2007, practically all municipalities could offer coverage rates of 95 to 100 percent for copper-based connections. Consequently, there is little variation to exploit for this technology. Our choice to study fiber-based access is thus partly due to the data situation.

Using the annual PTS surveys from 2007 to 2011, we construct a panel dataset consisting of municipal data on broadband availability across this period. The broadband availability variable measures the share of municipal households/workplaces that are located within 347 meters from an operational fiber-optic cable in a given year. The surveys are originally performed on a standardized nation-wide grid where each square is 250 by 250 meters. 347 meters represent the diagonal of such a square, i.e. the longest possible distance between a household/workplace and a fiber optic cable located in the same square. A household 300 meters away from a fiber-optic cable will likely incur a significant cost should they want to set up a connection. This means that our measure overestimates the share of households where a high-speed connection is accessible at a low cost. However, since this overestimation is likely to affect all municipalities equally, it will not interfere with our objectives. Later surveys do include a more precise measure, but limiting ourselves to these would come at the cost of a significant loss of observations.

Table 1: Data descriptions

Variable	Description	Mean	Standard deviation	Source
Fiber coverage, households	Share of municipal households with access.	0.3261	0.2482	PTS
Fiber coverage, workplaces	Share of municipal workplaces with access.	0.2675	0.2185	PTS
Employment	Share of adults ages 25 to 64 employed.	0.7989	0.0387	Statistics Sweden
Income (SEK)	Municipal average net income including transfers, adults ages 20 and above.	202786	31597	Statistics Sweden
Population	Number of residents in a municipality.	32064	63141	Statistics Sweden
Population density (2006)	Number of residents per square kilometer.	136	468	Statistics Sweden
Topographical variation	Standard deviation and variance of municipal elevation profile.	45.3616	37.0861	Google Maps, Open Street Map
Road density (2005)	Total length of roads (km) per square kilometer.	2.5539	1.4934	National Database on Roads
Education	Share of adults aged 25-64 with at least three years of tertiary education.	0.1870	0.0791	Statistics Sweden

All variables are observed during 2007-2011 unless otherwise stated.

Data on municipality-level employment are supplied by Statistics Sweden (SCB). The data is taken from a population-wide survey, based on registration for income tax purposes. Our models also include a number of controls. Population, population density and share of adults with at least three years of tertiary education are all provided by Statistics Sweden.

Our instrument is the topographical variation within the geographical borders of each municipality. We measure this variation as the standard deviation and the variance of terrain elevation above sea level. Elevation data is obtained by sampling the terrain along a path through each municipality. Coordinates for the end points of each path was extracted from open source data supplied by Open Street Map using a tailored Python script to access their public API (Application Programming Interface). Having obtained the coordinates for the end points of a path, our script accessed a Google Maps API for elevation data. Using the coordinates as inputs, a 200-point elevation sample was extracted along the path defined by our two sets of coordinates (figure 1). The process was automatically repeated for each of Sweden's 290 municipalities. After the initial run, coastal municipalities had their paths manually adjusted and resampled to reduce bias caused by excessive sampling of lake and sea beds. The data was exported to an Excel-compatible format and the sample standard deviation and variance was calculated for each municipality.

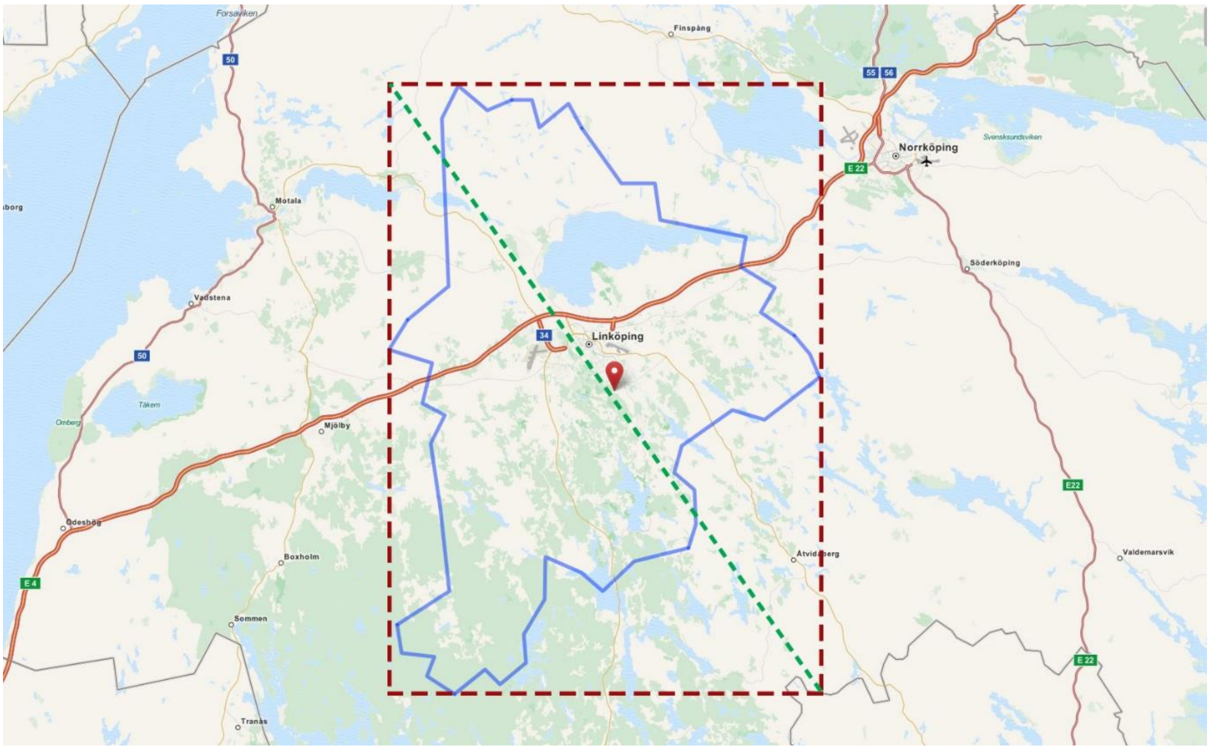

Figure 1: Illustration of municipal elevation profiles. The diagonal dotted line is the sampled path, the sample size is 200.

One of the caveats with our instrument is that topographical variation may have direct as well as indirect effects unrelated to fiber coverage on employment. Again, we follow Kolko's (2012) methodology to control for transportation costs, perhaps the most apparent channel though which terrain can affect employment. As a proxy for transportation costs, we use data on total road length per square kilometer. High transportation costs and a low road density are assumed to be positively correlated, and rough terrain is likely to have a negative impact on both. Data on road length is obtained from a database maintained by the Swedish transportation authority (Trafikverket). The set dates back to 2005, but can be assumed to remain static over the short run. A benefit of using data from an earlier period is that it can be assumed to be exogenous in relation to recent economic outcomes.

4 Empirical specification

By exploiting the differences in absolute levels of fiber coverage as well as coverage growth during 2007-2011, we can identify the relationship between employment and fiber coverage. An interesting feature of our dataset is the possibility to jointly estimate the effects of fiber coverage among households as well as the effect of coverage among workplaces. Here, we hypothesize that the two coverage variables are uniquely relevant, i.e. the effect of coverage among households is different from the effect of coverage among workplaces.

As for our choice of dependent variable, by choosing employment rate we follow Kolko's (2012) approach. This will provide a comparative aspect to our analysis. It is also intuitively appealing to use an outcome of the same type as fiber coverage, a relative share. However, the interpretation of a change in employment is not straight-forward. In addition to a net transfer from the pool of unemployed to the employed, a change could simply be the result of migration, i.e. citizens relocating to another municipality. With this in mind, if we were to focus on variables more closely relating to growth, we would be forced to use average disposable income as a municipal-level proxy. An increase in average disposable income could be the result of any of a number of changes to the local income distribution. We must also keep in mind that basic economic theory dictates a close relationship between employment and growth, at least when firms are considered. Higher productivity implies higher marginal productivity which in turn justifies hiring additional labor.

To find the true effect of fiber coverage on employment rates, we use a two-pronged approach. A fixed effects panel data model will be augmented by an instrumental variable approach. At the least, our IV model will constitute a robustness check to our panel data results. Depending on the quality of our instrument, we may be able to make inferences regarding any causal link between fiber coverage and employment.

4.1 Fixed effects specification

Our baseline model will be a fixed effects panel specified according to:

$$\mathrm{Employment}_{\mathrm{it}} = \alpha_{\mathrm{i}} + \gamma_{\mathrm{t}} + \beta_1 \mathrm{Fiber\ HH}_{\mathrm{it}} + \beta_2 \mathrm{Fiber\ WP}_{\mathrm{it}} + \beta_3 \mathrm{x}_{\mathrm{it}} + \varepsilon_{\mathrm{it}} \qquad (1)$$

Employment is the share of persons ages 25-64 who are employed, *Fiber HH* is the share of households with access to fiber, *Fiber WP* is the share of workplaces with access to fiber and x_{it} is a $\mathrm{K} * 1$ vector of controls. We also include municipality fixed effects, α, and time fixed effects, γ. This will control for all time-invariant heterogeneity across municipalities, as well as capture any general trend in employment across our sample period (2007-2011). We are interested in estimating the effects on employment associated with changes in fiber coverage among households as well as among workplaces, represented by β_1 and β_2 in (1).

As the sample covers just five years, specifying a model in first differences would come at the cost of losing 20 percent of our observations. We also lack any preconceived notion of whether effects are contemporaneous or whether there are lags to consider. Rather than mining for a credible model in first differences, we are content with estimating the effects of a general change in the level of broadband coverage.

A fixed effects specification controls for all unobserved time-invariant heterogeneity across municipalities. However, we still need to control for time varying factors correlated with broadband coverage and employment or our estimates will be inconsistent. Income, education and population are likely to be correlated with both demand for fiber access and employment. Consequently, we include these as controls in our model.

Industry mix, geography, demographics and all other factors assumed to be invariant within our short time frame are some of the characteristics unique to each municipality. While we treat them as static, we cannot treat our observations as independent. Observations on any given municipality are likely to exhibit autocorrelation which will jeopardize our inferences. As described in Verbeek (2012:389), the Newey-West method of estimating the parameter variance-covariance matrix is robust against autocorrelation within municipalities, as well as general forms of heteroscedasticity.

The possibility of weighting observations is briefly explored. On an individual level, fiber access, employment and tertiary education are all binary variables, e.g. a household or workplace either has or does not have access to fiber. Consequently, each observation, i.e. the share of people with access to broadband in a given municipality, can be viewed as an average. By adjusting our observations for the municipal population implicitly represented by these averages, we take into account the fact that effects on a large municipal population provides a greater contribution to the nation-wide effect associated with increased fiber coverage. However, analytical weighting is only a complement to our primary, non-weighted, model as it helps us answer a different question. We are not primarily interested in exploring the aggregate effect of broadband, we are interested in the conditionally expected effect for any given municipality. In contrast to Kolko (2012) who uses data on zip-code level and weight each observation in proportion to the number of employed residents, municipalities are a non-arbitrary sampling unit with respect to employment and fiber coverage. A municipality is a self-governed political entity and, as detailed in Section 2, Swedish municipalities exert a high degree of control with regards to local investments in broadband infrastructure.

While the baseline model allows us to explore the effects associated with increased coverage, it says nothing about the causal relationship between the two. The issues of endogeneity (i.e. violations of the OLS assumption that our independent variables are orthogonal to the error term) must be considered, as any violation of this assumption renders our estimates inconsistent. However, while it is often said that correlation does not imply causality, this simple rule of thumb is no reason not to be hardheaded about what can be concluded from our estimation. Economic literature commonly addresses three ways that problems of endogeneity can arise in a regression model. These are measurement error, omitted variables and reverse causality. We have no reason to suspect systematic measurement error in our variables, and even if we did, we have little recourse to take. Omitted variables can be an issue, here we have to rely on economic intuition to include all relevant controls, but the possibility of an unobservable variable correlated with fiber coverage as well as employment is hard dismiss. Our major obstacle is reverse causality. Employment can be assumed to have a causal effect not only on fiber coverage, but also on average income levels and education. This is the primary reason behind augmenting our fixed effects model with an instrumental variable model.

In a simplified setting, we can make an educated guess of what the bias due to reverse causality could look like. Following an example by Verbeek (2012:146), let us assume that fiber coverage and employment are jointly decided in a system of two equations (individual indices omitted for simplicity):

$$\mathrm{Emp} = \beta_0 + \beta_1 \mathrm{Fiber} + \sum_{n=2}^{N} \beta_n x_n + \varepsilon \qquad (2)$$

$$\mathrm{Fiber} = \alpha_1 \mathrm{Emp} + \sum_{n=2}^{N} \alpha_n x_n + u \qquad (3)$$

Where x_n is an exogenous control, ε and u are both i.i.d. β_1 is estimated in our baseline model (1) and α_1 is assumed to be positive but less than one, i.e. employment is assumed to have a positive effect on fiber coverage, but the partial derivative of fiber coverage w.r.t. employment is assumed to be less than unity. Inserting one equation into the other and solving for employment and fiber coverage yields two new equations:

$$\mathrm{Fiber} = \frac{\alpha_1\beta_0}{1-\beta_1\alpha_1} + \frac{1}{1-\beta_1\alpha_1}\sum_{n=2}^{N}(\alpha_n + \alpha_1\beta_n)x_n + \frac{\sum_{n=2}^{N}\alpha_n}{1-\beta_1\alpha_1}\varepsilon + \frac{1}{1-\beta_1\alpha_1}u \qquad (4)$$

$$\mathrm{Emp} = \frac{\beta_0}{1-\beta_1\alpha_1} + \frac{1}{1-\beta_1\alpha_1}\sum_{n=2}^{N}(\beta_n + \alpha_n\beta_1)x_n + \frac{1}{1-\beta_1\alpha_1}\varepsilon + \frac{\beta_1}{1-\beta_1\alpha_1}u \qquad (5)$$

Given our simple setup, it can be shown that the probability limit (see Verbeek, 2012:147) of the OLS estimate $\hat{\beta}_1$ is:

$$\mathrm{plim}\,\hat{\beta}_1 = \beta_1 + \frac{\mathrm{cov}(\mathrm{Fiber},\varepsilon)}{\sigma^2_{\mathrm{fiber}}} \qquad (6)$$

A simple expression for the probability limit of the effect associated with fiber coverage can be derived if we assume that our controls are independent:

$$\mathrm{Cov}(x_n, \varepsilon) = \mathrm{Cov}(x_n, u) = 0 \text{ for } n = 1,2,3 \ldots$$

$$E(x_m x_n) = 0 \text{ for } n \neq m$$

Given these assumptions, $\mathrm{Cov}(\mathrm{Fiber}, \varepsilon)$ and $\sigma^2_{\mathrm{fiber}}$ are reduced to:

$$\mathrm{Cov}(\mathrm{Fiber}, \varepsilon) = \frac{\sum_{\mathbf{n=2}}^{\mathbf{N}}\alpha_n}{1-\beta_1\alpha_1}\sigma^2_\varepsilon \qquad (7)$$

$$\sigma^2_{\mathrm{fiber}} = \left(\frac{1}{1-\beta_1\alpha_1}\right)^2 \sum_{\mathbf{n=2}}^{\mathbf{N}}(\alpha_n + \alpha_1\beta_n)^2\,\mathrm{Var}(x_n) + \left(\frac{\sum_{\mathbf{n=2}}^{\mathbf{N}}\alpha_n}{1-\beta_1\alpha_1}\right)^2 \sigma^2_\varepsilon + \left(\frac{1}{1-\beta_1\alpha_1}\right)^2 \sigma^2_u \qquad (8)$$

And consequently, the probability limit of β_1 reduces to:

$$\mathrm{plim}\,\hat{\beta}_1 = \beta_1 + \frac{\sum_{\mathbf{n=2}}^{\mathbf{N}}\alpha_n\sigma^2_\varepsilon}{\frac{1}{1-\beta_1\alpha_1}\left[\sum_{\mathbf{n=2}}^{\mathbf{N}}(\alpha_n+\alpha_1\beta_n)^2\mathrm{Var}(x_n)+\sum_{\mathbf{n=2}}^{\mathbf{N}}\alpha_n^2\sigma^2_\varepsilon+\sigma^2_u\right]} \qquad (9)$$

While cumbersome, this expression will prove useful for making an educated guess regarding whether or not we are likely to over- or underestimate the true fiber effect. However, we must keep in mind that while assuming that our controls (variables such as education and income) are exogenous is necessary for a feasible analysis, it is a highly unrealistic assumption.

Another issue with this approach is that we cannot exclude the possibility that there is an unobservable or omitted time-varying variable correlated with both fiber coverage and employment, rendering a causal interpretation impossible. To summarize: While the fixed effects model is relevant, the problems of endogeneity point to the need for a second estimation strategy.

4.2 Instrumental variable approach

A common approach to achieving consistent estimates in the presence of reverse causality is using instrumental variables. An instrument is a variable which is correlated with the endogenous regressor while being independent with respect to the dependent variable, i.e. the instrument is exogenous and does not explain the dependent variable through other channels than the endogenous regressor. Obviously, upon any variable which happens to fulfill these criteria is not enough, the case for using a particular instrument must be made using economically sound arguments. Verbeek (2012) provides a primer on the instrumental variables estimator. Consider a simple linear model:

$$y_i = x_i{}'\beta + \varepsilon_i \qquad (10)$$

$$x_i = z_i{}'\pi + u_i \qquad (11)$$

Where x and y are vectors of dimension $K * 1$. The OLS estimator, $\hat{\beta}_{OLS}$, is solved using K moment conditions, derived from the first order condition for minimizing the sum of squared differences:

$$E[(y_i - x_i{}'\beta)x_i] = 0 \qquad (12)$$

While we cannot observe the error term, these moment conditions requires us to impose:

$$E[\varepsilon_i{}'x_i] = 0$$

If this condition is violated due to endogenity of one or more x_i (as in the model represented by (2) and (3) above), we are no longer consistently estimating β. However, we can find $R \geq K$ instruments, z, for which the exogenity assumption hold:

$$E[\varepsilon_i{}'z_i] = 0$$

In this case, the estimates will be consistent, and a causal interpretation can be possible, i.e. given that we can credibly impose the exogenity assumption, our estimates represent our expectation conditional on all factors, observable as well as unobservable, remaining constant (Verbeek, 2012).

For R=K, the IV estimator can be solved from the sample moment conditions (compare with (12) above):

$$\frac{1}{N}\sum_{i=1}^{N}\left(y_i - x_i'\widehat{\beta_{IV}}\right)z_i = 0 \qquad (13)$$

$$\widehat{\beta_{IV}} = \left(\sum_{i=1}^{N} z_i x_i'\right)^{-1}\sum_{i=1}^{N} z_i y_i \qquad (14)$$

We will not consider the case where R>K here[4]. Note that the instruments can overlap with the covariates in the original specification, i.e. exogenous variables are their own instruments. In addition to the exogeneity assumption, for a set of instruments to be relevant it is required that not all elements of π are zero, i.e. there has to exist a non-zero correlation between the instruments and the endogenous regressor. An extension of these conditions is that non-overlapping instruments should not be significant when added to (10), i.e. they should not themselves explain y, only through our endogenous regressor. This is sometimes called "the exclusion restriction". For here on, I will refer to variables included in both z_i and x_i as exogenous covariates and denote our non-overlapping variables simply as 'instruments'.

A computationally easy way to obtain the IV estimates is by modifying the OLS estimator, performing a so called two-stage least squares regression. Consider (11). Using matrix notation, we can write an expression for the OLS predictions of our endogenous regressor (Verbeek, 2012).

$$\hat{\pi} = (Z'Z)^{-1}Z'X \qquad (15)$$

$$\hat{X} = Z(Z'Z)^{-1}Z'X \qquad (16)$$

Using these fitted values to regress X on Y yields the following expression for the OLS (IV) estimator:

$$\widehat{\beta_{IV}} = (\hat{X}'\hat{X})^{-1}\hat{X}'y (17)$$

Inserting (16) in (17) yields:

$$\widehat{\beta_{IV}} = [X'Z(Z'Z)^{-1}Z'X)]^{-1}X'Z(Z'Z)^{-1}Z'y$$

[4] Estimating this scenario involves using a weighting matrix to render the ZX' matrix invertible. See Verbeek (2012) for a detailed explanation

For the case where R=K, we assume that the $X'Z$ matrix is invertible, which reduces the expression to:

$$\widehat{\beta_{IV}} = (Z'X)^{-1}Z'y \quad (18)$$

As is evident, (18) is nothing more than the IV estimator (14) written in matrix notation.

Finding a good instrument is easier said than done. This thesis follows Kolko's (2012) approach by utilizing topographical variation as an instrument. Kolko (2012) uses data on average slope, we use the standard deviation and variance of a sample elevation profile constructed for each municipality. Our measure arguably does a better job at capturing relevant topographical features. Average slope says little about whether or not an observer would characterize the terrain as smooth or mountainous, which is what must be assumed to be the deciding factor behind the return to fiber investments in terms of coverage.

Our key assumption regarding topographical variation as an instrument is that more mountainous municipalities are at a disadvantage with respect to fiber coverage. Thus, we expect more mountainous regions to enjoy less fiber coverage progress during 2007-2011, ceteris paribus. There are two circumstances which adds to the validity of this hypothesis. Firstly, the initial round of government grants and subsidies ended in 2007. In the following years, government spending on broadband has been less abundant, and grants have primarily been implemented via a rural development program, placing an even greater focus on rural areas (PTS, 2012). While local political initiative is still an influencing factor, it has decreased in relative importance. Extending coverage in sparsely populated areas is arguably more vulnerable to rough terrain compared to dense urban neighborhoods. When considering which projects to support, the cost-benefit analysis is presumed to be skewed towards favoring flat areas, as the cost of extending the grid in these areas is likely to be lower.

Secondly, since the most "profitable" areas (flat and heavily urbanized) are likely to have been prioritized by both private as well as public service providers, by 2007 these areas likely already enjoyed coverage. Combined with the general economic downturn around this time, it is reasonable to expect that even public investors are looking for "bang for the buck" rather than simply investing to reach the coverage rate mandated by public policy, increasing the relevance of topographical variation as an instrument.

Since our instrument does not vary over time we cannot use a fixed effects panel specification, as the instrument would be perfectly collinear with the municipality-specific intercept. A random effects model is not appropriate due to probable correlation between the municipality specific error component and our regressors. Therefore, we move to a purely cross-sectional baseline model in first differences for our TSLS estimation, the second stage of which is specified as:

$$\mathrm{Emp}_{i,2011} - \mathrm{Emp}_{i,2007} = \beta_1 + \beta_2\left(\widehat{\mathrm{Fiber}_{i,2011} - \mathrm{Fiber}_{i,2007}}\right) + \beta_3' x_i + \varepsilon_i \qquad (19)$$

Where x_{it} is a vector of exogenous covariates (controls). The fitted values for the difference in fiber coverage is obtained in an auxiliary regression (the "first stage") where fiber coverage is explained by our exogenous controls as well as our instruments z_i:

$$\mathrm{Fiber}_{i,2011} - \mathrm{Fiber}_{i,2007} = \pi_1 + \pi_2' z_i + \pi_3' x_i + u_i \qquad (20)$$

$$\widehat{\mathrm{Fiber}_{i,2011} - \mathrm{Fiber}_{i,2007}} = \widehat{\pi_1} + \widehat{\pi_2'} z_i + \widehat{\pi_3'} x_i \qquad (21)$$

The intuition behind this procedure is that the fitted values of fiber coverage will be "cleansed" from endogenous variation (in our case assumed to be caused by changes in the employment rate) and should only explain the causal link between fiber coverage and employment.

4.3 Inference under weak instruments

An issue which has come under scrutiny in recent years is inference under weak instruments. Stock, Wright & Yogo (2002) provide an excellent explanation of the implications of weak instruments and well as a review of advances regarding robust inference in the presence of weak instruments. A weak instrument is formally defined as low values of a statistic calculated by dividing the so called concentration parameter[5] by K, the number of instruments. A low value implies that inferences based on asymptotic normality will not be correct.

In the case with a single endogenous regressor under i.i.d. disturbances, a partial F-test on the instruments in the first stage regression is a valid sample equivalent to the concentration parameter, $H_0: \pi_2 = 0$. This statistic is often referred to as the "first stage F-statistic". In the case of multiple endogenous regressors, the concentration parameter becomes a matrix. Cragg & Donald proposed using the eigenvalue for the sample equivalent of this

[5] Formally defined as $\mu^2 = \Pi' Z' Z \Pi / \sigma_u^2$ following the notation in (15) and (16).

matrix as an analogue to the first stage F-statistic in the case of multiple endogenous regressors (Stock et al, 2002).

With an endogenous variable in our model, it can be shown (Verbeek, 2012:147) that the OLS inconsistency depends on the correlation with the error term, ε, according to:

$$\operatorname{plim} \hat{\beta} = \beta + \frac{\operatorname{cov}(x, \varepsilon)}{\sigma_x^2}$$

As shown by Stock et al (2002), the expectation of the TSLS estimator using a completely irrelevant set of instruments, i.e. all elements of $\pi_2 = 0$ in (20), is the probability limit of the OLS estimator. With stronger instruments, the bias of the TSLS estimator decreases. The first stage F-statistic or Cragg-Donald statistic can be compared to a set of critical values for various tests of instrument weakness. In the presence of weak instruments, the TSLS estimator can be shown to have a distinctly non-normal distribution (Staiger & Stock, 1997, Stock et al, 2002). As common Wald tests are based upon point estimates, a sample standard error and an assumption of (asymptotic) normality, this may lead to incorrect inferences. This unfortunate property of weak instruments is not limited to small samples as Staiger & Stock (1997) notes.

As we are interested in hypothesis testing, we will use a method finalized by Stock & Yogo (2002) to gauge the strength of our instrument. The authors tabulate critical values for the Cragg-Donald statistic that are used to test if a standard Wald-test with a nominal significance level of five percent has an actual size (i.e. the probability of rejecting a true null) no greater than an arbitrary threshold. For example, if we are willing to accept a size of 15 percent for a standard t-test, $H_0\colon \beta_{TSLS} = 0$, and use a single instrument for a single endogenous regressor, the critical value for the first stage F-statistic is 8.96. If our obtained statistic is larger, we reject the null, i.e. we reject the hypothesis that the size of the t-test is actually 15 percent or greater. A way of achieving the correct actual size for our Wald-test would be to increase our nominal significance level (e.g. one percent instead of five). But this practice would be detrimental to the power of our test, i.e. a low probability of rejecting a false null. Andrews, Moreira and Stock (2007) examine the poor power properties of a standard Wald test in the presence of weak instruments. They note that due to the asymmetric distribution of TSLS-estimates with weak instruments, rejection rates are particularly low for negative values of β.

Moreira (2003) proposes a method for providing confidence intervals around estimates which are asymptotically robust against weak instruments. Andrews, Moreira &

Stock (2007:117) explains his method as *"the idea of implementing tests in IV regression not using a single fixed critical value, but instead using a critical value that is itself a function of a statistic chosen so that the resulting test has the correct size even if the instruments are weak"*.

As we shall see in the following section, weak instruments are indeed a present in our study, which makes Moreira's method intuitively appealing as a way of further gauging the strength of our instruments. Andrews, Moreira & Stock (2007) show that the underlying statistic has excellent power properties compared to other tests. But the asymptotic probability of rejecting a false null nonetheless decreases with low values of the first stage F-statistic as well as, of course, true values of β close to the null.

4.4 Instrument validity

Another issue is validity of our instrument. In this context, validity refers to the exclusion restriction: our instrument should not in and of itself cause changes in the dependent variable. This is problematic, since we would likely be wrong in assuming that our dependent variable, employment, is independent with respect to our instrument, topographical variation. Two trivial examples would be agriculture and manufacturing, both of which likely benefit from smooth terrain, as it enables large cohesive areas to be cultivated and reduces transportation costs. Using a differenced model alleviates some of our concerns, as we need only to be concerned about the channels (other than fiber coverage) through which topography can affect the change in employment while controlling for other factors correlated with employment and fiber coverage. Since terrain is static over time, its effect on short term changes in employment rates is likely rather small.

We follow Kolko's (2012) methodology and introduce municipal road density as a proxy for transportation costs in order to control for this in our IV model. However, we would like to argue that its implementation is more problematic than Kolko's straight-forward one. Road density is confounded by its close relationship with population density. Controlling for high transportation costs without taking population density into account is a bit of a backwards approach. We want to control for the fact that rough terrain can, by affecting transportation costs, have adverse effects on the growth in employment. However, high transportation costs will likely have very little effect on employment in areas where hardly any people live or work. Thus, we interact road density with a lag of population density to make sure it's relative importance diminishes in sparsely populated municipalities while retaining exogeneity in relation to the employment rate during 2007-2011.

5 Results

5.1 Panel data estimations

As there is little in the way of consensus around best-practices estimating the effects of high-speed internet, I set out to use a general-to-specific modelling approach. A range of interaction terms as well as quadratic terms was included in initial estimations, unfortunately with mostly nonsensical results, e.g. low values for a standard F-test of all non-zero coefficients. These preliminary results are not reported in this paper. An economically justifiable addition to our baseline model is fiber coverage in workplaces interacted with income. This interaction term captures whether or not high-income municipalities experience additional effects of having a high-speed internet connection at work. This potential relationship is related to the findings of Forman et al (2009) regarding the asymmetric distribution of wage increases following internet investment as well as the more general concept of skill-biased technological change, i.e. the idea that productivity increases primarily benefit skilled labor[6]. While we are not primarily examining such interactions, including this term seems both relevant and interesting.

The results from our panel data estimations are reported in table 2. Looking at the first column, the most interesting result is the significant negative effect associated with increased fiber coverage among households. The partial derivative has a straightforward interpretation:

$$\frac{\delta \mathrm{E}(\mathrm{Employment}_{\mathrm{it}} | \mathrm{x}_{\mathrm{it}})}{\delta \mathrm{Fiber\ HH}_{\mathrm{it}}} = -0.0124$$

A one unit increase in fiber coverage (equivalent to a leap from no coverage to full coverage) is associated with an expected 1.2 percentage unit decrease in municipal employment rate, ceteris paribus. Increased fiber coverage among workplaces is however associated with a positive effect on employment, although insignificant at any conventional level. Jointly, our two effects do not support the idea that optic fiber has a positive effect on employment. We

[6] See for example Violante (2006) for a detailed explanation of skill-biased technological change.

also do not find evidence that municipal income is a deciding factor behind the effects of increased workplace coverage on employment, as the interaction term is insignificant.

It should be noted that our two fiber coverage variables are both insignificant when included separately in the baseline model. Given our assumption that they both have a unique impact on employment, we would be introducing a source of bias by not including them jointly. The bias can be expected to be quite severe due to the high sample correlation between fiber coverage among households and workplaces[7].

5.2 Robustness checks

To examine the sensitivity of our results in the first column of table 2, two subsamples are estimated in addition to our main model. In the second column of table 2, we exclude a number of municipalities based on categories created and maintained by SKL (Sveriges kommuner och landsting, 2011). We exclude all major cities (with a population of 50,000 and above) and their surrounding commuter towns (where more than 50 percent of the population commute to another municipality). Since the labor market in a commuter town is likely affected by changes in fiber coverage in a neighboring city, towns like these may serve to weaken the link between municipal coverage and employment rate. Using this subsample, we see that both coverage effects are negative, but the magnitude of the workplace effect is far greater than the household effect. None of the two are significant. In general, we should expect to lose significance after dropping about a third of our total observations, but the reason for the reversal of the workplace effect is not clear.

In the third column, we weight our observations according to municipal population. This is a useful exercise, but not essential for this thesis as it serves to answer questions of a nationwide fiber effect. With weighting, both types of fiber coverage are insignificant, but the point estimates have reversed. Weighting by population places great importance on major cities, so what we could be seeing is simply that the market in large urban areas is more adapted to utilize increased fiber coverage among households. As for workplaces, by 2007 the firms enjoying the greatest productivity increases from fiber were probably already covered. However, as our modelling approach might not be optimal for examining nation-wide effects I can not draw further conclusions from these results. The fact that our two subsample

[7] The sample correlation coefficient is 0.97.

estimations lead us to very different results points to the possibility of heterogeneous effects of fiber coverage among municipalities.

Could it be that the negative effect associated with increased coverage among households in our main model is the result of omitting an unobservable variable correlated with both employment and fiber coverage? While we cannot dismiss this possibility, there are few candidates with opposing effects on our two variables interest, e.g. a variable with a negative effect on employment and a positive effect on fiber coverage among households. An obvious shortcoming is that we do not have data on the amount of government grants received by each municipality. Given that they are geared towards rural areas, the grants may represent an omitted variable. However, as detailed in section 2, government grants make up far from all fiber financing. In fact, we observe a positive correlation between the increase in household coverage 2007-2011 and average municipal population during the period[8]. On top of this, there is no clear relationship between average municipal population and the change in employment during the period[9]. Thus, we do not have any evidence indicating that our results are simply due to the fact that areas with low employment rates received more fiber coverage due the design of the government grants.

[8] The sample correlation coefficient is 0.1889.

[9] The sample correlation coefficient is 0.0217.

Table 2: Panel data estimations (2007-2011)
Dependent variable: Employment

Fiber HH	-0.0124 (0.0069)*	-0.0077 (0.0087)	0.0136 (0.0120)
Fiber WP	0.1641 (0.1269)	-0.6065 (0.5073)	-0.1454 (0.1852)
Population	-0.0292 (0.0229)	-0.1072 (0.0363)***	-0.0371 (0.0262)
Education	0.41079 (0.0779)***	0.3919 (0.1208)***	0.3766 (0.0942)***
Income	0.0990 (0.0186)***	0.0945 (0.0229)***	0.0939 (0.0233)***
Income $*$ Fiber WP	-0.0121 (0.0101)	0.0506 (0.0414)	0.0104 (0.0145)
No of observations	1450	980	1450
Weighted by population	No	No	Yes
Excluded municipalities	None	Major cities and surrounding commuter towns	None

****: significant at the 1 percent level*
***: significant at the 5 percent level*
**: significant at the 10 percent level*

Robust standard clustered around municipality in parentheses. All models include municipality specific as well as period specific fixed effects.
Population and income refer to the natural logarithms of these variables.
The second column excludes groups 1-4 in the SKL classification (SKL, 2011).
The third column places analytical weights on each observation according to average municipal population during 2007-2011.

Returning to our analysis of bias due to reverse causality in section 4. Our simplified setup and the resulting expression for the probability limit of the effect of fiber, repeated below for convenience, allows us to make an educated guess regarding the direction of our bias.

$$\text{plim}\,\hat{\beta}_1 = \beta_1 + \frac{\sum_{n=2}^{N} \alpha_n \sigma_\varepsilon^2}{\frac{1}{1-\beta_1\alpha_1}\left[\sum_{n=2}^{N}(\alpha_n + \alpha_1\beta_n)^2 \operatorname{Var}(x_n) + \sum_{n=2}^{N} \alpha_n^2 \sigma_\varepsilon^2 + \sigma_u^2\right]}$$

Focusing on fiber coverage among households, in our main model (first column of table 2) the point estimate of β_1 is negative. As for our controls, our point estimates of β_n are all positive, i.e. education and income are both positively correlated with employment. Let us further assume that employment, education, population and income all have a positive effect on fiber coverage among households, i.e. $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ are all positive. Given these assumptions, the probability limit of $\hat{\beta}_1$ will necessarily be greater than the true value of β_1. Thus, the negative effect associated with fiber among households could in fact be even more negative. However, note that this exercise is entirely dependent on the assumption that our controls are exogenous, a very strong assumption.

5.3 IV estimations

As stated previously, our static instrument necessitates a purely cross-sectional model. In the first stage of our two-stage least squares approach we obtain fitted values for our endogenous regressor (fiber coverage) using our instrument as well as exogenous covariates. See appendix for results from the first stage regressions. To control for non-linear effects of topography, we include both the standard deviation as well as the variance of our sample elevation profiles as instruments for fiber coverage. Although we technically only have a single instrument at our disposal, the inclusion of a quadratic term allows us to include both types of fiber coverage as endogenous regressors and still be able to fully identify all parameters. As we hypothesize that both types of fiber coverage are uniquely relevant, omitting one of them could introduce inconsistency.

A simple way of examining the strength of a potential instrument is the sample correlation between the instrument and the endogenous regressor. The correlation between topographical standard deviation and fiber coverage among households is stronger than the correlation with fiber coverage among workplaces[10]. This is expected, since households presumably have a greater geographical dispersion across the municipality compared to workplaces. Combined with the fact that our panel data model produced interesting results, suggesting a negative effect of fiber coverage among households, we will focus on fiber coverage among households in our IV approach to try to confirm the results from our panel data estimates.

One of our concerns from the panel data model is that none of our controls can be attributed exogeneity in relation to contemporaneous changes in employment rate. Therefore, we specify two models, differing only in the set of controls used. One in which we use a lag in levels as a way of achieving plausible exogenity. For example, since average municipal income in 2007 can be assumed to be exogenous with respect to the change in income between 2007

[10] The sample correlation coefficient is -0.1984 for households versus -0.1456 for workplaces.

and 2011, we can safely control for income without violating the assumptions of the TSLS estimator. We simply use lags in levels as a way of making our controls are their own instruments. We also estimate a model where the controls are differenced in the same way that we difference our dependent variable. While this practice violates the basic assumptions of using IV (the change in average income from 2007 to 2011 is not exogenous with respect to the change in employment during the same period), consistently differencing our controls could better explain the variation in employment and thus provide a more reliable estimate of the fiber effect. Models with the change in employment rate from 1999 to 2003 as a control for any unobservable municipal ability regarding employment were also estimated initially. This control did not significantly alter the results and as including lags of the dependent variable can be problematic when interpreting the results, these models are not included here.

Tables 3 and 4 present the results from our TSLS estimations. As way of examining the sensitivity of our results, four different specifications were estimated for each of our two baseline models. In the first columns of tables 3 and 4, fiber coverage among households as well as among workplaces were included as endogenous regressors. In the second column, coverage among households are included as an endogenous regressor while coverage among workplaces is treated as an exogenous control. This constitutes a violation the basic assumptions of IV since we have good reason to suspect that fiber coverage among workplaces is partially determined by employment, but it is nonetheless interesting to compare these results to those in the first column. The point estimate of the workplace effect does not seem to differ much from the instrumented estimate in the first column, this could be a sign that any correlation between workplace coverage and the error term is not enough to be a significant source of inconsistency. However, the Derbin-Wu-Hausman[11] test suggests that OLS is, as hypothesized, inefficient. In the third and fourth column, fiber coverage among workplaces and households are excluded, respectively. As the two are correlated, we are introducing a source of bias by not including them jointly. The comparatively small and insignificant estimation of the fiber can probably be attributed to the fact that one variable is left two explain the two diametrically different influences of fiber coverage.

Overall, the results from our two models are similar. Both appear to confirm the results from our main fixed effects model, i.e. a negative effect associated with household coverage

[11] IV extension of a standard Hausman (1978) specification test. Tests the null of OLS consistency against the alternative of OLS inconsistency by implementing the residuals from the first stage TSLS-regression as a variable in an auxiliary second stage regression. If the residuals are significant in the auxiliary regression, OLS is inconsistent.

and a positive effect of workplace coverage. As expected, the controls are far less significant when expressed as levels. Evidently the level of income, population and education are not good predictors of future short-term changes in employment. Disregarding endogeneity issues, the Cragg-Donald statistic suggests that the two sets of controls performs equally as first-stage predictors of the change fiber coverage. Worth noting is also the increased size and significance of the household coverage effect in the model with controls expressed as levels, this could suggest that there are issues of endogeneity present in the differenced model.

There are many reasons to be cautious when interpreting the results. The effects of fiber coverage are unrealistically large. The partial derivative of change in employment with respect to the change in household coverage over the period is about -0.19 to -0.24, i.e. a 100 percentage unit increase in coverage is associated with a 19-24 percentage unit decrease in municipal employment rate compared to no change in coverage during 2007-2011. However, it is useful to keep in mind that in a real world setting, this negative effect is confounded with the apparently positive effect associated with increase workplace coverage, which is greater in magnitude than the household effect in all our models. Compared to our panel data model, our IV models do not offer the same possibility to control for unobservable heterogeneity as we have no municipal-specific intercepts. Consequently, we run a greater risk of omitting a relevant control, observable or not.

Our main concern is the weakness of our instruments. Upon comparing the Cragg-Donald statistic to the tabulated values of Stock & Yogo (2005), it is apparent that the true size of any standard Wald-test on our estimates have a size greater than 25 percent, the largest size for which a critical value is commonly tabulated. However, Moreira's (2003) conditional confidence intervals does reject the hypothesis that the effect associated with fiber coverage is equal to zero at the five percent level in the model where only household coverage is instrumented (the third column of tables 3 and 4). Moreira's method cannot readily be applied to models with multiple endogenous variables, and for our two models with a single coverage variable (columns 3 and 4 in both tables), the standard errors are too large to provide a bounded interval. Taken both measures of instrument strength into account, our assessment is that topographical variation simply does not explain fiber coverage well enough to provide robust results. The magnitude of issues caused by our dependent variable being dependent on our instrument is difficult to judge, but this can likely be handled be a well-specified set of controls.

Table 3: IV estimations, controls in levels

$TSLS\ estimations, second\ stage. Dependent\ variable: \Delta Employment_{2007-2011}$

	Both fiber types of coverage are instrumented	Only household coverage is instrumented	Only household coverage is included	Only workplace coverage is included
$\Delta Fiber\ HH_{2007-2011}$	-0.2319 (0.09586)**	-0.2318 (0.0957)**	-0.0725 (0.0538)	
$\Delta Fiber\ WP_{2007-2011}$	0.2842 (0.1393)**	0.2778 (0.1131)**		0.0092 (0.0538)
$Population_{2007}$	0.0004 (0.0024)	0.0005 (0.0018)	0.0004 (0.0023)	-0.0022 (0.0014)
$Income_{2007}$	-0.0028 (0.0225)	-0.0014 (0.0142)	0.0281 (0.0162)	0.0121 (0.0145)
$Education_{2007}$	0.0137 (0.0267)	0.0140 (0.0264)	0.03539 (0.0237)	0.0392 (0.0163)**
$Road\ density_{2005} * Population\ density_{2007}$	0.00004 (0.0002)	0.00005 (0.0001)	0.00002 (0.0002)	-0.0002 (0.0001)
No of observations	289	289	289	289
Cragg-Donald statistic	1.333	4.654	1.770	1.333
Conditional C.I. (p-value for $H_0: \beta_{fiber\ HH} = 0$)		0.0004	Unbounded	
Pagan-Hall (p-value for H_0: homoskedastic errors)	0.8761	0.8814	0.9567	0.8404
Derbin-Wu-Hausman (p-value for H_0: instrumented fiber coverage is exogenous)	0.0024	0.0005	0.0444	0.9301

****: significant at the 1 percent level*
***: significant at the 5 percent level*
**: significant at the 10 percent level*

Standard errors in parentheses. Population, population density and income refer to the natural logarithm of these variables.

Table 4: IV estimations, controls in differences

$TSLS\ estimation, second\ stage. Dependent\ variable{:}\ \Delta Employment_{2007-2011}$

	Both types of coverage are instrumented	Only household coverage is instrumented	Only household coverage is included	Only workplace coverage is included
ΔFiber $HH_{2007-2011}$	-0.1949 (0.1119)*	-0.1895 (0.1065)*	-0.0519 (0.0531)	
ΔFiber $WP_{2007-2011}$	0.2102 (0.1411)	0.2293 (0.1248)*		-0.0005 (0.0507)
$\Delta Population_{2007-2011}$	0.0483 (0.0573)	0.0431 (0.0531)	-0.00008 (0.0436)	-0.02712 (0.0262)
$\Delta Income_{2007-2011}$	0.1499 (0.0366)***	0.1535 (0.0352)***	0.1502 (0.0351)***	0.1604 (0.0263)***
$\Delta Education_{2007-2011}$	0.2303 (0.1486)	0.2229 (0.1442)	0.3481 (0.1193)***	0.3578 (0.0928)***
Road $density_{2005}$ $*$ Pop $density_{2007}$	0.000033 (0.00025)	-.00005 (.0001)	0.0002 (0.0002)	-0.00003 (0.0002)
Number of observations	289	289	289	289
Cragg-Donald statistic	1.183	2.669	1.240	1.225
Conditional C.I. (p-value for H_0: $\beta_{fiber} = 0$)		0.0234	Unbounded	
Pagan-Hall (p-value for H_0: Homoskedastic errors)	0.7720	0.7728	0.8082	0.1134
Derbin-Wu-Hausman (p-value for H_0: instrumented fiber coverage is exogenous)	0.0687	0.0239	0.1671	0.8737

****: significant at the 1 percent level*
***: significant at the 5 percent level*
**: significant at the 10 percent level*

Standard errors in parentheses. Population, population density and income refer to the natural logarithm of these variables.

6 Discussion

The data set examined in this thesis is very different from the sets used previous research. Longitudinal data on actual broadband availability does not seem to be readily available in a majority of countries, and many researchers are forced to use proxies (Kolko, 2012 and Kim & Orazem, 2012). In addition to a good measure of actual availability, the primary advantage of our data is the separation of coverage between workplaces and households. Previous studies have used a single coverage variable, even though the internet applications utilized by businesses and the public sector are very different from those utilized by consumers. Our results indicate that there are indeed multiple effects at work. We have considered a few of the methodological pitfalls which cloud the possibilities of form conclusions, the evidence is there for the reader to judge. Or primary concerns are the weakness of our instruments and the apparent sensitivity to different specifications. These concerns can at least partially be attributed to our fairly small number of observations. As a comparison, Kolko (2012) has roughly 26 000 observations at his disposal. The fact that our sample covers years characterized by a general economic downturn also puts the generalizability of our results into question. We must also keep in mind that a decrease in municipal employment rate can be caused by a net inflow of unemployed rather than a decrease in the number of employed, which is why we control for municipal population. But we cannot fully dismiss the possibility that expanded fiber coverage is correlated with a number of unobserved factors jointly influencing the desirability of a municipality. A change in desirability could cause an in- or outflow of unemployed looking for work, as people can be assumed to be fairly mobile across municipalities. However, expanding fiber coverage is in itself a laborious task. Although any directly job-creating effects are most certainly insignificant in any major city, it does serve to further merit our results for more extensive research. Potential econometric issues caused by the high degree of correlation between our coverage variables is a source of error left unexamined here, but this is nonetheless an issue to be addressed by future research using coverage data with similar multicollinear properties.

Assuming there is truth behind our somewhat controversial finding of a negative effect associated with fiber coverage among households, an explanation might be found in the fact that we are studying a new generation of internet access technology. As such, our results must be interpreted in light of the set of internet applications made available by fiber. Our hypothesis is that the benefits of this new set of applications is geared towards consumers, large organizations and businesses in the field of high-tech. The latter two are arguably more likely to be established in major cities, which could explain any heterogeneous effects across municipalities, briefly explored here by weighting observations according to population. With weighting, household coverage is positively associated with employment, although weakly so. This is in line with Forman et al (2009) who found that internet investment is associated with increased wages primarily among skilled urban works.

As for consumer applications, if we recognize that all internet applications can be placed in one of two categories, productivity and entertainment, the argument is fairly straight-forward. Presumably, it is mainly the latter category which enjoys the benefits of the increases bandwidth and decreased latency associated with a fiber optic connection. Video on demand-services, online gaming, gambling and peer-to-peer file sharing are just a few examples of entertainment applications whose quality increases in proportion to bandwidth. On the productivity side, applications such as online job-searching, e-mail and teleconferencing are arguably less prone to increase in quality as connections are enhanced. Continuing along this line of reason, a possible explanation is that increased availability of high-speed internet has mainly boosted internet applications which are substitutes rather than compliments to labor, giving rise to an increased demand for leisure. This is a highly disputable statement, but it is nonetheless very likely that entertainment applications are universally consumed whereas the domestic use of “productive” applications is limited to a select few and highly dependent on one’s occupation, education and IT-literacy.

7 Conclusion

Our results weakly support the consensus surrounding the positive effect of IT on firm productivity, expressed in our study as the positive effect on employment associated with increased fiber coverage among workplaces. This effect is most evident in our IV model, although with varying degrees of significance. The effect fluctuates and even reverses across subsamples and model specifications, possibly due to the small sample employed.

What is difficult to reconcile with previous research is the observed negative effect associated with fiber coverage among households. While our instrumental variable approach does not provide conclusive support of the negative effect of fiber coverage among households observed in our fixed effects regression, presumably due to a weak set of instruments and/or too few observations, the possibility of such effects should not be dismissed. As studies such as Kolko (2012) have not separated household and workplace coverage, our results do not directly contradict previous results. They do however pose an interesting set of questions.

Methodological issues aside, speculating around why a negative effect might arise is not difficult. As the cost of fiber decreases, the technology is likely to be adopted by consumers to whom high speed internet have a low or even negative marginal productivity in terms of employment. In a broader perspective, we are facing difficult questions surrounding the economic effects of total internet diffusion after the initial broadband honeymoon is over.

As for future research topics, an alternative identification strategy and/or more data is needed in order to provide more robust results. Our approach does not control for the possibility that there is some time-varying variable affecting both changes in fiber coverage and employment, observable or not.

8 References

8.1 *Government and organizational*

Näringsdepartementet, (2009). Bredbandsstrategi för Sverige (N2009/8317/ITP)

Näringsdepartementet, (2011). Nyttan av bredband – en fallstudie av sex kommuner. Retrived from www.regeringen.se/sb/d/14679/a/167318 on August 20 2013.

OECD, (2012). Percentage of fiber connections in total broadband subscriptions, Dec. 2012. Data retrived from OECD broadband portal at: http://www.oecd.org/sti/broadband/oecdbroadbandportal.htm on August 22 2013.

Post- och telestyrelsen (PTS), (2007). Bredband i Sverige 2007 (PTS-ER-2007:17)

Post- och telestyrelsen, (2008). Svart fiber – marknad och konkurrenssituation (PTS-ER-2008:9)

Post- och telestyrelsen, (2010). Beslut om marknaden för nätinfrastrukturtillträde, 2010-05-24. (Dnr: 07-11757/23). Retrived from http://www.pts.se

Post- och telestyrelsen, (2012). Handläggning av bredbandsstöd. Accessed at: http://www.pts.se/sv/Privat/Internet/Bredband-ADSL/Bredbandsstod-for-landsbygden/ on August 22 2013.

Sveriges kommuner och landsting, (2011). Kommungruppsinledning. Accessed at: http://www.skl.se/kommuner_och_landsting/fakta_om_kommuner/kommungruppsindelning on August 25 2013

8.2 Academic

Andrews, D.W.K., Moreira, M.J., & Stock, J.H. (2007). Performance of conditional Wald tests in IV regression with weak instruments. *Journal of Econometrics*, (139), 116–132.

Crandall, R., Lehr, W., Litan, R. (2007). The effects of broadband deployment on output and employment: A cross-sectional analysis of U.S. data. *Issues in Economic Policy*, discussion paper no. 6. Brookings institution. Retrived from: http://www.brookings.edu/~/media/Files/rc/papers/2007/06labor_crandall/06labor_crandall.pdf

Forman, C., Goldfarb, A., Greenstein, S. (2009). The Internet and local wages: Convergence or divergence? NBER Working paper no. 14750.

Greenstein, S., & McDevitt, R. C. (2011). The broadband bonus: Estimating broadband internet's impact on U.S. GDP. *Telecommunications Policy,* (35), 617-632.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica,* (46), 1251-1271.

Jorgenson, D.W., Stiroh, K. J. (2000). Raising the speed limit: U.S. Economic growth in the information age. *Brookings papers on economic activity,* (1:2000)

Jorgenson, D.W., Ho, M.S. & Stiroh, K. J. (2008) A retrospective look at the U.S. productivity growth resurgence. *Journal of Economic Perspectives,* (22), 3-24.

Kim, Y., & Orazem, P. (2012). Broadband internet and firm entry: Evidence from rural Iowa. Working paper, Department of Economics, Iowa State University.

Kolko, J. (2012). Broadband and local growth. *Journal of Urban Economics*, (71), 100–113.

Moreira, M.J. (2003). A conditional likelihood ratio test for structural models. *Econometrica,* (71), 1027–1048.

Noh, Y., & Yoo, K. (2007). Internet, inequality and growth. *Journal of Policy Modeling*, (30), 1005-1016.

Staiger, D., Stock, J.H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, (65), 557–586.

Stock, J.H., Wright, J., & Yogo, M. (2002). GMM, weak instruments and weak identification. Unpublished paper prepared for the *Journal of Business and Economic Statistics* symposium on GMM.

Stock, J. H., Yogo, M. (2002). Testing for Weak Instruments in Linear IV Regression, NBER Technical Working Paper No. 284.

Verbeek, M. (2012). *A guide to modern econometrics* (4th ed.). John Wiley & Sons.

Violante, G. L. (2006). Skill-Biased Technical Change. In L. Blume and S. Durlauf (Eds.), *The New Palgrave Dictionary of Economics* (2nd ed.). Palgrave Macmillan.

9 Appendix

9.1 TSLS-estimations, first stage STATA print-outs

```
. ivreg2 emp_diff (fh_diff fw_diff = topo_stdev topo_var) l4.ln_pop l4.roads_pop l4.ln_inc l4.edu, first

First-stage regressions
-----------------------

First-stage regression of fh_diff:

OLS estimation
--------------

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

                                                      Number of obs =      289
                                                      F(  6,   282) =     8.74
                                                      Prob > F      =   0.0000
Total (centered) SS     =  8.937947542                Centered R2   =   0.1569
Total (uncentered) SS   =  23.23575931                Uncentered R2 =   0.6757
Residual SS             =  7.535958009                Root MSE      =    .1635

------------------------------------------------------------------------------
     fh_diff |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ln_pop |
         L4. |    .036475   .0138476     2.63   0.009     .0092172    .0637328
             |
   roads_pop |
         L4. |   .0017074   .0011327     1.51   0.133    -.0005222    .0039369
             |
      ln_inc |
         L4. |   .1257182   .1302477     0.97   0.335    -.1306629    .3820993
             |
         edu |
         L4. |   .0155951   .2401142     0.06   0.948    -.4570486    .4882388
             |
  topo_stdev |  -.0015026   .0008001    -1.88   0.061    -.0030775    .0000724
    topo_var |   6.93e-06   4.07e-06     1.70   0.090    -1.08e-06    .0000149
       _cons |  -1.642369   1.581761    -1.04   0.300    -4.755926    1.471188
------------------------------------------------------------------------------
Included instruments: L4.ln_pop L4.roads_pop L4.ln_inc L4.edu topo_stdev
                      topo_var
------------------------------------------------------------------------------

. ivreg2 emp_diff (fh_diff = topo_stdev topo_var) fw_diff l4.ln_pop l4.roads_pop l4.ln_inc l4.edu, first

First-stage regressions
-----------------------

First-stage regression of fh_diff:

OLS estimation
--------------

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

                                                      Number of obs =      289
                                                      F(  7,   281) =   335.27
                                                      Prob > F      =   0.0000
Total (centered) SS     =  8.937947542                Centered R2   =   0.8931
Total (uncentered) SS   =  23.23575931                Uncentered R2 =   0.9589
Residual SS             =   .955746028                Root MSE      =   .05832

------------------------------------------------------------------------------
     fh_diff |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     fw_diff |   1.178739   .0267988    43.98   0.000     1.125987    1.231491
             |
      ln_pop |
         L4. |   .0106618    .004975     2.14   0.033     .0008688    .0204547
             |
   roads_pop |
         L4. |   .0004995    .000405     1.23   0.218    -.0002977    .0012968
             |
      ln_inc |
         L4. |  -.1078641   .0467693    -2.31   0.022    -.1999268   -.0158014
             |
         edu |
         L4. |   -.043145    .085673    -0.50   0.615    -.2117874    .1254973
             |
  topo_stdev |   -.000467   .0002864    -1.63   0.104    -.0010308    .0000968
    topo_var |   8.18e-07   1.46e-06     0.56   0.575    -2.05e-06    3.69e-06
       _cons |   1.228106   .5680665     2.16   0.031     .1099001    2.346312
------------------------------------------------------------------------------
Included instruments: fw_diff L4.ln_pop L4.roads_pop L4.ln_inc L4.edu topo_stdev
                      topo_var
------------------------------------------------------------------------------
```

```
. ivreg2 emp_diff (fh_diff = topo_stdev topo_var) l4.ln_pop l4.roads_pop l4.ln_inc l4.edu, first

First-stage regressions
-----------------------

First-stage regression of fh_diff:

OLS estimation
--------------

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

                                                      Number of obs =      289
                                                      F(  6,   282) =     8.74
                                                      Prob > F      =   0.0000
Total (centered) SS     =  8.937947542                Centered R2   =   0.1569
Total (uncentered) SS   =  23.23575931                Uncentered R2 =   0.6757
Residual SS             =  7.535958009                Root MSE      =    .1635

------------------------------------------------------------------------------
     fh_diff |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ln_pop |
         L4. |    .036475   .0138476     2.63   0.009     .0092172    .0637328
             |
   roads_pop |
         L4. |   .0017074   .0011327     1.51   0.133    -.0005222    .0039369
             |
      ln_inc |
         L4. |   .1257182   .1302477     0.97   0.335    -.1306629    .3820993
             |
         edu |
         L4. |   .0155951   .2401142     0.06   0.948    -.4570486    .4882388
             |
  topo_stdev |  -.0015026   .0008001    -1.88   0.061    -.0030775    .0000724
    topo_var |   6.93e-06   4.07e-06     1.70   0.090    -1.08e-06    .0000149
       _cons |  -1.642369   1.581761    -1.04   0.300    -4.755926    1.471188
------------------------------------------------------------------------------
Included instruments: L4.ln_pop L4.roads_pop L4.ln_inc L4.edu topo_stdev
                      topo_var
------------------------------------------------------------------------------

. ivreg2 emp_diff (fw_diff = topo_stdev topo_var) l4.ln_pop l4.roads_pop l4.ln_inc l4.edu, first

First-stage regressions
-----------------------

First-stage regression of fw_diff:

OLS estimation
--------------

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

                                                      Number of obs =      289
                                                      F(  6,   282) =     8.89
                                                      Prob > F      =   0.0000
Total (centered) SS     =  5.631396369                Centered R2   =   0.1590
Total (uncentered) SS   =  15.99660603                Uncentered R2 =   0.7039
Residual SS             =  4.735925113                Root MSE      =    .1296

------------------------------------------------------------------------------
     fw_diff |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ln_pop |
         L4. |   .0218991   .0109776     1.99   0.047     .0002906    .0435075
             |
   roads_pop |
         L4. |   .0010247   .0008979     1.14   0.255    -.0007427    .0027922
             |
      ln_inc |
         L4. |   .1981629   .1032531     1.92   0.056    -.0050818    .4014076
             |
         edu |
         L4. |   .0498331   .1903492     0.26   0.794    -.3248525    .4245187
             |
  topo_stdev |  -.0008785   .0006343    -1.39   0.167    -.0021271      .00037
    topo_var |   5.18e-06   3.23e-06     1.61   0.109    -1.17e-06    .0000115
       _cons |  -2.435209   1.253932    -1.94   0.053    -4.903464    .0330454
------------------------------------------------------------------------------
Included instruments: L4.ln_pop L4.roads_pop L4.ln_inc L4.edu topo_stdev
                      topo_var
------------------------------------------------------------------------------
```

```
. ivreg2 emp_diff (fh_diff fw_diff = topo_stdev topo_var) ln_pop_diff l4.roads_pop inc_diff edu_diff, first

First-stage regressions
-----------------------

First-stage regression of fh_diff:

OLS estimation
--------------

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

                                                      Number of obs =      289
                                                      F(  6,   282) =     6.51
                                                      Prob > F      =   0.0000
Total (centered) SS     =  8.937947542                Centered R2   =   0.1217
Total (uncentered) SS   =  23.23575931                Uncentered R2 =   0.6621
Residual SS             =  7.850216656                Root MSE      =    .1668

------------------------------------------------------------------------------
     fh_diff |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 ln_pop_diff |   .4049197   .4089079     0.99   0.323    -.3999795    1.209819
             |
   roads_pop |
         L4. |   .0033343   .0010771     3.10   0.002      .001214    .0054546
             |
    inc_diff |  -.1245283   .4106107    -0.30   0.762    -.9327794    .6837227
    edu_diff |  -.2397411   1.424854    -0.17   0.867    -3.044441    2.564959
  topo_stdev |  -.0012427   .0008011    -1.55   0.122    -.0028197    .0003343
    topo_var |   5.48e-06   4.10e-06     1.34   0.183    -2.60e-06    .0000135
       _cons |   .2404599   .0537424     4.47   0.000     .1346728    .3462471
------------------------------------------------------------------------------
Included instruments: ln_pop_diff L4.roads_pop inc_diff edu_diff topo_stdev
                      topo_var
------------------------------------------------------------------------------

. ivreg2 emp_diff (fh_diff = topo_stdev topo_var) fw_diff ln_pop_diff l4.roads_pop inc_diff edu_diff, first

First-stage regressions
-----------------------

First-stage regression of fh_diff:

OLS estimation
--------------

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

                                                      Number of obs =      289
                                                      F(  7,   281) =   325.95
                                                      Prob > F      =   0.0000
Total (centered) SS     =  8.937947542                Centered R2   =   0.8903
Total (uncentered) SS   =  23.23575931                Uncentered R2 =   0.9578
Residual SS             =  .9800771016                Root MSE      =   .05906

------------------------------------------------------------------------------
     fh_diff |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     fw_diff |   1.171609   .0263983    44.38   0.000     1.119645    1.223572
 ln_pop_diff |    .338816   .1447469     2.34   0.020     .0538901    .6237418
             |
   roads_pop |
         L4. |  -.0001382   .0003892    -0.36   0.723    -.0009044    .0006279
             |
    inc_diff |   .0309501   .1453842     0.21   0.832    -.2552302    .3171304
    edu_diff |  -.7632073   .5044868    -1.51   0.131     -1.75626    .2298458
  topo_stdev |  -.0001473   .0002846    -0.52   0.605    -.0007076     .000413
    topo_var |  -5.26e-07   1.46e-06    -0.36   0.718    -3.40e-06    2.34e-06
       _cons |   .0270241   .0196214     1.38   0.170    -.0115994    .0656477
------------------------------------------------------------------------------
Included instruments: fw_diff ln_pop_diff L4.roads_pop inc_diff edu_diff
                      topo_stdev topo_var
------------------------------------------------------------------------------
```

```
. ivreg2 emp_diff (fh_diff = topo_stdev topo_var) ln_pop_diff inc_diff edu_diff l4.roads_pop, first

First-stage regressions
-----------------------

First-stage regression of fh_diff:

OLS estimation
--------------

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

                                                     Number of obs =      289
                                                     F(  6,   282) =     6.51
                                                     Prob > F      =   0.0000
Total (centered) SS     =  8.937947542               Centered R2   =   0.1217
Total (uncentered) SS   =  23.23575931               Uncentered R2 =   0.6621
Residual SS             =  7.850216656               Root MSE      =    .1668

------------------------------------------------------------------------------
     fh_diff |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 ln_pop_diff |   .4049197   .4089079     0.99   0.323    -.3999795    1.209819
    inc_diff |  -.1245283   .4106107    -0.30   0.762    -.9327794    .6837227
    edu_diff |  -.2397411   1.424854    -0.17   0.867    -3.044441    2.564959
             |
   roads_pop |
         L4. |   .0033343   .0010771     3.10   0.002      .001214    .0054546
             |
  topo_stdev |  -.0012427   .0008011    -1.55   0.122    -.0028197    .0003343
    topo_var |   5.48e-06   4.10e-06     1.34   0.183    -2.60e-06    .0000135
       _cons |   .2404599   .0537424     4.47   0.000     .1346728    .3462471
------------------------------------------------------------------------------
Included instruments: ln_pop_diff inc_diff edu_diff L4.roads_pop topo_stdev
                      topo_var
------------------------------------------------------------------------------

. ivreg2 emp_diff (fw_diff = topo_stdev topo_var) ln_pop_diff inc_diff edu_diff l4.roads_pop, first

First-stage regressions
-----------------------

First-stage regression of fw_diff:

OLS estimation
--------------

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

                                                     Number of obs =      289
                                                     F(  6,   282) =     5.88
                                                     Prob > F      =   0.0000
Total (centered) SS     =  5.631396369               Centered R2   =   0.1112
Total (uncentered) SS   =  15.99660603               Uncentered R2 =   0.6871
Residual SS             =  5.004958943               Root MSE      =    .1332

------------------------------------------------------------------------------
     fw_diff |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 ln_pop_diff |   .0564213   .3265013     0.17   0.863    -.5862678    .6991105
    inc_diff |  -.1327051    .327861    -0.40   0.686    -.7780705    .5126603
    edu_diff |   .4467928   1.137705     0.39   0.695     -1.79268    2.686266
             |
   roads_pop |
         L4. |   .0029639   .0008601     3.45   0.001     .0012709    .0046568
             |
  topo_stdev |   -.000935   .0006397    -1.46   0.145    -.0021941    .0003242
    topo_var |   5.12e-06   3.27e-06     1.56   0.119    -1.32e-06    .0000116
       _cons |   .1821733   .0429118     4.25   0.000     .0977053    .2666413
------------------------------------------------------------------------------
Included instruments: ln_pop_diff inc_diff edu_diff L4.roads_pop topo_stdev
                      topo_var
------------------------------------------------------------------------------
```

9.2 Python script to obtain municipal elevation profiles

```
import csv
import urllib.request
from xml.etree import ElementTree as ET

COORD_BASE_URL = "http://nominatim.openstreetmap.org/search/se"
ELEVATION_BASE_URL = "http://maps.googleapis.com/maps/api/elevation/xml"

def CsvOutput(url,mun_query):
    xml_string = urllib.request.urlopen(url).read()
    xml = ET.fromstring(xml_string)
    mun_list=[]
    mun_list.append(str(mun_query).strip("[]"))
    for e in xml.iter('elevation'):
        elev = ET.tostring(e, method="text", encoding='utf-8')
        mun_list.append(elev)
    return mun_list

def GetElevation(coords,mun_query):
    e=0
    i=1
    l = []
    while i < 5:
        first = coords.find('"',e)
        last = coords.find('"',first+1)
        l.append(coords[first+1:last-1])
        e=last+1
        i=i+1
    elevation_url = ELEVATION_BASE_URL + "?path=" + str(l[0]).strip('[]') + "," +
str(l[2]).strip('[]') + "|" + str(l[1]).strip('[]') + "," + str(l[3]).strip('[]') +
"&samples=200&sensor=false"
    print(elevation_url)
    return CsvOutput(elevation_url,mun_query)


def GetCoords(mun_query):
    url = COORD_BASE_URL + '?q=' + str(mun_query).strip('[]') + '&format=json'
    json = urllib.request.urlopen(url).read()
    jsonstr = str(json)
    begin = jsonstr.find('relation')
    end = jsonstr.find('administrative')
    if begin == -1:
        return None
    else:
        bb_long = jsonstr[begin:end]
        display_name = bb_long[bb_long.find("display_name"):]
        print(display_name)
        bb_short =
bb_long[bb_long.find('boundingbox')+12:bb_long.find('boundingbox')+93]
        return GetElevation(bb_short,mun_query)

data = []
with open("municipalities.csv", 'r', newline='') as csvinput:
    munreader = csv.reader(csvinput, dialect='excel')
    for row in munreader:
        print(row)
        elevation_list = GetCoords(row)
        with open("output200.csv", 'a', newline='') as csvout:
            wrtr = csv.writer(csvout, delimiter=';', dialect='excel')
                wrtr.writerow(elevation_list)
```