STOCKHOLM SCHOOL OF ECONOMICS
Department of Economics
5350 Master's thesis in economics
Academic year 2015 - 2016

Upstream land-use and downstream water
Using water treatment costs to value ecosystem services

Nils Westling (40603)

January, 2016

Abstract

Water purification is commonly referred to as an economically valuable ecosystem service. Yet, the effect of upstream land-use on downstream water quality is poorly understood, and the economic implications even less so. With panel data of upstream land-use, raw water quality and chemical costs for 76 municipal water treatment plants in Sweden, this MSc thesis studies how land-use affects water quality and, consequently, water treatment costs. The findings suggest that land-use affects levels of E. coli (a microbiological pollutant) in downstream water, but the effect of E. coli on treatment costs remains unclear. Instead, turbidity of water is found to increase treatment costs, but no significant effect of upstream land-use on turbidity is found, possibly because of limited data quality. Whereas a recently published study (Vincent et al., 2015) provided econometric evidence that upstream forests reduce water treatment costs, it was only implicitly assumed that water quality was the channel for this effect. By adding water quality data to the analysis, endogeneity concerns are addressed and a better understanding is gained of how ecosystems create value by affecting water quality.

Keywords: Ecosystem services, Drinking water, Pollution, Monetary valuation, External costs, GIS
JEL: D62, Q25, Q51, Q53, Q57

Supervisor: Johanna Wallenius
Date submitted: January 2, 2016
Date examined: January 13, 2016
Discussant: Edward Lindquist
Examiner: Chloé Le Coq

Acknowledgements

I would like to express my sincere gratitude to my supervisor, associate professor Johanna Wallenius, for helpful advice and for never being further than an e-mail away.

Furthermore, I am very grateful to professor Stephan Köhler, biogeochemist at the section for geochemistry and hydrology of the Swedish University of Agricultural Sciences in Uppsala, who helped me grasp some of the complexities of land-use and water quality; Viktor Kärnstrand, who showed me how python-coding could turn chaos into order and Hillevi Eklund, who taught me the basics of Geographic information Systems (GIS); Per Strömberg, environmental economist at the Swedish Environmental Protection Agency, who was always available for discussions; Rickard Sandberg, associate professor and center director at the Center for Economic Statistics and Ranjula Bali Swain, visiting professor to the Mistra Center for Sustainable Markets (MISUM), both at the Stockholm School of Economics, for acting against all incentives and taking time to discuss the complexities of panel data analysis with me. All errors are my own.

Contents

1 Introduction

> *"The things which have the greatest value in use have frequently little or no value in exchange; on the contrary, those which have the greatest value in exchange have frequently little or no value in use. Nothing is more useful than water: but it will purchase scarcely anything; scarcely anything can be had in exchange for it."*
>
> *Adam Smith (1901)*
> *An Inquiry Into The Wealth Of Nations*

This is a study about clean freshwater and about the value of the ecosystem services that help provide it. More specifically, the aim of this thesis is to study if and how the value of the natural processes that bring reasonably clean surface water to the intakes of municipal water treatment plants in Sweden can be econometrically inferred, by studying the treatment costs in these plants. Assuming that cleaner raw water is less costly to treat, it is quite intuitive that land-uses from which the run-off water is clean, do indeed provide a valuable service to water treatment plants and, hence, for society at large. This study explores whether the monetary value of such ecosystem services can be quantified by studying them as intermediate inputs into the water treatment plants' production processes. A more thorough discussion on the concept of ecosystem services is provided in section 2.1.

A safe supply of clean water is fundamental to human prosperity, and was highlighted as one of the UN's sustainable development goals, adopted by world leaders in the fall of 2015.[1] Understanding the ecological processes that determine water quality is key to ensuring a safe supply of clean drinking water, but environmental scientists are better suited than economists for that work. However, economists can apply their knowledge of firm and consumer behaviour to understand how these ecological processes create value and how this value can be included in decision-making and economic theory (Atkinson, Bateman, and Mourato, 2012).

The value of nature, and of the natural processes that are beneficial to society, is the key focus of environmental economics. For a long time, environmental economists have built a foundation of knowledge about the causes of environmental degradation (Dasgupta, 2008), its costs to society (Holmes, 1988; Dearmont, McCarl, and Tolman, 1998; Mjelde et al., 1984) and possible cures (Arrow et al., 1996; Daily et al., 2000; Jones and Vossler, 2014). The conceptual understanding of environmental resources as public goods (Hardin, 1968) and their degradation as a consequence of un-internalised external costs (Pigou, 1924) is widely accepted and a rapidly growing body of empirical research is increasingly filling these concepts with meaning and content. The increasing popularity of the concept of ecosystem services has changed the focus from estimating the costs of environmental degradation to assessing the benefits of nature, whether it is degraded or not. Many studies, the current one among them, aim to infer a monetary value of environmental goods and services for which market prices do not reflect the benefits to individuals and society, so-called *valuation studies* (see for example Gallai et al. 2009; Ghermandi et al. 2009; Byström 2000; Vincent et al. 2015). This area of research is highly policy relevant, as many societies are making it a key priority to address market failures in order to halt environmental degradation and to find ways of living up to the promise of *sustainable development* (Brundtland et al., 1987).

Traditionally, many valuation studies have used so-called *stated preference* and *revealed preference* methods. Since not least stated preference methods, which depend on survey-based data on stated willingness-

[1] The goals on the UN's website: http://www.un.org/sustainabledevelopment/sustainable-development-goals/

to-pay for different hypothetical environmental goods and services, have been criticised for being unreliable (Kahneman and Knetsch, 1992; Daily et al., 2000), there has recently been calls for research using other methods. Production and cost function based methods combine firm-level production and cost data with environmental quality data to study how environmental quality affects consumer and producer surpluses. In their meta-study on valuations of ecosystem services related to watersheds in the Nordic countries, Barton et al. (2012) explicitly call for more studies using the production cost approach. This study responds to that call, as its aim is to estimate the effect of upstream land-use on downstream water treatment costs. This is done in two stages, where first the effect of land-use on water quality is estimated and then the effect of water quality on water treatment chemical costs.

Four different types of land-use are in focus in this study; forest, logged forest, agricultural and urban areas — and two indicators of water quality; turbidity and E. coli. The first, turbidity, is a measure of suspended solids in the water, which has been used in previous studies on the effects of water quality on treatment costs (Dearmont, McCarl, and Tolman, 1998; Moore and McCarl, 1987; Holmes, 1988). The second water quality parameter, E. coli, is a microbiological pollutant and a commonly used indicator of pathogens, which is highly relevant in a water treatment context because of the current focus on mitigating the risk of disease outbreaks. In 2010, a parasite known as Cryptosporidium made almost half the population, approximately 27,000 individuals, of the Swedish town Östersund ill after it had been spread through the public water supply (Widerström et al., 2014). After this event, several Swedish municipalities have made or are considering investments in more modern, and often very expensive, water treatment techniques.

A recently published study that uses the cost function method to value ecosystem water purification was conducted in Malaysia by Vincent et al. (2015). This study, the first of its kind according to the authors, used Geographic Information Systems (GIS) based land-use data and monthly data on short-run costs from treatment plants in the Perak region of Malaysia to estimate the effect of land-use on treatment costs. The study reported robust evidence that forests, especially virgin forests, in a water treatment plant's catchment area reduced the plant's treatment costs relative to other upstream land-uses. The current study employs a similar approach, but adds to the literature by studying water quality as the intermediate link between land-use and treatment cost. Besides providing relevant information about how land-use affects water quality, this approach also addresses an econometric concern in the Vincent et al. study. Since water quality may not only affect treatment costs directly but also the quantity of water that can be treated, water quantity is possibly an endogenous explanatory variable in their model.[2] By explicitly controlling for water quality in the cost function specification, such endogeneity concerns are mitigated.

The current study is, to my knowledge, the first cost function based valuation study of ecosystem water purification services in Sweden. Empirical results, from random effects as well as generalised method of moments (GMM) estimations, indicate that forests reduce levels of E. coli in downstream water, but the results can not be considered as very robust. No significant effect of upstream land-use on the turbidity of water is found in this study, possibly because data on turbidity levels is too incomplete. In the second stage of the analysis, where the effect of water quality on treatment costs is analysed, it is found that the turbidity of raw water has a significant effect on chemical costs, whereas the effect of E. coli remains unclear.

It is possible that low data quality is hampering inference of the effects studied in this paper. The water quality data is an unbalanced panel dataset, something that can only partially be compensated for by imputations. Another part of the dataset was created by the author, using Geographic Information Systems (GIS). Combining maps from several different sources, data was obtained for the exact locations

[2] Vincent et al. (2015) are aware of this concern and use lagged values of water quantity as an instrumental variable to address it.

that were studied. The availability and quality of GIS data is, however, quite limited, which could be one reason why the results are not more robust. Publicly available GIS data is currently becoming better as well as more prevalent, which implies that future research on the effects of land-use will possibly be able to benefit from higher precision maps.

Besides adding to the body of empirical research on ecosystem service valuations, this study provides some input to the methodological development in the area of water purification service valuations. Nevertheless, more research is clearly needed in order to understand how ecosystems affect the quality of drinking water, in Sweden and elsewhere, as well as how land-use matters.

The rest of this thesis is structured as follows: Section 2 presents the backdrop to the study and describes the concept of ecosystem services as well as how drinking water provision is governed in Sweden. Section 3 provides an overview of the academic literature about the economics of ecosystem services, with a focus on water purification services, and how monetary valuations of these services can be used to enhance decision making and complement national accounts. Section 4 specifies the research questions and in section 5 the method is described as well as the data collection process. The results of the empirical analysis are given in section 6 and section 7 concludes.

2 Background

In this section I will briefly explain the concept of ecosystem services and discuss the policy relevance of deepening the understanding of how ecosystems contribute to the economy and how this can be measured. I will also give some background information about how public drinking water provision is governed in Sweden.

2.1 What are ecosystem services and why should we care about their value?

The concept of ecosystem services is about acknowledging that not all of what is produced in an economy can be attributed to the productivity of human beings. For example, even the most talented and hard-working farmer will find that he or she is in trouble if pollinators don't do their job. Even if pollination is a service that is given to farmers for free, it can not be considered as lacking value.[3] The reason why ecosystem service valuations have become the object of so much research (see section 3), is that environmental economists increasingly argue that our understanding of the economy is only partial if we ignore the production of ecosystems (Dasgupta, 2008).

The Millennium Ecosystem Assessment (UN, 2005), following an initiative launched by the UN secretary-general Kofi Annan in 2000, was hugely influential in putting ecosystem services on the political agenda and making the concept more known and used. The project mapped those ecosystem services that most substantially enhance human well-being, evaluated their status and assessed if they were being used sustainably.[4] It was found that 15 out of the 24 services that were studied were not being sustainably used and that the service in focus of the present study, freshwater purification, was among those that were used in an unsustainable way. Since clean freshwater is a life necessity, it must be considered highly relevant to find ways of ensuring the sustainability of the ecosystem services that sustain it. The fact that clean water was highlighted as one of the UN's 17 sustainable development goals, adopted by world leaders in the fall of 2015, ensures that it will remain a political priority.

Another important contribution of the Millennium Ecosystem UN (2005) was that it categorised ecosystem services in the way that is most often used today. These categories are *provisioning services* (such as the provision of freshwater), *regulating services* (such as the purification of water), *cultural services* (such as a breathtaking view) and *supporting services* (the underlying processes that support the entire ecosystem, such as the nutrient cycle).

A much cited example of how ecosystems provide services, which is highly related to this current study, is highlighted by Chichilnisky and Heal (1998) and regards the water provision in New York City. According to their paper, the city was facing a multi-billion-dollar cost of a new filtration plant in order to reach the Environmental Protection Agency's (EPA) water standards, but avoided the cost by restoring the upstream ecosystems in the Catskills watershed, north of the city. The authors have been criticised for not correctly describing the actual situation in New York (Sagoff, 2002), but the case is mentioned here as an illustration of the type of effect that could possibly exist.

Most water runs great distances through various types of landscapes before it is collected into a water treatment plant in order to become drinking water. It is therefore highly complex to sort out exactly how different ecosystems affect water quality and how their respective characteristics and locations matter. Even though previous studies have shown that forests purify water and that agricultural land-uses reduce water quality (see for example Tong and Chen 2002; Vincent et al. 2015), they have only roughly accounted for the heterogeneity of forests and agriculture.

[3] For an actual assessment of the monetary value of all pollination in agriculture worldwide, see Gallai et al. (2009).

[4] Sustainability is defined, in this UN report, as current use that is not jeopardising future use, aligned with the definition in Brundtland et al. (1987).

Ecosystem services are, by their very definition, beneficial for people.[5] The benefits can, however, be direct or indirect and values can be for use or non-use of the services. In the field of environmental economics, polar bears are commonly referred to in the context of non-use values. Clearly, there are people who would be willing to pay something to save the polar bears, even though they have never seen one in life and don't intend to change that — only the knowledge that there are polar bears in the world is a source of utility for these individuals. It could be argued that clean water can also be valued for its existence and not only its use. However, the methodological approach in this study — valuing ecosystem services based on how they enter a firm's cost function — can only capture use values.

2.2 Drinking water in Sweden

With almost 100,000 lakes, freshwater is abundant in Sweden. There are approximately 1900 treatment plants that produce drinking water for the public, and most of them are run by municipalities or municipally owned firms. In some cases, several municipalities share the ownership of a firm that produces drinking water for their inhabitants. Many of the firms that run water treatment plants also run the wastewater treatment and some are also in the energy business. Consumers pay for their water use, but prices are set by municipalities not on a marginal cost basis, but to cover all costs of water delivery as well as wastewater treatment. Prices can thus not be expected to convey much information about treatment costs.

In this paper, I have only studied treatment plants that treat surface water and thus all water quality data refers to surface water as well. The majority of water treatment plants in Sweden treat groundwater, but many of these plants are very small. In fact, half of the population drink treated surface water, including the population of Sweden's three largest cities — Stockholm, Gothenburg and Malmö. The quality of the surface water is thus of high importance for a large share of the population, and it should be kept in mind that the supply of drinking water is also an issue of public safety. Large-scale outbreaks of waterborne disease are uncommon, but as late as November 2010 approximately 27,000 inhabitants of Östersund were affected by disease after a parasite known as *Cryptosporidium* was spread through the public water supply from one of the city's treatment plants (Widerström et al., 2014). Since then, the Swedish water treatment industry has had a strong focus on how to minimise the risk of such events and several municipalities are considering costly investments in modern treatment techniques.[6]

2.3 Policy relevance

> "In the quest to increase GDP, we may end up with a society in which citizens are worse off."

This quote is from the book *Mismeasuring our lives* by Stiglitz, Sen, and Fitoussi (2010), probably one of the best examples of the political importance of the discussion around welfare measures and the role of GDP in recent years. The book is the result of a project initiated by the former French president Nicolas Sarkozy, who in 2008 asked Nobel prize winners Joseph Stiglitz and Amartya Sen as well as the French economist Jean-Paul Fitoussi to form a commission and write a report about the shortcomings of GDP and propose how to produce more relevant indicators to measure social progress.

The ambition to include the value of ecosystem services in decision-making is becoming increasingly important as a theme in policy around the world. A report demanded by the Swedish government in 2013 proposed several reforms to visualise and integrate the value of ecosystem services and the EU's so-called *Biodiversity strategy* has as one of its objectives to assess the economic value of ecosystem services and

[5] Ecosystem Services are defined in The Millennium Assessment as "the benefits people obtain from ecosystems".

[6] For the purpose of this study, two study visits were made at water treatment plants, to provide insight into the treatment process. During these visits, conversations were held with plant technicians and issues of investments in new treatment technique were discussed. The study visits were made in Järfälla and Västerås during the fall of 2015.

integrate these into accounting and reporting systems at EU and national level by 2020 (Barton et al., 2012).[7] On a global scale, there is the UN's System of Environmental-Economic Accounting (SEEA), which is a satellite system to the System of National Accounts (SNA). Studies of ecosystem service values are clearly necessary to underpin these policy goals.

[7]The Swedish government report, in Swedish, can be found at: http://www.regeringen.se/contentassets/ba53cd9f18b74f348eb0ff31e8280d60/ synliggora-vardet-av-ekosystemtjanster-sou-201368. The EU's biodiversity strategy can be found at http://ec.europa.eu/environment/nature/biodiversity/comm2006/2020.htm

3 Literature overview

The purpose of this study is to econometrically infer the value of the ecosystem service of water purification, measured as an input to the production of drinking water in municipal water treatment plants in Sweden. In order to guide the reader through the relevant research fields, environmental and ecological economics, and to position the current study in these fields, I will use this section to introduce some basic concepts and, thereafter, give a brief overview of the body of environmental valuation studies made to date. I will specifically focus on different conceptual approaches to environmental valuation studies, introduce the cost function method and present some relevant empirical results. The literature will be related to the current study along three different dimensions: (i) Methodological choices, where the focus has often been on so-called *stated preference* methods but where calls for research have emphasised production/cost function approaches; (ii) the different ways in which environmental valuation studies are used and integrated into economic thinking; and (iii) the emphasis on some specific ecosystem services in previous empirical research — and where there are gaps.

3.1 Basic concepts in environmental economics

The field of environmental economics is to a large extent based on the concept of external costs: when an economic actor pollutes the air, the water or the ground without having to pay for the damage he or she is inducing on others, the cost is spread out among all those who use the air/water/ground, hence *external*. Making the polluter pay for the full cost of his or her actions (i.e. *internalising* the cost) requires that the external costs of pollution and other environmental degradation are assessed, which has often been the objective of research in environmental economics. The pioneer behind the concept of external costs was A.C. Pigou, who pointed out situations where, in his words, "marginal trade net product is greater than marginal social net product" (Pigou, 1924).

Another important concept in environmental economics is that of public goods, most famously framed by Hardin (1968) in his hugely influential *The Tragedy of the Commons*. The title refers to a pasture open to all herdsmen of some area (a commons), where they can let their cattle roam. The commons is a public good, which is why the herdsmen, being rational and seeking to maximise their own gain, will have very weak incentives to abstain from adding another animal to the herd even if there are so many cattle in the pasture that it is being overgrazed. In Hardin's words: "Each man is locked into a system that compels him to increase his herd without limit — in a world that is limited. Ruin is the destination toward which all men rush, each pursuing his own best interest in a society that believes in the freedom of the commons" (Hardin, 1968, p. 1244). Lakes and streams are clearly public goods and, even though there are regulations about what people can and cannot do with Swedish water bodies, property rights are absent when it comes to the water that runs through the Swedish landscape. Hardin's concept is thus very relevant in the context of this study.

3.2 The value of the environment and the methods to find it

Dasgupta (2008) relates Hardin's claim about public goods to the concept of ecosystem services by saying that "property rights to natural capital are often either vaguely defined or weakly enforced, meaning that nature's services are underpriced in the market" (p. 2). Whereas Hardin and Pigou focused mainly on the costs of environmental degradation, Dasgupta highlights the benefits of the services that nature constantly provides. These are mirror images, and the true cost of environmental degradation must reflect the benefit that is lost. This is why much research in environmental economics has been focused on understanding the types of benefits people enjoy from the environment, and it is the reason why so-called valuation studies have become an increasingly important part of environmental economics (Atkinson, Bateman, and Mourato, 2012).

Several methods have been used in environmental valuation, all with their specific advantages and drawbacks. In their thorough literature review of the field of ecosystem service valuations, Atkinson, Bateman, and Mourato (2012) list the most commonly used methods for environmental valuation as being *stated preference methods*, *revealed preference methods*, *adjusted market prices* and *production function methods*.

In valuation studies with stated preference methods — many times a method called contingent valuation is used — respondents are typically asked about their willingness to pay for some non-market good. Since respondents are asked about a hypothetical payment, their responses might differ from payments that they would actually make — hence the term *hypothetical bias*. A related critique has been put forward by Kahneman and Knetsch (1992) who claimed that the validity of contingent valuation methods suffered from respondents' insensitivity to scope. For example, Kahneman and Knetsch found that residents of Toronto expressed a willingness to pay to prevent the drop in fish populations in all Ontario lakes that was only slightly higher than the willingness to pay to preserve the fish stocks in only a small area of the province. They call this problem the *embedding effect*. In a meta-analysis of 28 stated preference valuation studies, Murphy et al. (2005) confirm that biases exist, but argue that they can be mitigated by study design and calibration techniques.

Two well-known revealed preference methods are *travel cost* and *hedonic pricing*. In a study using the travel cost method, a lower bound of the value of a nature area would be inferred by observing what visitors are willing to pay to travel there (assuming that they get no utility from the travelling experience in itself). An example of a study using hedonic pricing, which is relevant in the context of the current study, is given by Leggett and Bockstael (2000). They study how waterfront real estate prices in a county in the Chesapeake bay on the US east coast are affected by local water quality, and find that an increase in fecal coliform bacteria in the water significantly depresses property prices. This is thus another way to estimate the value of the ecosystem service studied in the current paper and it shows how different methods value different aspects of value. Even in the study by Leggett and Bockstael, the estimated value of higher water quality should be seen as a lower bound since, for example, the benefit of visitors to the area is not included.

Clearly, there is a range of problems and limitations in ecosystem valuation studies. A few of those are mentioned by Daily et al. (2000), who claim for example that contingent valuation surveys, in which individuals are asked to value hypothetical environmental changes, are "improving but still notoriously unreliable" (Daily et al., 2000, p. 4). The authors mention that all methods that rely on individual preferences to construct social values, i.e. stated and revealed preferences, are heavily influenced by how much individuals know about the environment, hence the outcome is not more informed than the people whose values are being assessed. However, Daily et al. remind the reader of an important point: even market prices are flawed as an indication of value, if externalities mean that prices are not reflecting the full social costs of production. To conclude, no valuation method is perfect and the best choice depends on the object and purpose of the specific study. The plethora of valuation techniques probably reflects the complexity of valuing environmental change.

3.2.1 Production/cost function based methods

One way to infer the value of environmental goods and services, and the method that will be deployed in the current study, is to identify situations in which these are factors in the production function of a market good. In these cases, estimated production functions or cost functions can be used to link the physical effects of environmental change to changes in market prices and quantities and ultimately to producer and consumer surpluses.

One example of such a study is given by Mjelde et al. (1984), who use data on ambient ozone levels (a pollutant known for its negative effect on crop yields) and firm-level data from crop-producing farms in the state of Illinois, US, to estimate a production function with air quality as one of the arguments besides capital and labor, assuming that farmers adjust the levels of inputs based on observed yield in order to maximise profit. They find that, at the average level of ambient ozone, a 1% increase in ozone leads to a 0.408% decrease in profits. They find that total cost of ozone pollution was USD 226 million for crop producing farmers in Illinois in 1980.

Neeliah and Shankar (2010) use a similar framework to study the effect of ozone on crop yields in the UK, adding to the analysis that farmers may adjust both inputs and outputs (choice of crop) in response to observed yields. They use a panel dataset of 66 UK cereal farms over 6 years and include fixed effects to calculate profit elasticities with regards to ozone levels. Their findings suggest that a 10% increase in ozone, above its mean value, would lead to a 1.3% decrease in variable farm profits (significant at the 10% level).

Daily et al. (2000) claim that methods based in cost functions (such as "valuing natural water purification at the cost of its technological alternative, a filtration plant for instance", p. 4) provide only partial, lower bound indications of value, because it does not take into account the value of e.g. clean lake water for swimming in the summer or the improved habitat for fish. This must clearly be true in the case of ozone and agricultural production as well, since low ozone levels must also be valued for its health implications, for example.

A range of studies have used cost function based methods to study the costs of water pollution (Dearmont, McCarl, and Tolman, 1998; Moore and McCarl, 1987; Holmes, 1988) and the study by Vincent et al. (2015) uses a similar approach to value upstream ecosystem services. These studies will be treated more thoroughly in section 3.5.1, below.

3.3 Taking the value of the environment into account

As a consequence of external costs and the public good characteristic of many environmental benefits, market forces cannot be expected to lead to the most economically efficient use of environmental resources. That is why environmental valuation studies often come to be used as information inputs into decision making regarding resource and environmental management. If the monetary value of the environmental benefits of, say, a wetland area are greater than the economic value of the same area being drained and used for some other purpose, then social welfare is clearly not maximised if the area is drained. According to A. M. I. Freeman, Herriges, and Kling (2014, p. 10), if the objective of policy-making is to maximise net social economic value, then benefit-cost analysis becomes the tool to reach optimum management and decision-making becomes almost mechanical. Clearly, this is a very idealised view of decision-making and the process of correctly estimating marginal benefit and marginal cost curves remains a challenge.

Arrow et al. (1996) argue for the use of benefit-cost analysis as one of the fundamental criteria for evaluating proposed regulations in such fields as environment, health and safety. They highlight an example of a drinking water regulation that the US Environmental Protection Agency (EPA) implemented in 1979 that was assessed to save lives at a cost of USD 200,000 per life saved. Compared with another regulation regarding wood-preserving chemicals that was estimated to save lives at a cost of USD 6.3 trillion, benefit-cost analysis should tell policy-makers to make the drinking water regulation stricter and the one regarding wood-preserving chemicals less strict (given that both regulations are in place primarily to save lives). Not only does benefit-cost analysis illuminate such trade-offs in policy-making, it can also tell decision-makers how much regulation is "enough".

An illustrative example of a location-specific benefit-cost analysis is given by Barbier (2007) in a study where the economic value of mangroves in Thailand is estimated and social values for services such as reduction of coastal flooding are compared with private values if the mangroves are converted for shrimp farming. Whereas private profits double from around USD 600 to USD 1,200 if a representative hectare of mangrove is converted, the social value of a conserved mangrove is estimated at around USD 12,000.

This is a situation where an environmental valuation integrated in a benefit-cost analysis would lead to a more economically efficient outcome.

On a micro-level, valuation studies of the environment and its services could be used as an aid in decision-making and public policy. But it could also play a role on a much larger scale, in national income statistics. GDP is by far the most commonly used statistic to evaluate the health of an economy, but it is no surprise to economists that GDP is a rather imperfect measure of economic performance (Dasgupta, 2009), and only a crude proxy of human well-being (see for example Deaton 2007 for a discussion about how GDP correlates with self-reported life satisfaction and health).

The value of the environment and the goods and services it provides may seem negligible in official statistics on national income, such as GDP. That is because GDP only reflects the value of production at market prices and, as previously discussed, markets tend to be non-existent for most environmental goods and services. Consequently, argues Dasgupta (2008), national income statistics would look quite different if shadow prices for ecosystems and ecosystem services were included. This relates to the current study in the sense that when treatment plants pay for chemicals, this adds to GDP, but if ecosystem services substitute chemicals, this is not accounted for.

3.4 Ecosystem service valuation

From an economic point of view, ecosystem services can be any contribution of the natural world to the generation of goods and services which people value. This can be physical products (such as food) as well as services that generate use values (such as recreational areas that people can visit) and even non-use goods that are valued simply for their continued existence (use and non-use values are introduced in section 2.1).

Economic valuations of ecosystems require some understanding of the natural (biological, ecological, chemical or physical) processes that are in play. For this, natural scientists are irreplaceable. They can explain *how* upstream land-use affects water purification or *why* a wetland might be helpful in reducing eutrophication of a lake or the sea. But to put an economic value on these ecosystem services, we need also to understand how these services come into play in peoples' utility functions and firms' cost or production functions. Atkinson, Bateman, and Mourato (2012) gives the somewhat obvious example that the recreation value of a woodland depends on where it is located, since it is likely to give rise to higher social value the more proximate it is to a potential visiting population and the larger the amount of visitors the higher is the aggregated utility. The amount of economic valuation studies on ecosystem services has grown "at an almost exponential rate" during recent years (Atkinson, Bateman, and Mourato, 2012, p. 22) and a few contributions will be mentioned in the following paragraphs.

An especially illustrative ecosystem service is pollination. Compared with e.g. the nutrient cycle, which just passively goes on in the background, pollination can more intuitively be understood as a service, with pollinators such as honeybees actively transferring pollen and upholding the productivity of plants. Gallai et al. (2009) estimated the value of pollination for agricultural production worldwide and found the total economic value to be of the order of EUR 153 billion (quite clearly, a number that should be taken as indicative).

Costanza et al. (1997) aimed to estimate the value of 17 ecosystem services for the entire biosphere. Using previously conducted valuation studies, they estimated an average value of a certain ecosystem service per biome, then multiplying the average value with the total area on the globe of that specific biome.[8] They claimed that the value of the ecosystem services studied were in the range of USD 15-64 trillion per year, with an average of USD 33 trillion (global gross national product was around USD 18 trillion per year at

[8] Biomes is a very rough classification of the earth into different types of areas. In Costanza et al. (1997) the planet consists of 16 biomes, such as *desert*, *open ocean* and *tropical forest*.

the time). Even though the authors stressed the fact that their study only gave a "crude initial estimate" of the total value of ecosystem services, the study was heavily criticised. One example of this critique comes from Bockstael et al. (2000). They approve of the general ambition of Costanza et al. — showing the aggregate value of the natural world — but argue that the approach makes little sense. Whereas Costanza et al. seem to talk about the economic value of ecosystems as if the choice were between having them as they are or not having them at all, Bockstael et al claim that "economic value is about tradeoffs and as such requires defining the alternatives clearly". In their view, ecosystem service valuation must be about evaluating well-defined changes to ecosystems. The approach of the current study is in line with this view, as the purpose is to find how water quality and treatment costs are affected by marginal changes of different upstream land-uses.

3.4.1 Ecosystems as natural assets

Several authors, including Daily et al. (2000), Dasgupta (2008) and Barbier (2011), describe ecosystems as *natural capital* that, if properly managed, yields a flow of services such as the production of goods (e.g. timber), life-fulfilling conditions (such as beauty and serenity) and life supporting processes (e.g. water purification). When human intervention degrades an ecosystem, the effect on human well-being can have an inter-temporal dimension: there is a potential change in peoples' future prospects and the change in the future flow of services should affect the value of the natural capital asset (Atkinson, Bateman, and Mourato, 2012). Natural capital is thus comparable to other types of capital in that the net present value of the stock reflects current and expected benefit flows.

Barbier (2011, p. 234-239) adopts this view and claims that natural, physical and human capital together determine the economic opportunities that are available and obtainable for present and future generations. As any capital owner or manager, society must decide how to manage this capital stock so as to maximise current benefit flows as well as the future value of the stock. It may, however, not necessarily be the case that only the aggregate stock matters, but also the composition. The question is thus, according to Barbier, whether the three types of capital are perfectly substitutable.

Dasgupta (2008) uses the same framework with three types of capital, but claims that natural capital is different from human and physical capital in the way in which it depreciates. Whereas all three types of capital depreciate if they are misused or overused, natural capital (i) can depreciate irreversibly, (ii) might not be replaceable and (iii) can collapse abruptly. The view of ecosystems as natural capital assets leads to an interesting interpretation of the word *sustainability*: for any act that harms ecosystems, and hence depletes natural capital, the same act must lead to the build up of physical or human capital with the same value or larger — otherwise the act is making society poorer and development is not sustainable. For this reason, Dasgupta claims that standard national accounting measures will not detect if economic development is sustainable or not, since the value of services obtained from natural capital are missing in standard economic accounts.

The concept of ecosystems as natural capital, besides physical and human capital can be linked to the current study, as I will examine the cost at which the water purification service provided by ecosystems in Swedish landscapes (natural capital) can be substituted by water purification in treatment plants (physical capital).

3.4.2 Geographical variability of ecosystem services

An important aspect of environmental valuation is that location is likely to matter. For this reason, Bateman et al. (2002) argued for the use of Geographic Information Systems (GIS) in environmental and resource economics. In their view, spatial relationships are often significant elements of the problems analysed in environmental economics and GIS can contribute to the sophistication of such analyses. One example where the spatial relationships are given much consideration is provided by Polasky et al.

(2011), who use a GIS-based model to study the economic effects of five different land-use scenarios for the state of Minnesota, USA, comparing returns to land-owners from timber and agriculture production as well as urban development with estimated social values of water quality, carbon sequestration and habitat quality. With the spatial precision of the model they were able to integrate the local returns of agriculture and the flows of agricultural run-off to local pollution levels of the Mississippi river. They find that an agricultural expansion is the scenario that generates the highest private returns but is also the scenario that generates the lowest social benefit, due to declines in water quality and carbon storage as well as declines in habitat quality. The spatial sophistication of the study does not only lead to better understanding of local environmental changes, but also means that trade-offs between different scenarios are more credible.

For many ecosystem services, location close to population centers is an important determinant of value. Brander and Koetse (2011) conduct a meta-analysis of a number of valuation studies focused on open space in urban areas. They include results from studies using two different valuation methods, contingent valuation and hedonic pricing, and find that population density increases open space value, regardless of the valuation method. In a similar meta-analysis for wetlands, Brander, Florax, and Vermaat (2006) also found a significant relationship between population density and the value of wetlands. These findings are intuitive, since ecosystem services are valued on the basis of their contribution to human welfare, hence their value should increase as more individuals can benefit from them. Quite clearly, the value of ecosystems that purify water should be greater the larger the population is that can benefit from the cleaner water.

Another characteristic of ecosystems that is specifically important in densely populated areas is resilience to shocks. Safe water provision is important to study in this context, since e.g. droughts and waterborne disease outbreaks are threats that grow with the population density. The contribution of ecosystems to resilience against such shocks is sometimes referred to as *insurance value* (Gomez-Baggethun and Barton, 2013). Even though it is beyond the scope of the current study, the drinking water purification of upstream ecosystems could clearly be evaluated in an *insurance value* framework. If a waterborne disease is spread through the public water supply of a larger city, like the case in Östersund 2010 (Widerström et al., 2014) discussed in section 2.2, the costs can be enormous and the value of minimising the risk should reflect that.[9] Hence, if ecosystem services help protect water bodies from microbiological pollution and pathogens, they should be valued for the risk reductions they provide. Clearly, this value should reflect the amount of people that depend on the specific water body that is protected. Consider the lake Mälaren, west of Stockholm, which supplies drinking water to approximately 2 million people only in the greater Stockholm area (Ledesma, Köhler, and Futter, 2012) and to several other cities as well. Although difficult to estimate, the value of keeping this lake safe from health hazards can not be ignored.

One important objective of the current study is to find which types of land-use that correlate with water quality. Tong and Chen (2002) might give some insight about what to expect. They used a GIS-based model to study how agricultural, residential, commercial and agricultural land-uses correlated with a number of water quality variables in a watershed in Ohio, USA. They found that nitrogen, phosphorus and coliform bacteria were positively correlated with commercial, residential and agricultural land-uses, but negatively correlated with forests. However, the effect of land-use on water quality is complex and not only does type of land-use matter but also the location of the land-use relative to the lake or stream where the water quality is measured. A common theme in the water quality literature is *riparian zones*. These are zones along the water bodies that, under certain conditions and with the right land-uses, can stop pollutants, not least nutrients, from reaching the water (Lowrance et al., 1997). According to Sliva and Williams (2001) there is an ongoing scientific dispute about whether land-use of the entire catchment

[9] An even more catastrophic example is from Milwaukee, Wisconsin where, in 1993, an estimated 400,000 people became ill and more than 100 died after Cryptosporidium had passed through one of the local water treatment plants and entered the drinking water supply (Mac Kenzie et al., 1994).

area or only that of the riparian zone is a better predictor of water quality. With a statistical and GIS-based approach they found that catchment area land-use predicted water quality slightly better than what buffer zones did, but the results did not give any clear answers. According to their results, the run-off from urban land is the most important source of water pollution, whereas forests are negatively correlated with several pollutants. The effect of agricultural land-use was not found to be strong, which contradicts findings from several other studies.

3.5 The value of water

Valuation studies of ecosystem services regarding water are quite common in the literature. An especially recurring object for valuation studies is wetlands, a likely explanation being that wetlands is a type of nature that provides a broad range of ecosystem services but also that wetland areas have been declining quite dramatically — according to Carlsson, Frykblom, and Liljenstolpe (2003), more than 90% of the original wetlands in southern Sweden have been eradicated to make place for urban and agricultural land-uses.

The declining amounts and areas of wetlands have in some countries been countered by political instruments to construct new wetlands. Ghermandi et al. (2009) provide a meta-analysis of wetland valuation studies, comparing the values of natural and constructed wetlands. They do find higher monetary valuations for constructed wetlands, which might not be very surprising considering that these wetlands were, most likely, constructed in some specific location in order to carry out some specific biological or ecological service (unlike natural wetlands that are spread out across the landscape in a more random fashion).

In Sweden, one especially important service provided by wetlands is the abatement of agricultural run-off, especially nutrients. The eutrophication problems of the Baltic sea is a likely reason for why wetlands have become the focus of so much research. Byström (2000) used the replacement cost method to estimate the value of wetlands for abatement of agricultural nitrogen runoff to the Baltic Sea. With this method, a value of wetlands was obtained by comparing the cost of using wetlands for nitrogen abatement to the second most cost-effective measure to reduce nitrogen levels. It was shown that, in order to reduce nitrogen run-off from agricultural sources by 50%, wetlands provided substantial cost-savings and the value of this service was estimated at SEK 210 million (Byström, 2000). Gren (1995) also studied investments in wetlands for nitrogen abatement, this time on Gotland, an island off the Swedish east coast, and found that the utility of a marginal investment in wetlands is higher than the utility of a marginal investment in sewage treatment plants. The difference is mainly attributed to (i) the natural growth of wetlands whereas sewage treatment plants depreciate; and (ii) that wetlands also provide other benefits.

3.5.1 Drinkable water

Most water runs great distances through various landscapes before it is collected into a water treatment plant. Even though it is intuitively easy to accept that some ecosystems are better than others at providing clean water, the exact impact of upstream land-use on downstream water treatment is quite poorly understood (Vincent et al., 2015; Tong and Chen, 2002). The importance of this knowledge is often mentioned, however, for example by Davies and Mazumder (2003) who argue that questions of drinking water quality is too much focused on raw water treatment, and that more attention should be given towards watershed management — thus relying on natural capital rather than physical capital to provide clean water (for a discussion on these concepts, see section 3.4.1). This clearly calls for more research on (i) how catchment area characteristics, such as upstream land-use, affects water quality and (ii) the substitutability between ecosystem water purification and regular water treatment. The current study investigates these issues.

A recent development in this field is the study by Vincent et al. (2015), where a panel dataset with water treatment costs and upstream land-uses for 41 water treatment plants in the Perak region of Malaysia is used to derive the monetary value of water purification by tropical forests. With monthly observations of water treatment costs over 14 years, they have enough observations to control for time-invariant treatment plant specific effects as well as time-specific effects. In their statistical specification, land-use can be of three different types — virgin forest, logged forest and non-forest — and the value of both types of forest land-use are estimated relative to all non-forest land-uses.[10]

They find that both virgin and logged forests have a significantly negative effect on treatment costs, the effect being slightly stronger for virgin forests. The marginal value of the ecosystem service is defined as "the annual reduction in treatment cost that results from avoiding the conversion of one hectare of virgin forest to a non-forest land-use" (Vincent et al., 2015, p. 23). This is what the treatment plant should be willing to pay for the conservation of an additional hectare of forest, if the ecosystem service was provided by a market actor. A downward sloping curve is estimated, which implies that the treatment plants should be willing to pay diminishing amounts to protect additional hectares of forest. The contribution of virgin forests, valued at the marginal reductions in treatment costs that they generate, was found to be worth more, on average, than a third of the treatment plants' aggregate expenditures on labor, energy, chemicals and maintenance. This signals that upstream ecosystem services are highly relevant for the production of drinking water in Malaysia.

An econometric concern in the study by Vincent et al. (2015) follows from the fact that water quality is not included in the analysis — it is only implicitly assumed to be the factor linking land-use to treatment costs. Since water quality may not only affect treatment costs but also influence the quantity of water that can be treated in a given time-frame, water quantity, which is used as an explanatory variable in the cost function specification, may in fact be endogenous. To address this issue, the authors use the second lag of water volume as an instrumental variable. Their results are robust to the IV specification and a formal test, Wooldridge's robust score test, does not reject the null hypothesis that water volume is exogenous.

Dearmont, McCarl, and Tolman (1998) studied how water treatment chemical costs were affected by turbidity levels of raw water for a sample of surface water treatment plants in Texas. Like Vincent et al. (2015) they had access to monthly data from the treatment plants on amount of treated water and chemicals used and carried out a panel data regression in which chemical costs per 1000 gallons of water were predicted by the amount of treated water and the turbidity and pH levels, with annual rainfall as control variable. They find that a 1% reduction of the turbidity level can be expected to reduce chemical costs by 0.27%. Extrapolating this result, they estimate that a 1% reduction of turbidity levels across Texas, would save approximately USD $70,000$ annually in chemical costs for plants that treat surface water.

In a similar type of study, Holmes (1988) estimated a water treatment plant's cost function but, in contrast to Dearmont, McCarl, and Tolman (1998), he studied all short-run treatment costs and how they were affected by the turbidity level of the raw surface water. He found a substantially lower cost-turbidity elasticity than Dearmont, McCarl, and Tolman, at 0.07. However, extrapolating the result, it was estimated that suspended solids, measured as turbidity, induced costs of USD 458-661 million across all surface water treatment plants in the US. Holmes (1988) linked the water quality issue to agricultural run-off, but did not provide empirical evidence for this link. Moore and McCarl (1987) studied how the costs of two water treatment chemicals, alum and lime, as well as sediment removal costs, were affected by turbidity levels in the Willamette Valley of Oregon, USA, and reported an estimated elasticity of approximately 0.33. These three studies were all conducted before the concept of ecosystem services gained popularity. However, it should be noted that these *damage function* studies, where the costs of

[10] In the study by Vincent et al. (2015) *logged* forest refers to secondary forest, which has re-grown after timber harvest. Their use of the word *logged* should thus not be confused with the way it is used in the current study, where it refers to land that has recently been logged and is currently in a transitional state.

pollution were estimated, can be seen as a mirror image of the benefits of pollution controlling ecosystems.

A paper which uses an approach very similar to that of the current study has been published by the American non-profit Trust for Public Land (J. Freeman, Madsen, and Hart, 2008).[11] The authors study how upstream land-use affects water quality, as indicated by total organic carbon (TOC), alkalinity and turbidity of the water and how these water quality variables affect treatment costs. Their cross-sectional analysis of data from 60 US water treatment plants show indications that urban and agricultural land-uses correlate with increased levels of turbidity and that forests correlate with lower levels of turbidity and TOC. In this study, turbidity was not found to significantly affect treatment chemical costs, but taken together, the three water quality variables studied correlated positively with chemical costs.

3.6 Contributions of the current study

In this section I have discussed the evolution of valuation studies, its methodological development and its importance in policy-making and national accounts as well as the concept of ecosystems as natural capital. I have also referred to a range of studies about how ecosystems help generate clean water and how the value of this service can be estimated.

It has been found that production/cost function methods have been applied to value ecosystem water purification in previous research, most recently by Vincent et al. (2015). Before that, several studies have investigated the economic costs of raw water turbidity (Holmes, 1988; Dearmont, McCarl, and Tolman, 1998). However, I have found no similar study conducted in Sweden and the only study, to my knowledge, that deploys a combined approach of studying both the effects of land-use on water quality and of water quality on water treatment costs is a non peer-reviewed paper published by the Trust for Public Land, an American non-profit (J. Freeman, Madsen, and Hart, 2008). An advantage of this combined approach is that it addresses the endogeneity concern raised by Vincent et al., by explicitly controlling for water quality in the cost function specification.

Moreover, in their survey of ecosystem valuation studies in the Nordic countries, Barton et al. (2012, p. 52-53) identify a knowledge gap in the area of regulating ecosystem services, such as water pollution control, and a lack of studies using production function based methods. The current study is intended to contribute to filling these identified gaps.

Research is also called for by Atkinson, Bateman, and Mourato (2012) about how ecosystem services are related to spatial variability. Adding to that, Polasky et al. (2011) claim that knowledge of how land-use affects ecosystem service provision and values is a precondition for efficient land-use. The current study is taking steps to contribute to this type of understanding.

[11] It should be noted that the study by J. Freeman, Madsen, and Hart (2008) is not a peer-reviewed article.

4 Research question

This MSc thesis is intended to add to the literature on ecosystem service valuations in the Swedish context, by studying the effect of upstream land-use on downstream water treatment costs. Specifically, it will try to answer the following questions:

1. Does upstream land-use significantly affect pollution levels in downstream water?
2. Does pollution levels in surface water significantly affect the chemical costs of water purification in Swedish municipal water treatment plants?
3. Can the marginal values of the contributions from different types of ecosystems to water purification be estimated by studying their effect on chemical costs of municipal water treatment plants?

4.1 Theoretical predictions

This subsection summarises some results from previous research on the effect of land-use on water quality and the effect of water quality on treatment costs. Most of these studies are treated more thoroughly in the literature review (section 3) but are mentioned here to briefly introduce the expected results.

Forest ecosystems are mentioned in the literature as a positive factor for water quality (Saarikoski et al., 2015; Tong and Chen, 2002; Vincent et al., 2015), hence that is the expected result in the current study as well. Consequently the clearing of forest can be expected to be negative for water quality. It is quite established that removal of forest increases long-term water discharge and leads to higher flows in streams and rivers (Bosch and Hewlett, 1982), which can lead to higher run-off of sediment and pollutants. Costa, Botta, and Cardille (2003) study a large catchment area in Brazil that has experienced a dramatic conversion from forest to agricultural land-use and find significantly higher stream flows (they do not study the effect on water quality). Thus, a negative sign is expected for forest land-uses and a positive effect for logged forests on downstream water pollution levels and, consequently, for treatment costs.

It should be noted, however, that Sweden does not experience the same type of land-use conversion as Brazil and many other parts of the developing world. Even though forestry is an important economic activity in Sweden, the Swedish forest cover is rather constant because logging and re-growth is balanced. The lower rate of land-use change in Sweden is likely to make statistical inference more difficult, which will be discussed in section 5.

Tong and Chen (2002) study the effect of four types of land-uses — residential, commercial, forest and agriculture — on downstream water quality in one watershed in southeastern Ohio, USA. They find, among other things, that agriculture, residential and commercial land-uses significantly increase levels of nutrients and fecal coliform. My dataset allows me to test the effect on water quality of urban land-uses, as well as forest and agriculture. Following Tong and Chen (2002), agricultural land-uses are thus expected to lead to higher levels of microbiological pollution. It should be noted that agriculture and forest constitute a much larger share of the Swedish landscape than urban areas, and the effect of urban land-use might therefore be harder to infer statistically.

Regarding the effect of water quality on treatment costs, Dearmont, McCarl, and Tolman (1998) as well as Holmes (1988) and Moore and McCarl (1987), have estimated treatment cost elasticities to turbidity. All have reported significant effects, but the estimated elasticities range from 0.07 to 0.33. I have not found studies that report how water treatment costs are affected by levels of E. coli or other microbiological pollutants.

5 Method and data

In this section I will describe the methodology used in this study, as well as the data and the data collection process. The econometric specification is in part based on the cost function of municipal water treatment plants in Sweden, which is why the methodology section below will treat the theory of how ecosystem services can be valued as a factor input of production, and then describe how the theory can be applied to the specific situation and the data at hand.

My approach to the valuation of ecosystem water purification in Sweden is to assess how ecosystem services in Swedish landscapes affect the quality of raw water and how this, in turn, affects costs in water treatment plants. The relationship between land-use and treatment costs is thus divided into two different processes that are studied separately. With monthly panel data over 14 years for 76 different water treatment plants and their upstream catchment areas, the aim of this study is to quantify the effect of upstream land-use on two water quality measures, turbidity and E. coli (the first process), and then use these water quality measures as independent variables to study their effect on chemical costs in the treatment plants (the second process).[12]

5.1 The cost function approach to environmental valuation

The theory of how to derive welfare measures for changes in an environmental quality parameter, which enters directly into the production function of a firm, is given by A. M. I. Freeman, Herriges, and Kling (2014). An increase in this parameter, let's call it q, is assumed to increase the output attainable with any given set of inputs — or vice versa, an increase in q decreases the amount of inputs required for the production of a given amount of the final product.

In many instances, valuation with the production function approach is based on the change in producer and consumer surplus that follows after a change in q, which requires knowledge of supply and demand functions. However, no such knowledge is necessary if the following two assumptions can be made: (i) q is a perfect substitute for other inputs of production and (ii) that price elasticity of demand is close enough to zero, so that the change in production costs does not lead to changes in output. If these assumptions are true, then the change in production cost fully captures the value of the change in q, and can be calculated as long as the substitution relationship between q and other inputs of production are known.

The motivation for why the first assumption can be made in the context of raw water quality and water treatment is that the intake water at most treatment plants is always monitored and computer software adjusts the chemical dosage instantaneously when the water quality changes.[13] Hence, an increase in raw water quality will lead to an instant reduction in chemical costs.

Regarding the second assumption, consumer prices for drinkable water in Sweden are so low that price elasticity of demand can be assumed to be close to zero. Höglund (1997) used water price and consumption data from the period 1980-1992 to estimate the price elasticity of demand for Swedish municipal water and found the elasticity to be approximately -0.10 for marginal price changes and approximately -0.20 for average price changes, which is significant but small.[14] Consequently, almost the entire benefit of increased raw water quality will be captured by cost savings in water treatment plants. These cost savings should, in the long run, be transferred back to consumers, since production of drinking water is a municipal responsibility and not-for-profit in Sweden. It should be noted, however, that many municipalities have decided to let publicly owned firms run the water treatment facilities, which may introduce other incentives

[12] The two water quality variables, E. coli and turbidity, are thoroughly introduced in section 5.2.2.

[13] Chemical dosage was one of the issues discussed during the study visits at treatment plants, held in Järfälla and Västerås during the fall of 2015.

[14] The pricing policy varied across municipalities during the study period, but in general kept consumers uninformed about marginal price. This is why, according to the author, elasticity is higher for average price than for marginal price.

than in treatment plants that are owned directly by the municipality.

5.1.1 The water quality model

Acknowledging that I have very much better data, in terms of quality as well as quantity, of water quality measures than of treatment plant costs, the effect of land-use on water quality (the first process) will be studied in more detail than the cost function (the second process). Whereas the cost data from treatment plants is yearly, the dataset on water quality contains observations of a rather irregular frequency. This is because treatment plants are allowed to report their test results whenever they conduct a water quality test. Some of the treatment plants in the sample do this very regularly, others do not. I constructed a dataset with monthly observations, despite quite a few missing values. If more than one water quality test was conducted during a specific month, the average value for that month was used.

The water quality model has the following form:

$$ln(Q_{im}) = \beta_1 ln(L_{iy}) + \beta_2 ln(r_{im}) + \beta_3 ln(w_{im}) + v_i + \theta_{m,y} + \epsilon_{im}$$

where Q_{im} is water quality at plant i, varying by month, L_{iy} is a matrix of land-use measures varying by year, r_{im} is rainfall and w_{im} is total water flow, v_i is time-invariant characteristics that are specific to each upstream area and $\theta_{m,y}$ is a time specific fixed effect that can be month or year specific. Furthermore, ϵ_{im} is an error term and the β are parameters to be estimated. The time subscripts y and m signify whether the variable varies per month or year. All variables are introduced in section 5.2.

5.1.2 The cost function

The second part of this study is based on the chemical cost function of treatment plants. In this case, data is yearly and the model is specified as

$$ln(C_{iy}) = \beta_1 ln(Q_{iy}) + \beta_2 ln(vol_{iy}) + \beta_3 ln(r_{iy}) + \beta_3 ln(w_{iy}) + v_i + \theta_y + \epsilon_{iy}$$

where C_{iy} is chemical costs in treatment plant i during year y, Q_{iy} is a matrix of water quality parameters (year averages) and vol_{iy} is treated water volume. The control variables are total rainfall during the year, r_{iy}, and average yearly water flow, w_{iy}. The panel structure of the data allows me to estimate v_i as time-invariant plant-specific characteristics and θ_y as plant-invariant annual characteristics. Besides this, ϵ_{iy} is an error term and the β are parameters to be estimated.

5.2 Data collection

Most Swedish water treatment plants are run by their respective municipalities and cost data is only obtainable at the actual plant, municipality or municipal company running it.[15] Thus, a smaller number of plants were sampled and asked to provide cost data for this analysis. Out of the 194 surface water treatment plants in Sweden, 77 have provided water quality data to the Swedish Geological Survey (SGU) since the year 2000, which made these plants a natural sample.

Water treatment plants can take either ground water or surface water as their main input. The reason I chose to only study surface water treatment plants is twofold: (1) they tend to use the same water treatment techniques, so comparability is high, and (2) surface water can be expected to be more directly affected by land-use than ground water. Regarding reason number 1, surface water treatment plants tend to be more dependent on chemicals in the treatment process, which makes them more suitable for the current study.[16]

[15]In several cases, it turned out, cost data was not obtainable at all, since reorganisations and/or bad administration had lead to people not knowing where to find it.

[16]During the data collection process, I visited two water treatment plants to understand the process and talk to those

5.2.1 Cost data

The quality of raw water can affect treatment costs in several ways. Factors such as chemical dosage, electricity use and the need for extra personnel can all be affected in the short run, when the water quality changes. One factor that could possibly be even more important is that many municipalities currently consider large investments in new treatment equipment, such as membrane filters worth several hundred million SEK. This is a consequence of the cryptosporidium outbreak in Östersund of 2010 as well as expectations that water quality will be reduced in the future due to climate change and/or a current trend of more organic material in many northern European water bodies.[17] However, treatment plant operators were quite clear that chemical use, more than any other factor, varies with raw water quality in the short run. For this reason, only chemical costs will be used in the cost function specification.

E-mails were sent out to all water treatment plants in the sample, asking how much money had been spent on water treatment chemicals per year between 2000 and 2014. I also asked how many m^3 drinking water had been delivered during each year, in order to calculate per-unit chemical cost. Clearly, this is overly simplistic, since choice of chemicals and price per unit of chemical is likely to differ both over time and across treatment plants (unit and year fixed effects in the panel data regressions can possibly control for some of these effects). Assuming that all treatment plants try to minimise chemical cost per unit of treated water, I chose to use this simplistic procedure. I received data from 13 different treatment plants, in total 171 yearly observations.[18] Chemical costs were adjusted for inflation using the Swedish consumer price index.[19] Some summary statistics for the cost data is given in table 1.

Table 1 **Summary statistics chemical costs**

Variable		Mean	Std.Dev.	Min	Max	Observations
Chem. cost (SEK/annum)	overall	2,072,629	2,610,605	59,870	9,865,648	N=171
	between		2,513,724			n=13
	within		588,445			
Water quantity (m^3/annum)	overall	11,889,381	16,255,643	123,856	46,100,000	N=171
	between		16,100,000			n=13
	within		992,277			
Chem. cost per unit (SEK/m^3)	overall	0.296	0.176	0.073	.979	N=171
	between		0.172			n=13
	within		.069			

5.2.2 Water quality data

The data available from SGU limited the sample of potential treatment plants to include in the study to 77 plants. These were the only plants with time-series going back to the year 2000 for their raw water test results data, which is also the year when the first land-use data is available for Sweden (see section 5.2.3). One of these plants was excluded, because it had not been used for the last six or seven years and its catchment could not be properly identified in the GIS software.

Water treatment plants feed in new data into the SGU database as often as they please. They are allowed to introduce new parameters to the database, which has given a rather broad range of parameters. The treatment plants in the sample had collectively reported 335,577 test results, distributed over 1,464

running it. I also visited the Swedish University of Agricultural Sciences (SLU) to talk to Stephan Köhler, a Landscape biogeochemist and specialist on the effects of land-use on water quality.

[17] This was discussed during study visits at treatment plants. For more information on the trend of more organic material in the water, see for example Haaland et al. (2010).

[18] Some treatment plants could only retrieve data for shorter period of time, which is why the panel data set is unbalanced.

[19] The CPI is available online at http://www.scb.se/pr0101/.

different parameters.[20] A parameter is here defined both by what is measured and how it is measured. For example, the concentration of E. coli was measured in eight different ways, which resulted in eight different parameter specifications because clearly, concentrations measured with different methods can not be compared to each other.

The data provided the opportunity to study a rather large group of water quality parameters, even though none of them provided a complete panel dataset, with observed values for all treatment plants and all months in the study period. The potentially relevant variables, for which there were enough observations to enable statistical analysis, were microorganisms, coliform bacteria, E. coli, nitrite, ammonium, chemical oxygen demand (COD), color and turbidity.

Based on two criteria, data availability and relevance in previous studies, I chose to study E. coli (measured in membrane filter) and the turbidity of the water. E. coli, which is a bacteria and a commonly used indicator of microbiological pollution, is measured by the number of bacteria per 100 ml of raw water. It is also the water quality variable for which the SGU data provided the largest amount of observations. Turbidity is a commonly used indicator of water quality, since it indicates the presence of sediment, suspended clay, silt, finely divided organic matter, algae and other microorganisms. It is measured in a unit called Formazin Nephelometric Unit (FNU), which is based on how infrared light is scattered in the water, and has been used in previous studies of how water quality affects chemical costs in water treatment plants (Dearmont, McCarl, and Tolman, 1998; Holmes, 1988; Moore and McCarl, 1987). Some descriptive statistics for the water quality variables as well as rainfall and total water flow, which will be used as control variables, are found in table 2.

Table 2 **Summary statistics: Water quality variables and rain**

Variable		Mean	Std.Dev.	Min	Max	Observations
E. coli (No./100ml)	overall	4.529222	17.10525	0.5	453.5	N=7981
	between	12.01337	0.5	78.68938		n=76
	within	13.39359	-61.66016	390.0122		T-bar=105.013
Turbidity (FNU)	overall	2.1745	3.897188	.025	130.285	N=4076
	between	2.145615	.1510484	10.23045		n=76
	within	3.191816	-7.145954	123.899		T-bar=53.6316
Rain (mm/month)	overall	64.75694	38.91686	0.7	268	N=12768
	between	12.23182	47.0006	97.40119		n=76
	within	36.97109	-28.34425	259.945		T=168
Total water flow (m^3/sec)	overall	60.16456	157.4863	0	1770	N=12768
	between	141.4305	0.0201961	597.125		n=76
	within	71.14082	-316.9604	1400.193		T=168

Even though only two pollutants are in focus, it can be argued that these two are indicative of other water pollution as well, since several water quality parameters are strongly correlated, especially within their "type", i.e. microbiological pollutants, nutrients and a third group with COD, color and turbidity. I studied these correlations using fixed effects regressions and the results are found in table 3. Another important thing to note in this table is that there are quite few observations in the regressions, because the observation is dropped entirely if there is any missing data. Out of 12768 (the dataset if it were complete), only 1,276 observations include measurements of turbidity, COD and color simultaneously.

In the original data, some labs have not reported values of zero for E. coli and turbidity, but instead specified the values at the smallest detectable concentration. Since, in this case, the true value lies somewhere between zero and the smallest detectable concentration, I included the parameter in two

[20]The data was extracted from the original Excel file and assembled into new spreadsheets using a Python code, which sorted the data according to id-code of the treatment plant, the date of the water quality test and the observed value. In the final datasets, yearly and monthly averages were used for each treatment plant.

forms, one where all those measurements were coded as zero and one where they were coded at the smallest detectable concentration. All econometric analysis in this study is based on the mean of those two values.

Table 3 **Correlations between pollutants**

VARIABLES	(1) ln (ecoli)	(2) ln (ammonium)	(3) ln (color)
ln (microorganisms)	0.0608***		
	(0.0140)		
ln (coliform bact.)	0.226***		
	(0.0243)		
ln (nitrite)		0.335***	
		(0.0656)	
ln (COD)			0.567***
			(0.0731)
ln (turbidity)			0.268***
			(0.0389)
Constant	-0.126*	-2.586***	2.330***
	(0.0748)	(0.340)	(0.129)
Observations	5,092	2,951	1,276
R-squared	0.215	0.093	0.441
No. of treatment plants	75	76	46

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

5.2.3 Land-use data

Data on land-use was obtained as shape-files from three different sources.[21] I received a highly detailed map of Sweden, from the Swedish mapping, cadastral and land registration authority.[22] This map includes all buildings in Sweden and their type, which means that water treatment plants can be separated from other buildings. Since this map is only available in its last updated form, I used another source of land-use data that made it possible for me to extract time-series: the Corine Land Cover (CLC) maps with land-use in 44 different categories for most European countries (a list of the categories is found in appendix B).[23] The CLC maps are provided for 1990, 2000, 2006 and a not-yet-validated update for 2012. I use the three latter versions (Sweden was not included in the first version), which means that I can follow land-use changes over 13 years. However, the fact that I have high frequency (approximately one observation per month) data on water quality, but only three measurements on land-use over the entire time period poses a dilemma. Following Vincent et al. (2015) I chose to linearly interpolate the land-use changes between the measurement years, allowing variation between but not within years. This interpolation is likely to cause two problems:

Measurement error: Clearly, all land-use changes in Sweden between 2000 and 2006 as well as 2006 and 2012 can not have happened exactly linearly over time, so linear interpolation will cause measurement error. I find no reason to believe that this error should be systematic, so assuming that the error is random it will cause *attenuation bias*: regression coefficients will be biased towards zero (Wooldridge, 2008, p. 319-320). This would imply that regression coefficients are more likely to be conservative and less likely

[21]A type of data used in geographical information system (GIS) software. I used the free software QGIS to extract the data into spreadsheet format.

[22]In Swedish: Lantmäteriet.

[23]These maps are made available by the Copernicus land monitoring services, coordinated by the European Environment Agency (EEA).

to be exaggerated estimates of the true effect.

The Moulton problem: When the dependent variable is allowed to vary at a higher level (per month in this case) than the explanatory variables (which only vary per year), there is a risk that standard errors are underestimated. This downward bias on the standard errors was shown by Moulton (1986), but can be addressed by clustering standard errors on a higher level. Following Vincent et al. (2015), I cluster standard errors by treatment plant.

Shape-files with catchment areas was collected from the Swedish Meteorological and Hydrological Institute (SMHI). One file contains the most detailed breakdown of Sweden into 52,778 basins, in which all run-off water has one unique drainage point. This file is linked to another file in which each basin is represented including all its upstream basins. Since the entire upstream area can be very large, I needed to limit it. One important reason for this is that some catchments stretch into neighbouring Norway and Finland and the CLC maps can differ between countries. I decided to include only land-use changes within a circle with radius 100 kilometers from the water treatment plant. It is not obvious how distance matters in this context and even in the water research community there is a scientific debate about whether the entire catchment or only zones closer to water (riparian zones) matter more for downstream water quality (Sliva and Williams, 2001). The distance limit of 100 km in the current study was thus rather practical than scientific. An advantage of this limit, however, is that several of the treatment plants in the sample used raw water from the same lake (eleven of the treatment plants in the sample take their raw water from lake Vättern for example). With no limitation of the catchment area, these treatment plants would be considered as equally affected by the same land-use changes, whereas with the limitation, land-use differed according to the treatment plant's location around the lake. To study the how the distance between land-use changes and water treatment plants matters, a test was conducted in which a subset of the treatment plants were studied with a limit of the upstream area to 30 km from the treatment plant. Almost 50% of the treatment plants had small enough catchments that the entire upstream area was included in the 30 kilometre circle.

Table 4 **Summary statistics for land-use**

Variable	Mean	SD	SD (between)	SD (within)	b/w ratio	Median	Min	Max
Catchment area (ha)	252821	373926	-	-	-	60875	126	1870823
Urban %	0,0123	0,0179	0,0180	0,0004	45	0,0077	0,0000	0,1131
Agriculture %	0,1112	0,1052	0,1056	0,0002	528	0,1041	0,0000	0,4089
Wetland %	0,0153	0,0256	0,0257	0,0000	-	0,0084	0,0000	0,1749
Waterbodies %	0,1806	0,1225	0,1230	0,0000	-	0,1446	0,0000	0,4493
Forest %	0,5725	0,1607	0,1600	0,0216	7.4074	0,5978	0,0983	0,9749
Logged %	0,0667	0,0626	0,0591	0,0215	2.7488	0,0528	0,0000	0,3752
Other %	0,0410	0,1493	0,1499	0,0001	1499	0,0000	0,0000	0,8485

The process of extracting land-use data from maps is described in appendix A and summary statistics for the land-use variables for the 76 studied catchment areas are presented in table 4. An important thing to note is that variation in land-use within catchment areas over time is very low, especially for urban and agricultural land-uses (see figures 1 and 2). In fact, the parameter estimates for urban and agricultural land-uses are likely to be difficult to estimate with fixed effects regressions, since more than half of the water treatment plants have zero within variation for these two variables (in a fixed effects regression, time-invariant variables are differenced out, see section 5.3). Wetlands have often been the focus of studies on water purification. However, as can be seen in table 4, the within-variation in wetlands is zero, and its effect on water quality can thus not be studied in a fixed effects framework.

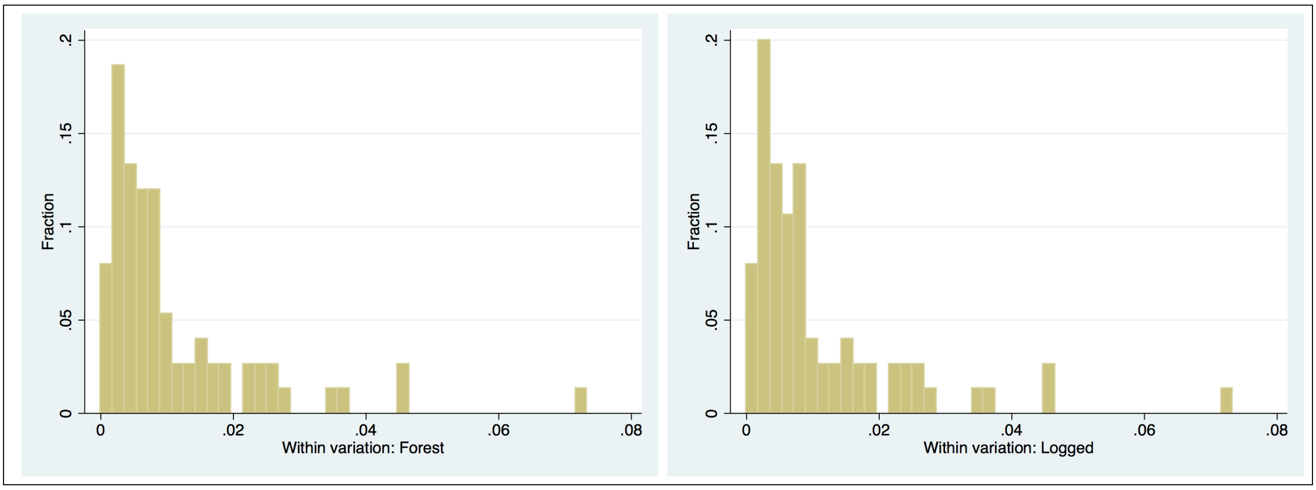

Figure 1 **Within variations for forest and logged forest.**

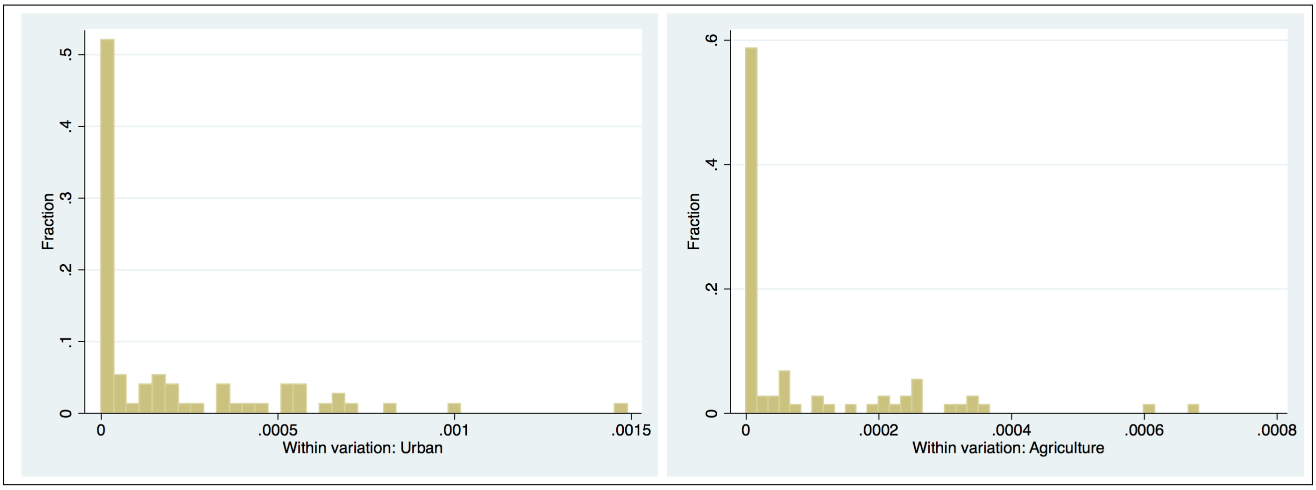

Figure 2 **Within variations for urban and agricultural land-uses.**

5.2.4 Control variables

An important variable to control for is rain, which is expected to increase run-off of pollutants into waterways and increase turbidity of the water. For the same reason, total water flow in the water body is expected to affect pollution levels as well. Flow is to some extent a consequence of rain, but e.g. snowmelt can have dramatic effects as well, especially in the spring. Model based data for rainfall as well as total flow is available from SMHI, for each of the basins in the most detailed breakdown.[24] Rainfall is thus only at the most detailed level, but flow takes all upstream flow into account. All variables that will be used in the econometric analysis are briefly presented and defined in table 5 and CLC land-use categories are found in appendix B.

[24]The data is available for download at http://vattenwebb.smhi.se/modelarea/.

Table 5 **Variable definitions and expected signs**

Variable name	Description	Expected sign
ln(E. coli)	Observed values, # of bacteria per 100 ml water (ln).	N/A
ln(turbidity)	Observed values, measured in FNU (ln).	N/A
ln(1 + forest)	Share of upstream forest land-use (CLC categories 311, 312 and 313) plus 1 (ln).	Negative.
ln(1 + logged)	Share of upstream cleared forest land-use (CLC category 324) plus 1 (ln).	Positive.
ln(1 + agriculture)	Share of upstream agricultural land-use (CLC categories 211, 222, 231 and 242) plus 1 (ln).	Positive.
ln(1 + urban)	Share of upstream urban land-use (CLC categories 111 - 142) plus 1 (ln).	Ambiguous.
ln(rain)	Modeled value of rainfall (mm/month) in locality of treatment plant (ln).	Positive.
ln(flow)	Modeled value of total water flow (m^3/sec) in locality of treatment plant (ln).	Positive.

For a discussion on expected signs see section 4.1.

5.2.5 Data quality

The data suffers from a few limitations. The most important one is probably the irregularity with which water treatment plants have reported water quality data to the SGU database. This makes the panel dataset unbalanced and means that for some water quality parameters, statistical inference can suffer from too few observations, especially when dynamic relationships are studied and lagged values included. It also introduces a possible bias, since quality tests can have been conducted only at times when concentrations of pollutants were especially high or low — a type of bias that is less likely in cases where treatment plants have reported data on a regular basis. It is, however, unlikely that treatment plant operators could manipulate the data, since all water quality tests are conducted by labs separate from the treatment plants and it is the labs, not the treatment plants, that deliver the data to the SGU database. This issue will be discussed further in section 5.3.2.

The most recently released CLC land cover data is used in this analysis in order to assure the longest possible time series. The 2012 release has not yet been validated by the data provider, however, and it is therefore possible that there are errors in the land-use data. Assuming that this error is random, it should cause estimated regression coefficients to be more conservative (see paragraph about measurement error in section 5.2.3). It should also be noted that the level of detail is not as high in the CLC maps as it is in maps that are available from the Swedish mapping, cadastral and land registration authority. However, this authority does not currently make it possible to obtain a time-series of maps.

The cost data was collected by the author, using a very simplistic e-mail survey. Treatment plants were first asked for a broad set of data regarding chemical costs and usage, but almost none of the plants could accommodate the requests. I therefore asked only for the aggregated costs of water treatment chemicals per year. This is clearly too simplistic to give anything more than indications of how treatment costs correlate with water quality.

5.3 Panel data analysis

The fact that I have panel data over 76 water treatment plants and their upstream areas over the years 2000-2013 allows me to control for unobserved unit-specific as well as time-specific effects (by year or month). In this subsection I will describe the most common methods to analyse panel data, as well as discuss advantages and disadvantages of different model specifications and the trade-offs between them.

Consider a model of the form

$$y_{it} = \alpha + \beta x_{it} + v_i + \epsilon_{it} \tag{1}$$

where v_i is a time-invariant, unit-specific, error term and ϵ_{it} is a "regular" error term with mean zero, that is uncorrelated with x and itself. If the equation above is true, it must also be true that

$$\bar{y}_i = \alpha + \beta \bar{x}_i + v_i + \bar{\epsilon}_i \tag{2}$$

where $\bar{y}_i$, $\bar{x}_i$ and $\bar{\epsilon}_i$ are time-averages for each unit i. Subtracting (2) from (1) gives

$$(y_{it} - \bar{y}_i) = \beta(x_{it} - \bar{x}_i) + (\epsilon_{it} - \bar{\epsilon}_i) \tag{3}$$

thus, the time-invariant, unit-specific, error term has been subtracted out of the equation (and the same would be true for any variable that did not vary over time). Equation (3), estimated by OLS, is known as the fixed-effects estimation.

One way of estimating panel data, without disregarding the effects of time-invariant variables, is to use random-effects. This is a weighted average of the estimates produced by equation (2) (also known as the between-effects) and (3), given by

$$(y_{it} - \theta\bar{y}_i) = (1-\theta)\alpha + \beta(x_{it} - \theta\bar{x}_i) + [(1-\theta)v_i + (\epsilon_{it} - \theta\bar{\epsilon}_i)]$$

where θ is a function of σ_v^2 and σ_ϵ^2. If $\sigma_v^2 = 0$ then $\theta = 0$ as well and equation (1) can be estimated by OLS directly. The opposite, $\sigma_\epsilon^2 = 0$ and $\theta = 1$, gives equation (3) and a fixed-effects model. The crucial point for the choice of model is thus what can be assumed about the individual specific effect, v_i. If v_i is correlated with the explanatory variables, x_{it}, then the random effects model is inconsistent and biased. If v_i is random, the model can be specified as

$$y_{it} = \alpha + \beta x_{it} + \eta_{it}$$

where $\eta_{it} = v_i + \epsilon_{it}$ with $E(\eta_{it}) = 0$, and consistently estimated with a random effects model. A formal test for this specification choice is the Hausman test, in which the null hypothesis is that the random effects model is consistent.[25] It should be noted that the fixed effects model is consistent under both the null and alternative hypotheses.

One factor to take into account when choosing which model to use, is that fixed effects models do not allow the estimation of time-invariant variables. A related, but less recognised, problem of fixed effects is its inefficiency in estimating the effect of variables that have very limited variation over time (Plümper and Troeger, 2007). As can clearly be seen in table 4, variation within panels is strongly dominated by variation between panels in the case of land-use data. This can hamper inference. Not only does it imply low levels of statistical significance but estimates can also be imprecise (Wooldridge, 2002, p. 286). Using Monte Carlo simulations, Plümper and Troeger study how the root mean squared errors (RMSE), thus taking both bias and inefficiency into account, of a regular fixed effects model depends on the ratio of *within* and *between* variation (b/w ratio). They conclude that the regular fixed effects model becomes less reliable when the b/w ratio increases, and the choice of model should take this ratio into account.[26] According to the authors, the choice of model becomes a trade-off between the ability to compute estimates of rarely changing variables and avoiding omitted variable bias. This is also emphasised by Beck (2001) who states that "If variables of interest are being lost because of the inclusion of fixed effects, the researcher must weigh the gains from including fixed effects against their costs. [...] Like most interesting issues, this is a matter of judgement, not slavish adherence to some 0.05 test level" (Beck, 2001, p. 285).

Another potential problem in the econometric analysis is autocorrelation. If the autocorrelation is an AR(1) process, in which case the error term is given by $\epsilon_{it} = \rho\epsilon_{i,t-1} + u_{it}$ with $|\rho| < 1$ and u_{it} being IID

[25] There are two versions of the Hausman test. I mainly used the version that is based on (Wooldridge, 2002, p. 290-291), which can be used with robust and clustered standard errors. When I needed to test specifications that were adjusted for autocorrelation, I used the version based of Hausman (1978).

[26] To counter the problem of rarely changing variables, Plümper and Troeger (2007) propose an estimation technique that they call *fevd*, fixed effects vector decomposition. This proposed method was heavily criticised by Breusch et al. (2011) who claim that *fevd* is just a Hausman-Taylor instrumental variable approach that is avoiding making important assumptions about exogeneity. The critique is directed towards the proposed *fevd* estimation approach, not Plümper and Troeger's motivation for why a new estimation approach is needed.

$(0, \sigma_u^2)$, it can be analysed with a method derived by Baltagi and Wu (1999). This method estimates ρ and uses this estimate to transform the data and remove autocorrelation. To remove the fixed effects with the transformed data, the first observation of each panel is dropped. An important advantage of this method is that Baltagi and Wu (1999) developed it explicitly to work with unbalanced panels with gaps, such as the dataset in the current study. A disadvantage is that it is unable to control for heteroskedasticity.

5.3.1 Dynamic panel data estimation

One advantage of panel data is that it allows a type of instrumental variable estimation, even in cases where suitable instrumental variables are hard to find. The generalised method of moments (GMM), allows the use of lagged values of variables in the model as instruments. These types of models can also control for autocorrelation and heteroskedasticity, while the instrumental variables handle endogeneity. Several techniques have been proposed for this type of estimation — one that is commonly used was developed by Arellano and Bond (1991), and it is pedagogically described by Roodman (2009b). Even though the estimator was designed for short panels (large N, small T), it will be used here as a robustness check rather than the main specification.

In order to remove fixed effects, as well as handle dynamic panel bias (which is not likely to be present in panels with large T, as in the current study), the data is transformed. Two such transformations are proposed by Roodman (2009b): the first-difference transform and the forward orthogonal deviations transform, originally developed by Arellano and Bover (1995). Since first-difference transformations magnifies gaps in unbalanced panels, forward orthogonal deviations is a more appropriate choice in the current study. Instead of subtracting the previous observation from the current, it subtracts the average of all future available observations of the variable. An additional advantage is that it leaves all lagged observations outside the formula, which means that they are available as instruments.

5.3.2 Handling missing data

The water quality data from SGU is not intended to be a panel dataset of any particular frequency, but should rather be seen as a collection of comparable water quality test results over a long time. Fitting them into a panel dataset with monthly observations thus yields a panel with many missing values, since several treatment plants report their results at a lower frequency than monthly. Another explanation is that some treatment plants report a test result for e.g. E.coli measured by another method than most of the other plants. Even though most econometric packages and estimation commands can run with missing values in the data, dropping all time periods which contain missing data would deprive the analysis from substantial information.[27] One way to handle missing data is to impute values. This is not uncommon in the panel data context, one reason being that respondents often drop out of survey panels. One famous example is the Michigan Panel Study of Income Dynamics (PSID), which suffered a loss of almost 50% of its initial sample between 1968 and 1989. Fitzgerald, Gottschalk, and Moffitt (1998) study the effect of this attrition on the reliability of results. Even though they find that attrition is significantly related to such characteristics as income and education level of respondents, they do not find that it has seriously distorted the representativeness of the panel. Another example is the Student Teacher Achievement Ratio (STAR) project in Tennessee, a large-scale randomised experiment where approximately 11,000 students in 79 schools were followed. This study suffered from 20% to 30% attrition rates annually and imputation was used by e.g. Dee (2004) to handle missing test scores.

Both these examples show rates of missing data that are comparable to the current study. Arguably, it can be considered less problematic in this case, since missingness is very unlikely to be correlated to upstream land-uses and raw water quality. Rather, missingness can be expected to be attributed to some

[27] The standard procedure is often to perform so-called *complete case analysis*, in which all observations with missing values are dropped and estimation is based on those that remain.

characteristic of the treatment plant or the municipality that runs it, such as lack of funding for test labs or a lack of interest in reporting test values to the SGU register. Thus the risk of selection bias, such as may be expected in the choice to withdraw from a survey, should be smaller.

Regardless of why data is missing, it is still likely to reduce estimation efficiency — a problem that can be mitigated by imputation. Most imputation methods are based on the assumption that the missing data is *missing at random* (MAR) or *missing completely at random* (MCAR), where the first indicates that missingness can somehow be predicted by the variables in the dataset and the second means that the missing values are entirely random. If the missing data is *missing not at random* (MNAR), imputation is more troublesome (Nicoletti, 2006). Quite clearly, it is almost impossible to prove anything about the missing values with certainty, since they cannot be observed. What remains for the researcher to do, is to study the patterns of missing data and make assumptions about what caused the missingness. I created a new variable that would take the value 1 if E. coli was missing and 0 otherwise, and studied the correlations between this new variable and the other variables in the dataset. If data would be missing systematically for episodes with high or low levels of E. coli in the water, there would probably be a correlation with rainfall, since rain is a very strong predictor of E. coli concentrations. However, this correlation is found to be -0.0156, which suggests that missing values for E. coli should not be systematically higher or lower than the reported test results.

A common method traditionally has been to fill out the missing values with the mean of the variable in question, so-called *mean imputation*. This rather unsophisticated method may actually predict missing data quite well, but distorts estimated variance and correlations (Schafer and Graham, 2002). With artificially low variance, there is a risk that significance is overstated when the data is analysed. There are, however, several other methods to impute data, each with their advantages and drawbacks. I have chosen to use multiple imputation (MI), recommended by e.g. Schafer and Graham (2002). Whereas one imputation represents one set of plausible values for the missing data, MI represents multiple sets of plausible values drawn from a distribution, acknowledging the fact that we can only make educated guesses of what the true values are. Traditionally, quite few sets of imputed data have been used, Royston (2004) mentions 3-5 sets for example, but recent research has recommended larger numbers. Each complete dataset is analysed independently and parameter estimates are averaged across the copies to give a single estimate.

5.4 Limitations

Quite clearly, to fully capture the value of ecosystem water purification as an intermediate input to municipal water treatment all costs — short-run and long-run — in Swedish water treatment plants should be studied. It is not unlikely that the long-run cost function of treatment plants could be influenced equally or even more by ecosystem service water purification than the short-run cost function. The current discussions in several Swedish water treatment plants about whether to invest in membrane filters or other modern techniques, worth several hundred million SEK, was mentioned in section 5.2.1. It would be of very high interest to study whether the need for municipal investments in water treatment equipment is correlated to upstream ecosystem services. Because of data constraints this did not seem possible, however.

Other aspects that could affect municipal water treatment costs are e.g. the ownership structure of treatment plants (discussed briefly in section 5.1) or which political party that happens to be in power in the specific municipality, since political preferences are likely to influence investments in water treatment. This is beyond the scope of the current study, however.

6 Empirical analysis

In the following section I will present the results of the econometric analysis. The first part of the section will be devoted to studying how land-use affects water quality and the second part will analyse the effect of water quality on treatment chemical costs.

6.1 Does land-use affect water quality?

Regressions on water quality measures and land-use were made using the dataset with monthly observations from January 2000 to December 2013 (T = 168) over 76 water treatment plants. I will first present results of how land-use affects E. coli, which is the variable with the largest amount of observations, and then turbidity, which has been shown in previous research to be an important determinant of water treatment costs.

6.1.1 Concentrations of E. coli

Results from fixed effects regressions with E. coli as the dependent variable are found in table 6. Quite clearly, coefficient estimates vary with the specification — except the coefficient for rainfall, which remains in the range 0.125 - 0.165 and significant at the 1% level regardless of the specification. The R^2 is very low in these fixed effects regressions, especially when no lagged dependent variables are included, which implies that the within-variation explains very little of the variation in the dependent variable. The only land-use effects that are significantly different from zero are found using the Baltagi-Wu regressions in column 4 (the method is described in section 5.3) adjusting for autocorrelation of type AR(1). These results are not robust to heteroskedasticity, however, which makes it quite likely that significance is overestimated. Baltagi-Wu and Beck-Katz regression results are included, because they each provide different methods to handle serial correlation.

The modified Wald test for groupwise heteroskedasticity in fixed effects models rejected the null hypothesis of homoskedasticity at the 1% level. To counter heteroskedasticity, as well as the Moulton problem (mentioned in section 5.2.3), standard errors are calculated using so-called Huber-White sandwich estimates, clustered at treatment plant level, in specifications 1-4. In column 6, using Beck and Katz's panel-corrected standard errors, panel-level heteroskedastic errors are assumed. I used Wooldridge's test for autocorrelation in panel data to distinguish whether the errors are serially correlated or not.[28] The null hypothesis of no first order autocorrelation was rejected at the 1% level, so this clearly needs to be controlled for.

The fact that levels of E. coli are persistent can be seen in columns 4 and 6, with two lags of the dependent variable (both significant at the 1% level) included as regressors. With the lags included, parameter estimates of forest and logged land switch signs to negative. This is the expected sign for forests but unexpected for logged land. Including lagged dependent variables comes at the cost of losing quite a few observations since, if two lags are included, each gap in the dataset means that three observations are lost. One great advantage of the Baltagi-Wu approach in this case is that serial correlation can be addressed without losing a large number of observations. This is a reason why the results of column 4 are interesting, in spite of the risk of underestimated standard errors.

Another risk when including lagged dependent variables in a fixed effects regression, as in columns 3, 4 and 6 of table 6, is the so-called dynamic panel bias, identified by Nickell (1981). This bias arises when time-means are subtracted from each observation, see equation (3) in section 5.3, because lagged dependent variables become correlated with the error. The bias is not mitigated by increasing N, but it

[28]Drukker (2003) performed a simulation study of this test and found no indication that the test would fail to identify autocorrelation in unbalanced panels or in the presence of gaps in the data. This information is relevant here, since the panels are unbalanced in the current study.

Table 6 **E. coli fixed effects estimations**

ln(E. coli)	(1) FE	(2) FE	(3) FE	(4) FE	(5) BW-FE	(6) BK-FE
ln (1 + forest)	12.50	17.93	-0.176	-2.606	-1.156***	-2.112
	(12.24)	(16.26)	(8.883)	(9.495)	(0.373)	(7.816)
ln (1 + logged)	7.414	12.16	-0.408	-2.862	-1.885*	-2.685
	(8.818)	(11.44)	(6.412)	(6.676)	(1.047)	(5.293)
ln (1 + agriculture)		290.5	305.3	90.47	1.095	-174.2
		(454.0)	(451.9)	(428.0)	(1.304)	(203.3)
ln (1 + urban)		199.1	173.5	111.0	-13.45**	24.10
		(203.0)	(197.5)	(183.2)	(5.716)	(95.91)
ln (rain)	0.142***	0.153***	0.153***	0.165***	0.125***	0.169***
	(0.0249)	(0.0253)	(0.0229)	(0.0244)	(0.0159)	(0.0165)
ln (flow)	0.0952***	0.100***	0.0600**	0.0494**	0.107***	0.0468***
	(0.0332)	(0.0339)	(0.0251)	(0.0237)	(0.0137)	(0.0126)
ln (E. coli) t-1			0.308***	0.271***		0.253***
			(0.0312)	(0.0282)		(0.0161)
ln (E. coli) t-2				0.0943***		0.0710***
				(0.0163)		(0.0163)
Constant	-7.912	-41.72	-32.14	-9.194	0.398***	3.427
	(7.359)	(47.77)	(46.40)	(43.77)	(0.0733)	(5.908)
Year FE	No	Year	Year	Year	Year	Year
Observations	7,549	7,549	6,369	5,860	7,473	5,860
R-squared	0.020	0.028	0.120	0.126		0.583
No. of treatment plants	76	76	76	75	76	75

Robust standard errors in parentheses, except BW.
BW is Baltagi-Wu AR(1)-adjustment. BK is Beck-Katz with panel specific AR(1).
*** p<0.01, ** p<0.05, * p<0.1

is with increasing T. With a long panel, as in the current study, bias is thus likely to be less of a problem. Nevertheless, a dynamic panel data estimation will be performed in section 6.1.5 below, not only because of the risk of dynamic panel bias, but also because it is a way to address several potential estimation issues, including serial correlation, heteroskedasticity and endogeneity.

It should be noted that coefficients cannot be interpreted straightforwardly from the regression table, since land-use is specified as the logaritm of $1 + L/A$, where L is the specific land-use and A is total upstream area.[29] A measure of elasticity is thus given by $\hat{\beta}\left(\frac{L/A}{1+L/A}\right)$. It must also be mentioned again that the coefficients for agricultural and urban land-uses might be less credible as a consequence of the very small within-variance. In columns 2-4, parameter estimates for these two variables are much larger in magnitude than what can be expected to be true and they are very far from significant.

More precise estimates could possibly be estimated by a random effects model, provided that the assumption holds that the unit specific effects, v_i are uncorrelated with the explanatory variables, x_{it}. To test this assumption, I conducted two versions of the Hausman test, first for the model in column 2 of table 6. Here, the test statistic is $\chi^2 = 8.67$ (p-value 0.0699). For the same test with the model in column 4, thus including year dummies as well as lagged dependent variables, the test statistic is $\chi^2 = 7.51$ (p-value 0.1111). These two results can be taken as indications that random effects can be used. As discussed in section 5.3, it can also be argued that the choice between random and fixed effects models can be a trade-off between the costs and benefits of using the two models, rather than a strict rule-based decision. In the current case, the benefits of using random effects is that the cross-sectional variation contains substantial information that is disregarded with fixed effects. The cost of using random effects is the risk

[29] The reason for this specification is that the share of some land-uses are zero in several areas, in which case the logarithm is not defined.

of omitted variable bias — a risk that should be kept in mind when studying the results.

For random effects regression results, see table 7. Coefficients estimated with random effects have the expected signs and estimates change less depending on specification, even though they still do not appear as very robust. Clearly, it cannot be proven that omitted variable bias is not present in the random effects model, but parameter estimates are more in line with expected results and the negative effect of forests and the positive effect of agriculture on levels of E. coli in the water are statistically significant.

Table 7 **E. coli random effects estimations**

ln(E. coli)	(1) RE	(2) RE	(3) RE	(4) RE	(5) BW-RE	(6) BK-RE
ln (1 + forest)	-1.002	-0.349	-0.824*	-0.747*	-0.541	-1.109***
	(1.194)	(0.953)	(0.480)	(0.425)	(0.831)	(0.276)
ln (1 + logged)	-1.361	0.255	0.259	0.139	0.580	-0.251
	(1.198)	(1.131)	(0.474)	(0.386)	(0.965)	(0.229)
ln (1 + agriculture)		1.856	1.169**	0.930**	1.949**	0.922***
		(1.510)	(0.583)	(0.441)	(0.965)	(0.244)
ln (1 + urban)		5.142	1.126	0.832	3.552	0.370
		(5.316)	(2.102)	(1.677)	(4.680)	(0.876)
ln (rain)	0.144***	0.157***	0.193***	0.206***	0.129***	0.202***
	(0.0249)	(0.0254)	(0.0309)	(0.0291)	(0.0158)	(0.0172)
ln (flow)	0.0954***	0.0939***	0.0107	0.00518	0.0985***	-0.00198
	(0.0300)	(0.0306)	(0.0130)	(0.0102)	(0.0131)	(0.00573)
ln (E. coli) t-1			0.610***	0.442***		0.434***
			(0.0546)	(0.0281)		(0.0156)
ln (E. coli) t-2				0.273***		0.212***
				(0.0277)		(0.0158)
Constant	0.117	-0.596	-0.398	-0.452	-0.420	-0.223
	(0.716)	(0.584)	(0.296)	(0.277)	(0.534)	(0.197)
Year FE	No	Year	Year	Year	Year	Year
Observations	7,549	7,549	6,369	5,860	7,549	5,860
R-squared						0.441
No. of treatment plants	76	76	76	75	76	75

Robust standard errors in parentheses, except BW.
BW is Baltagi-Wu AR(1)-adjustment. BK is Beck-Katz with panel specific AR(1).
*** p<0.01, ** p<0.05, * p<0.1

6.1.2 Multiple imputation to handle missing data

As discussed in section 5.3.2, inference can possibly improve if missing values are imputed. For E. coli, test values are missing for approximately 38% of the treatment plant-month pairs. The missing values are not evenly distributed across treatment plants, however. For those treatment plants with a large share of empty cells, imputation is less reliable. Dropping the treatment plants that have more than 50% of the observations missing (23 treatment plants) reduces the rate of missing values to approximately 26%.

The amount of imputations needed for a certain efficiency has been shown to be given by $(1 + \lambda/m)^{-1}$, where λ is the proportion of missing values and m is the number of imputations (Schafer and Graham, 2002, p. 165). With $\lambda = 0.26$ it should thus suffice with $m \approx 5$ imputations to reach 95% efficiency. In their review of imputation methods, Schafer and Graham (2002) provide an example in which 80% of the data is missing and multiple imputation was conducted with $m = 20$.

Imputing values for E.coli based on predictive mean matching, I obtained an almost balanced panel for 54 treatment plants. Predictive mean matching uses a normal regression to obtain linear predictions, then uses these predictions to form a set of *nearest neighbours* from the non-missing values. Finally, it randomly draws an imputed value from this set of neighbours. The choice of number of nearest neighbours is a trade-off between possible bias, which increases with the number of neighbours, and variance, which can be too low if the number is small.

I used a regression with modeled values of rainfall and total water flow, percentages of forest, logged, agricultural and urban land-uses and estimated treatment plant fixed effects as explanatory variables to make 20 imputations, each making a random draw from 5 nearest neighbours. Fixed effects regression results with the imputed data are found in table 8. The results show no significant effect of land-use at all, but rain and total water flow are significant and the magnitude of their coefficients are in the same magnitude as in the earlier estimations. The parameter estimate for the forest variable has the unexpected positive sign in all fixed effects estimations (columns 1-3), but a negative sign when random effects are assumed in column 4. However, it is far from significant in all 4 specifications.

Quite clearly, imputations did not help with statistical inference. One possible reason is that a significant share of the treatment plants are dropped in these regressions, so the larger amount of observations per plant might be balanced out by the smaller number of plants. All in all, only a slightly larger number of observations are gained.

Table 8 **E.coli estimations with multiple imputations**

	(1)	(2)	(3)	(4)
ln(E. coli)	FE	FE	FE	RE
ln (1 + forest)	6.154	10.47	8.211	-1.780
	(8.944)	(11.16)	(9.444)	(1.942)
ln (1 + logged)	4.085	7.978	6.333	-0.359
	(6.662)	(8.134)	(6.891)	(1.454)
ln (1 + agriculture)	201.9	172.9	142.5	0.639
	(358.2)	(387.9)	(336.5)	(1.511)
ln (1 + urban)	95.35	120.7	105.7	-1.203
	(149.3)	(180.2)	(155.1)	(3.060)
ln (rain)	0.126***	0.136***	0.141***	0.172***
	(0.0264)	(0.0270)	(0.0254)	(0.0364)
ln (flow)	0.108***	0.112***	0.0942***	0.0122
	(0.0286)	(0.0295)	(0.0263)	(0.0196)
ln (E. coli) t-1			0.168***	0.516***
			(0.0251)	(0.0567)
Constant	-26.20	-26.26	-21.57	0.384
	(38.31)	(43.61)	(37.62)	(1.258)
Time FE	No	Year	Year	Year
Observations	8,383	8,383	8,295	8,295
No. of treatment plants	53	53	53	53

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

6.1.3 Turbidity levels

Turbidity is a highly relevant parameter to study, since previous research has found that it has a notable effect on chemical costs in treatment plants — for example, Dearmont, McCarl, and Tolman (1998) found that a 1% increase in turbidity of raw water increased chemical costs in treatment plants by 0.27%. However, the SGU data for turbidity is much less complete than the data for E. coli, so the relatively smaller number of observations is likely to reduce estimation precision and reliability. Estimation results are found in table 9. Just like the estimations with E. coli as the dependent variable, R^2 is very low in these fixed effects regressions, especially when no lagged dependent variables are included. The within-variation thus explains very little of the variation in turbidity levels. The only significant effect of land-use is found using the Beck-Katz fixed effects regression in column 4, in which urban land-use is estimated to reduce turbidity (significant at the 10% level). As can be seen in columns 3 and 4, where a lagged dependent variable is added as a regressor, turbidity is persistent and just like in the case of E. coli the parameter estimates for forest and logged land switch sign to negative when controlling for this persistence. This comes at a cost of losing additional observations, however. Somewhat surprisingly, Wooldridge's test for autocorrelation in panel data yields an F-statistic of 3.453 (p-value = 0.0700) so the null hypothesis of no first order autocorrelation cannot be rejected at the 5% level. Therefore, no AR(1)-corrected models are included.

As can also be seen in columns 3 and 4, when the Beck-Katz regression is not controlling for panel-specific serial correlation of type AR(1), estimated coefficients are identical to regular fixed effects. Standard errors are smaller however, which is because they are not calculated using Huber-White sandwich estimates. The Hausman test for fixed or random effects (performed on the model of column 3 in table 9) yields the test statistic $\chi^2 = 41.36$ (p-value = 0.0000), implying that the random effects model should not be used. No such estimation results are thus presented here. As in the case for E. coli, rainfall and total water flow are found to significantly increase turbidity levels, regardless of the model specification.

Table 9 **Turbidity fixed effects estimations**

ln(turbidity)	(1) FE	(2) FE	(3) FE	(4) BK-FE
ln (1 + forest)	9.262	18.52	-0.381	-0.381
	(11.11)	(13.58)	(22.23)	(15.42)
ln (1 + logged)	3.957	10.13	-3.073	-3.073
	(7.154)	(8.900)	(13.91)	(9.798)
ln (1 + agriculture)		147.5	-19.72	-19.72
		(211.6)	(390.9)	(228.4)
ln (1 + urban)		61.69	-104.1	-104.1
		(109.9)	(189.3)	(92.51)
ln (rain)	0.0933***	0.0918***	0.0633**	0.0633***
	(0.0265)	(0.0271)	(0.0316)	(0.0196)
ln (flow)	0.0648*	0.0647*	0.0890**	0.0890***
	(0.0333)	(0.0330)	(0.0361)	(0.0134)
ln (turbidity) t-1			0.276***	0.276***
			(0.0798)	(0.0261)
Constant	-5.624	-26.80	4.191	0.670
	(6.431)	(21.92)	(43.38)	(10.08)
Time FE	No	Year	Year	Year
Observations	3,994	3,994	1,912	1,912
R-squared	0.020	0.032	0.135	0.790
No. of treatment plants	76	76	65	65

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

6.1.4 Testing the implication of shorter distance to treatment plants

The analysis above is conducted on all land-use changes that happened upstream of any of the water treatment plants in the sample within a circle of 100 km from the plant. Since there is no consensus in previous studies about how to make the choice of this distance (there is rather a scientific debate on the matter, see Sliva and Williams 2001), a test was conducted in which I studied a subset of plants where the maximum distance between land-use and plant was reduced to 30 km. The amount of studied water treatment plants were reduced to 62 and the number of observations dropped accordingly. Estimation results are found in table 10.

Intuitively, land-use changes closer to the treatment plants are expected to matter more for water quality of the plants' raw water, but on the other hand observations are fewer and land-use changes even less than when a larger area is studied. The specification in column 1 of table 10 corresponds to that in column 4 of table 6. Signs remain unchanged and all land-use effects are equally insignificant. In column 2, which is a random effects model that corresponds to column 4 of table 7, the effect of forest changes from significantly negative to practically zero, logged and agricultural land switch sign from positive to negative and urban is found to be significantly positive. Columns 3 and 4 of table 10 correspond to columns 2 and 3 of table 9, with and without lagged values of the dependent variable as regressors. Signs remain similar in column 3, and far from statistically significant. In column 4, however, agricultural and urban land-uses switch sign to positive, which is the expected direction. Agriculture is even significant at the 10% level, but the magnitude of the estimated coefficient is unrealistically large. The coefficient can interpreted as the elasticity by recalculating according to $\beta\frac{L/A}{1+L/A}$. Evaluated at the average share of agricultural land-use (approximately 11%) the elasticity would thus be 21.8, which does not seem reasonable. A likely explanation to the unrealistic coefficient estimate is the minimal within-variation for agricultural land-uses.

Table 10 **FE and RE estimations with shorter distance to plants**

VARIABLES	(1) ln(E. coli) FE	(2) ln(E. coli) RE	(3) ln(turbidity) FE	(4) ln(turbidity) FE
ln (1 + forest)	-4.900	-0.00468	5.710	-13.24
	(9.110)	(0.497)	(11.88)	(23.37)
ln (1 + logged)	-4.589	-0.308	1.930	-11.65
	(6.479)	(0.433)	(7.978)	(14.81)
ln (1 + agriculture)	156.1	-0.0221	186.4	217.9*
	(291.8)	(0.385)	(145.8)	(129.3)
ln (1 + urban)	58.89	4.167**	56.96	13.16
	(146.7)	(1.930)	(78.77)	(59.96)
ln (rain)	0.154***	0.189***	0.0944***	0.0626*
	(0.0262)	(0.0317)	(0.0294)	(0.0335)
ln (flow)	0.0477	0.0290*	0.0664*	0.0902**
	(0.0288)	(0.0155)	(0.0363)	(0.0388)
ln (E. coli) t-1	0.287***	0.454***		
	(0.0296)	(0.0282)		
ln (E. coli) t-2	0.0990***	0.273***		
	(0.0168)	(0.0305)		
ln (turbidity) t-1				0.278***
				(0.0843)
Constant	-15.93	-0.737**	-27.09	-20.80
	(35.18)	(0.373)	(19.40)	(23.04)
Observations	4,943	4,943	3,575	1,753
R-squared	0.135		0.032	0.138
No. of treatment plants	61	61	62	56

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

One possible explanation for why most of the previous results do not hold when studying the area closer to the treatment plants is that a smaller area studied means that even less land-use change is observed over the time-series. It should also be remembered that the estimations in table 10 are based on a subset of the treatment plants. These estimations are thus based both on fewer observations and less longitudinal variation. These results should not be taken to imply that land-use closer to the treatment plants matters less for water quality. It is likely that the situation in Sweden, with relatively low rates of land-use change, requires that larger areas are studied in order to find significant effects. Not least with fixed effects panel data estimations.

On a related note, it is likely that land-use in riparian zones — those areas nearest to water bodies — matter more for water quality than does land-use further from water bodies. Studying the effect of land-use in riparian zones would be of high interest, and GIS-data is available. However, land-use regulations are more strict in these zones than in most other areas, so variation over time is likely to be even lower in these zones, which might make fixed effects estimation impossible.

6.1.5 Dynamic models

Since the estimations in sections 6.1.1 and 6.1.3 showed that both E. coli and turbidity levels are highly persistent, the data is analysed using Arellano-Bond GMM with the data transformed by forward orthogonal deviations. This should also be seen as a robustness check, even though the results presented so far have not been very robust. The estimator, described in section 5.3.1, is based on the assumptions that the data-generating process (1) is dynamic, so current realisations of the dependent variable depends on past ones, (2) that some regressors may be endogenous and (3) that the idiosyncratic disturbances may have unit-specific heteroskedasticity and serial correlation but that they are uncorrelated across units.

Whereas the land-use variables are treated as potentially endogenous, flow and rain are assumed to be strictly exogenous. Lags of the land-use variables are thus used as GMM-type instruments, but the length of the panel (T = 168) means that the amount of lags used must be limited, since there can easily be too many. It has been recommended in the literature to test the sensitivity of the model to different amount of lags, which is why I include lags 2-4 in one case and in another case I collapse the instrument matrix to one column of instruments, as is described by Roodman (2009a, p. 148). The instrument count is presented together with estimation results in table 11.

As can be seen in table 11, forests are found to significantly reduce levels of E. coli in the water — a result that is not robust to all changes in the instrument count and number of lags of the dependent variable, but both magnitude and sign remains somewhat the same. The effect of agriculture on levels of E. coli has the expected positive sign in all specifications but is not significant. The effects of logged and urban lands are ambiguous and switch signs. The effect of rain is positive and highly significant as in most other specifications throughout this study.

In line with the previous analysis, the GMM estimation results do not tell us much about the effects of land-use on turbidity levels. Whereas forests are found to significantly reduce levels of turbidity in the first specification (column 4), the sign is positive in the third specification (column 6). In fact all parameter estimates switch sign depending on specification, except logged (always negative, never significant) and rain and flow (always positive, but not always significant). A likely explanation is, as before, that we have far too many missing values for turbidity, a problem that is magnified when lags of the dependent variable are included as regressors.

6.2 Does water quality affect treatment costs?

Acknowledging that the data for chemical costs in water treatment plants is very limited — an unbalanced panel with yearly observations from 13 treatment plants over the years 2000-2014 — I studied how costs were affected by the raw water quality, measured by the indicators studied in section 6.1, turbidity and E. coli. Fixed effects regression results are found in table 12.

In columns 1 and 2 of table 12, the per-unit cost of chemicals is used as the dependent variable and in column 3 and 4, aggregate cost is used and quantity of treated water is added as a regressor. The fact that water quantity is not significant is quite surprising. Since only aggregate chemical costs are studied

Table 11 **Arellano-Bond GMM estimations**

VARIABLES	(1) ln (E. coli)	(2) ln (E. coli)	(3) ln (E. coli)	(4) ln (turbidity)	(5) ln (turbidity)	(6) ln (turbidity)
ln (1 + forest)	-1.068***	-0.773	-1.011***	-2.971***	-0.551	0.205
	(0.240)	(0.478)	(0.389)	(0.539)	(0.889)	(0.989)
ln (1 + logged)	-0.0949	-0.0345	0.220	-1.472	-3.219	-0.300
	(0.322)	(0.680)	(1.426)	(0.986)	(3.728)	(6.675)
ln (1 + agriculture)	0.304	0.234	2.916	-1.578**	0.261	3.985
	(0.363)	(0.445)	(1.939)	(0.766)	(1.773)	(3.342)
ln (1 + urban)	-1.676	1.096	-22.99**	0.553	0.888	-31.34
	(1.916)	(1.386)	(11.39)	(1.762)	(5.870)	(26.97)
ln (rain)	0.165***	0.129***	0.184***	0.404***	0.124	0.0475
	(0.0292)	(0.0408)	(0.0297)	(0.0863)	(0.0810)	(0.0322)
ln (flow)	0.00576	0.000525	0.0313	0.00252	0.0429	0.0773**
	(0.00959)	(0.0125)	(0.0280)	(0.0227)	(0.0700)	(0.0303)
ln (E. coli) t-1	0.791***	0.703***	0.278***			
	(0.0361)	(0.0654)	(0.0309)			
ln (E. coli) t-2		0.111**	0.103***			
		(0.0449)	(0.0181)			
ln (turbidity) t-1				0.884***	0.685***	0.298***
				(0.0451)	(0.103)	(0.0971)
ln (turbidity) t-2					-0.0348	0.104***
					(0.116)	(0.0380)
Time FE	Year	Year	Year	Year	Year	Year
Lags	2-4	2-4	collapsed	2-4	2-4	collapsed
No. of instruments	1463	1779	840	1406	1430	837
Observations	6,369	5,860	5,860	1,912	1,515	1,515
No. of treatment plants	76	75	75	65	50	50

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

and I have not controlled for the amounts used, year fixed effects are included in columns 2 - 4 and should to some extent control for price changes on chemicals that affect all treatment plants simultaneously. In column 3 I dropped the rain variable, which made the parameter estimate for turbidity less significant, in spite of adding some more observations to the analysis (this is because rain statistics for 2014 was not yet available from SMHI at the time of writing). Looking at the R-squared, as well as the estimated coefficients for rain and the constant, year fixed effects ameliorates the model significantly.

Judging from these results, it seems quite clear that turbidity is the most relevant pollutant to study in order to predict chemical costs in water treatment plants. Since both costs and turbidity levels are expressed in logaritms, the coefficient can be interpreted as an elasticity, implying that a 1% increase in turbidity levels increase chemical costs in water treatment plants by 0.10 - 0.13 %. This is lower than what was found by Dearmont, McCarl, and Tolman (1998), who reported a chemical cost elasticity of 0.27 but higher than the figures estimated by Holmes (1988), who found an elasticity of 0.07 for all short-run water treatment costs. This result is thus in line with previous findings.

Yearly rainfall is included in this regression as a control variable, but it is interesting in itself. The estimated elasticity of approximately 0.2 means that a year during which rainfall is 1% above average, chemical costs can be expected to be 0.2% above average. Considering that most climate change models predict up to 40% more precipitation in Sweden in the future (Arheimer et al., 2005), the effects on water quality and water treatment costs should not be neglected.

Table 12 **Cost function estimations**

VARIABLES	(1) ln (cost/m^3)	(2) ln (cost/m^3)	(3) ln (total cost)	(4) ln (total cost)
ln (m^3 water)		-0.542	0.400	0.458
		(0.329)	(0.350)	(0.329)
ln (turbidity)	0.126***	0.108***	0.0950**	0.108***
	(0.0222)	(0.0323)	(0.0370)	(0.0323)
ln (ecoli)	-0.00131	-0.0293	-0.0302	-0.0293
	(0.0183)	(0.0216)	(0.0229)	(0.0216)
ln (rain)	-0.00823	0.210**		0.210**
	(0.104)	(0.0948)		(0.0948)
Constant	-1.346*	5.155	7.451	5.155
	(0.703)	(4.715)	(5.187)	(4.715)
Time FE	No	Year	Year	Year
Observations	138	138	150	138
R-squared	0.070	0.291	0.221	0.265
No. of treatment plants	13	13	13	13

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

6.3 Further steps to be taken

It is quite likely that the fixed effects analysis is hampered by the low variation over time of land-use variables. To some extent, this is an unavoidable consequence of a slowly changing landscape in a country with strict regulations on land-use. But it is plausibly also a question of data quality. With higher precision maps, smaller land-use changes can be detected, and with more land-use classes the changes between different types of forests, for example, could be studied. Fortunately, GIS-data is currently developing quite rapidly, with both quality and quantity of maps increasing quite rapidly, so future studies can possibly benefit from more and better data on land-use change. A longer dataset, including land-use changes over more years, could possibly ameliorate inference of the effects of land-use as well. At this point, I have not found data sources that would allow longer time-series or higher precision.

Forest ecosystems have been highlighted by previous research as important for water purification. Forests are, however, heterogenous. Vincent et al. (2015) separate logged forest from virgin forest and find that the latter has a more positive effect on water quality.[30] The CLC land-use data did not allow such a separation between forest types. Considering that Vincent et al. find that virgin forest is significantly more efficient than other forest-types at providing clean water, a separation of the forest variable into e.g. production forest and non-production forest could possibly have yielded more robust results.

Another methodological issue that needs to be resolved is that of which land-use that is most relevant — the land-use that is (i) close to the treatment plant, (ii) close to upstream water bodies or (iii) all upstream land-use. This issue clearly points to the need for interdisciplinary work. Environmental economists need to team up with hydrologists, biologists and geographers to better understand this issue in ecosystem service valuations.

[30] A reminder: In the study by Vincent et al. (2015), "logged forest" refers to secondary forest. In the current study it refers to land that has recently been logged and is currently in a transitional state.

7 Discussion and conclusions

The aim of this study has been to econometrically infer a monetary value of the water purification services of ecosystems, using a cost function based valuation method. The dataset, constructed by the author and combining cost data from treatment plants, land-use data from publicly available GIS-files and water quality data from the Swedish Geological Survey, allowed fixed and random effects as well as GMM analysis of how upstream land-use affects water quality and how water quality, in turn, affects treatment costs. Unfortunately, the data is of limited quality, with a rather large proportion of missing values, which is one likely explanation to the lack of robust results. In spite of data limitations, indications have been found in random effects specifications as well as Arellano-Bond GMM specifications that upstream forests do reduce levels of E. coli downstream. However, logging of forests is not found to significantly increase levels, which would be expected given the effect of forests. One possible explanation lies in the fact that a large proportion of Swedish forests are production forests, thus regular logging is a part of this ecosystem. With data that could separate primary and secondary forest, it is likely that results would be different. If so, then that would also be in line with the findings of Vincent et al. (2015), who show that virgin forests reduce water treatment costs more than what secondary forests do.

Studying how upstream land-use affects levels of E. coli, the results are ambiguous for the effects of urban and agricultural land-uses. One likely explanation is that the within-variation is very low for these land-uses. With a random effects specification, thus including cross-sectional variation in the analysis, agricultural land-uses are found to significantly increase levels of E. coli (see table 7).

Results are even less robust when studying the effect of upstream land-use on turbidity levels. This is likely to be explained at least to some extent by the even more limited data availability for turbidity. Consequently, not much can be said in this study about how land-use affects turbidity of downstream water. This is clearly problematic when the ambition is to study the effect on water treatment chemical costs, because the cost function estimates in section 6.2 show that turbidity is an important determinant of chemical costs, whereas no effect is found for E. coli. In order to plausibly estimate how upstream land-use affects water treatment costs, it is essential that the effect of upstream land-use on turbidity levels in Swedish water bodies is studied further. Better data is surely needed for this purpose.

Even though it is possible that the benefits to society of decreased levels of microbiological pollutants in water bodies cannot be captured by the chemical cost function valuation method explored in this study, it surely does not mean that reducing levels of E. coli is without value. Considering the immense costs of waterborne disease outbreaks, carried through the public water supply, this service might indeed be more valuable than the reduction of chemical costs in water treatment plants. This study has pointed out that forests seem to reduce levels of E. coli downstream, implying that forest ecosystems do provide this service to society. What remains is to find ways of quantifying the value of that service.

Safeguarding water resources is a political priority, in developing as well as developed countries. Several previous studies have argued that safe and clean drinking water is not only an issue of water treatment technology but also of watershed management. This can be connected to the academic discussion about the substitutability between physical and natural capital, referred to in section 3.4.1. The fear of future water quality degradation is currently making Swedish municipalities invest in modern, and expensive, water treatment equipment. With better understanding of how upstream land-use affects water quality, investments in watershed management could be compared to these water treatment investments to see what is more cost-efficient. In developed countries, however, land tends to be expensive and the public sector can often afford these types of investment. In many developing countries, the opposite is more likely to be true, which is why watershed management is likely to be a more economically viable alternative in these countries. Studies like the current one can be well placed to inform decision-makers about these types of trade-offs.

This study was inspired by the recently published and first-of-its-kind valuation study by Vincent et al. (2015), which studied the effects of upstream land-use on downstream water treatment costs in Malaysia. Whereas Vincent et al. were able to present robust econometric evidence that upstream forests reduce water treatment costs, they barely touched the question of how forests actually affect water quality — it was only implicitly assumed that water quality was the link between land-use and treatment costs. The current study, which cannot establish the link assumed by Vincent et al., illuminates the importance of studying both parts of this relationship.

In order for studies like this one to be of higher relevance for ecosystem service valuation, more geographic precision is needed. If a significant relationship between upstream land-use and downstream water quality can be robustly proven, the next step is to discuss how the location of the specific land-use matters. One aspect that can be of importance is topography of the landscape, for example whether a certain land-use is in a slope or not. It is not unlikely that the land-use changes that have been studied in this paper appeared too far from water bodies to have a significant effect on water quality. An interesting next step would be to reduce the amount of land studied, and narrow down the focus towards areas closer to water, e.g. riparian zones. A test was conducted in section 6.1.4, where the maximum distance between a land-use change and the treatment plant was reduced to 30 km, instead of 100 km. This was only done for a subset of the treatment plants and their upstream areas, however. The fact that this analysis did not give any significant results (except that the coefficient for urban land-use was found to have a significant effect on levels of E. coli in a random effects regression) should not be taken as an indication that land-use closer to treatment plants matters less for water quality at the plants' raw water intake. Rather, it is likely to be explained by the even lower within-variation of land-use. Since land-use in proximity to water is quite heavily regulated in Sweden, it is likely that panel data estimations of land-use in these zones will suffer from very low longitudinal variation and fixed effects estimations might not be possible.

Rain and total water flow have been used as control variables throughout the econometric analyses in this study. Almost regardless of the specification and the variables studied, rainfall has been found to significantly increase levels of water pollutants as well as treatment costs. This is interesting in its own right, not least since climate change is likely to increase precipitation and total water flow in Sweden. Thus, municipal water providers, and consumers, need to prepare for higher water treatment costs in the future. These costs should not be disregarded when estimating the economic consequences of climate change, and this could possibly be an important avenue of future research.

This study is part of a growing body of ecosystem service valuation studies, an area of which the policy relevance is growing, as a consequence of the increasing ambitions to include the value of environmental goods and services in decision making and national accounts. From the early understanding of environmental degradation as a consequence of public goods and external costs (Pigou, 1924; Hardin, 1968), emanated the view that ecosystems constantly contribute to the economic system and that it should be considered an integral part of it (Dasgupta, 2008). Consequently, just as the net present value of human and physical capital is determined by current and expected flows of income, ecosystems can be treated as economic assets that yield a flow of economically valuable benefits. To know the value of the stock, we must understand these benefit flows (Barbier, 2011). The current study has contributed to this type of understanding in the context of water purification in Sweden.

Several methods have been proposed for the valuation of ecosystem services, each with its advantages and disadvantages (Daily et al., 2000). In the area of watershed ecosystem services in the Nordic countries, there have been calls for research using the cost function approach to valuation (Barton et al., 2012, p. 52-53). This study has responded to that call.

Since clean water is a prerequisite for life, the importance of finding cost efficient ways to safeguard this resource can hardly be exaggerated. The current study has contributed to this area of research by discussing methodological development and highlighting complexities such as how to best study the effect of land-use (such as choosing the distance to treatment plant or water body), which water quality parameters that matter most (only turbidity or E. coli as well) and which economic effects in society (water treatment costs, treatment equipment investments or waterborne disease outbreaks) that should be studied in relation to the water quality parameters chosen. Even though much work remains in order to credibly estimate the value of the water purification services of ecosystems and to include this value in decision-making and national accounts, steps have been taken toward this end.

References

Arellano, Manuel and Stephen Bond (1991). “Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations”. In: *The review of economic studies* 58.2, pp. 277–297.

Arellano, Manuel and Olympia Bover (1995). “Another look at the instrumental variable estimation of error-components models”. In: *Journal of econometrics* 68.1, pp. 29–51.

Arheimer, Berit et al. (2005). “Climate change impact on water quality: model results from southern Sweden”. In: *AMBIO: A Journal of the Human Environment* 34.7, pp. 559–566.

Arrow, Kenneth J et al. (1996). “Is there a role for benefit-cost analysis in environmental, health, and safety regulation?” In: *Science* 272.5259, pp. 221–222.

Atkinson, Giles, Ian J Bateman, and Susana Mourato (2012). “Recent advances in the valuation of ecosystem services and biodiversity”. In: *Oxford Review of Economic Policy* 28.1, pp. 22–47.

Baltagi, Badi H and Ping X Wu (1999). “Unequally spaced panel data regressions with AR (1) disturbances”. In: *Econometric Theory* 15.6, pp. 814–823.

Barbier, Edward B (2007). “Valuing ecosystem services as productive inputs”. In: *Economic Policy* 22.49, pp. 178–229.

– (2011). *Capitalizing on nature : Ecosystems as natural assets*. Cambridge, UK: Cambridge University Press. ISBN: 9781107007277.

Barton, David N et al. (2012). *Valuation of ecosystem services from Nordic watersheds.* Nordic Council of Ministers.

Bateman, Ian J et al. (2002). “Applying geographical information systems (GIS) to environmental and resource economics”. In: *Environmental and Resource Economics* 22.1-2, pp. 219–269.

Beck, Nathaniel (2001). “Time-series-cross-section data: What have we learned in the past few years?” In: *Annual review of political science* 4.1, pp. 271–293.

Bockstael, Nancy E et al. (2000). “On measuring economic values for nature”. In: *Environmental Science and Technology* 34.8, pp. 1384–1389.

Bosch, JM and JD Hewlett (1982). “A review of catchment experiments to determine the effect of vegetation changes on water yield and evapotranspiration”. In: *Journal of hydrology* 55.1, pp. 3–23.

Brander, Luke M, Raymond JGM Florax, and Jan E Vermaat (2006). “The empirics of wetland valuation: a comprehensive summary and a meta-analysis of the literature”. In: *Environmental and Resource Economics* 33.2, pp. 223–250.

Brander, Luke M and Mark J Koetse (2011). “The value of urban open space: Meta-analyses of contingent valuation and hedonic pricing results”. In: *Journal of environmental management* 92.10, pp. 2763–2773.

Breusch, Trevor et al. (2011). “FEVD: just IV or just mistaken?” In: *Political Analysis* 19.2, pp. 165–169.

Brundtland, Gru et al. (1987). “Our Common Future”. In:

Byström, Olof (2000). “The replacement value of wetlands in Sweden”. In: *Environmental and Resource Economics* 16.4, pp. 347–362.

Carlsson, Fredrik, Peter Frykblom, and Carolina Liljenstolpe (2003). “Valuing wetland attributes: an application of choice experiments”. In: *Ecological Economics* 47.1, pp. 95–103.

Chichilnisky, Graciela and Geoffrey Heal (1998). “Economic returns from the biosphere”. In: *Nature* 391, pp. 629–630.

Costa, Marcos Heil, Aurélie Botta, and Jeffrey A Cardille (2003). "Effects of large-scale changes in land cover on the discharge of the Tocantins River, Southeastern Amazonia". In: *Journal of Hydrology* 283.1, pp. 206–217.

Costanza, Robert et al. (1997). "The value of the world's ecosystem services and natural capital". In: *Nature* 387.

Daily, Gretchen C et al. (2000). "The value of nature and the nature of value". In: *Science* 289.5478, pp. 395–396.

Dasgupta, Partha (2008). "Nature in economics". In: *Environmental and Resource Economics* 39.1, pp. 1–7.

– (2009). "The welfare economic theory of green national accounts". In: *Environmental and Resource Economics* 42.1, pp. 3–38.

Davies, John-Mark and Asit Mazumder (2003). "Health and environmental policy issues in Canada: the role of watershed management in sustaining clean drinking water quality at surface sources". In: *Journal of environmental management* 68.3, pp. 273–286.

Dearmont, David, Bruce A McCarl, and Deborah A Tolman (1998). "Costs of water treatment due to diminished water quality: a case study in Texas". In: *Water Resources Research* 34.4, pp. 849–853.

Deaton, Angus (2007). *Income, aging, health and wellbeing around the world: Evidence from the Gallup World Poll.* Tech. rep. National Bureau of Economic Research.

Dee, Thomas S (2004). "Teachers, race, and student achievement in a randomized experiment". In: *Review of Economics and Statistics* 86.1, pp. 195–210.

Drukker, David M (2003). "Testing for serial correlation in linear panel-data models". In: *Stata Journal* 3.2, pp. 168–177.

Fitzgerald, John, Peter Gottschalk, and Robert A Moffitt (1998). "An analysis of sample attrition in panel data: The Michigan Panel Study of Income Dynamics". In: *National Bureau of Economic Research (Cambridge, Mass., USA)*, pp. 1–85.

Freeman, A. Myrick III, Joseph A. Herriges, and Catherine L. Kling (2014). *The measurement of environmental and resource values : theory and methods.* 3. ed. ISBN: 9780415501576 (hardback).

Freeman, Jade, Rebecca Madsen, and Kelley Hart (2008). "Statistical analysis of drinking water treatment plant costs, source water quality, and land cover characteristics". In: *Trust for Public Lands.*

Gallai, Nicola et al. (2009). "Economic valuation of the vulnerability of world agriculture confronted with pollinator decline". In: *Ecological economics* 68.3, pp. 810–821.

Ghermandi, Andrea et al. (2009). "The values of natural and constructed wetlands: a meta-analysis". In: *TI (Tinbergen Institute)* 80.3.

Gomez-Baggethun, Erik and David N Barton (2013). "Classifying and valuing ecosystem services for urban planning". In: *Ecological Economics* 86, pp. 235–245.

Gren, Ing-Marie (1995). "The value of investing in wetlands for nitrogen abatement". In: *European Review of Agricultural Economics* 22.2, pp. 157–172.

Haaland, S et al. (2010). "Quantifying the Drivers of the Increasing Colored Organic Matter in Boreal Surface Waters". In: *Environmental Science and Technology* 44.8, pp. 2975–2980.

Hardin, Garrett (1968). "The tragedy of the commons". In: *science* 162.3859, pp. 1243–1248.

Hausman, Jerry A (1978). "Specification tests in econometrics". In: *Econometrica: Journal of the Econometric Society* 46.6, Econometrica: Journal of the Econometric Society.

Höglund, Lena (1997). "Estimation of Household Demand for Water in Sweden and its implications for a potential Tax on Water Use". In: *Studies in Environmental Economics and Development (Sweden).*

Holmes, Thomas P (1988). "The offsite impact of soil erosion on the water treatment industry". In: *Land Economics* 64.4, pp. 356–366.

Jones, Luke R and Christian A Vossler (2014). "Experimental tests of water quality trading markets". In: *Journal of Environmental Economics and Management* 68.3, pp. 449–462.

Kahneman, Daniel and Jack L Knetsch (1992). "Valuing public goods: the purchase of moral satisfaction". In: *Journal of environmental economics and management* 22.1, pp. 57–70.

Ledesma, José LJ, Stephan J Köhler, and Martyn N Futter (2012). "Long-term dynamics of dissolved organic carbon: Implications for drinking water supply". In: *Science of the total Environment* 432, pp. 1–11.

Leggett, Christopher G and Nancy E Bockstael (2000). "Evidence of the effects of water quality on residential land prices". In: *Journal of Environmental Economics and Management* 39.2, pp. 121–144.

Lowrance, Richard et al. (1997). "Water quality functions of riparian forest buffers in Chesapeake Bay watersheds". In: *Environmental Management* 21.5, pp. 687–712.

Mac Kenzie, William R et al. (1994). "A massive outbreak in Milwaukee of Cryptosporidium infection transmitted through the public water supply". In: *New England journal of medicine* 331.3, pp. 161–167.

Mjelde, James W et al. (1984). "Using farmers' actions to measure crop loss due to air pollution". In: *Journal of the Air Pollution Control Association* 34.4, pp. 360–364.

Moore, Walter B and Bruce A McCarl (1987). "Off-site costs of soil erosion: a case study in the Willamette Valley". In: *Western Journal of Agricultural Economics* 12.1, pp. 42–49.

Moulton, Brent R (1986). "Random group effects and the precision of regression estimates". In: *Journal of econometrics* 32.3, pp. 385–397.

Murphy, James J et al. (2005). "A meta-analysis of hypothetical bias in stated preference valuation". In: *Environmental and Resource Economics* 30.3, pp. 313–325.

Neeliah, Harris and Bhavani Shankar (2010). "Ozone pollution and farm profits in England and Wales". In: *Applied Economics* 42.19, pp. 2449–2458.

Nickell, Stephen (1981). "Biases in dynamic models with fixed effects". In: *Econometrica: Journal of the Econometric Society* 49.6, pp. 1417–1426.

Nicoletti, Cheti (2006). "Nonresponse in dynamic panel data models". In: *Journal of Econometrics* 132.2, pp. 461–489.

Pigou, A. C. (1924). *The economics of welfare.* 2. ed. London: Macmillan.

Plümper, Thomas and Vera E Troeger (2007). "Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects". In: *Political Analysis* 15.2, pp. 124–139.

Polasky, Stephen et al. (2011). "The impact of land-use change on ecosystem services, biodiversity and returns to landowners: A case study in the State of Minnesota". In: *Environmental and Resource Economics* 48.2, pp. 219–242.

Roodman, David (2009a). "A Note on the Theme of Too Many Instruments". In: *Oxford Bulletin of Economics and Statistics* 71.1, pp. 135–158.

– (2009b). "How to do xtabond2: An introduction to difference and system GMM in Stata". In: *Stata Journal* 9.1, pp. 86–136.

Royston, Patrick (2004). "Multiple imputation of missing values". In: *Stata Journal* 4.3, pp. 227–241.

Saarikoski, Heli et al. (2015). “Exploring operational ecosystem service definitions: The case of boreal forests”. In: *Ecosystem Services*.

Sagoff, Mark (2002). “On the value of natural ecosystems: the Catskills parable”. In: *Politics and the Life Sciences*, pp. 19–25.

Schafer, Joseph L and John W Graham (2002). “Missing data: our view of the state of the art”. In: *Psychological methods* 7.2, pp. 147–177.

Sliva, Lucie and D Dudley Williams (2001). “Buffer zone versus whole catchment approaches to studying land use impact on river water quality”. In: *Water research* 35.14, pp. 3462–3472.

Stiglitz, Joseph E, Amartya Sen, and Jean-Paul Fitoussi (2010). *Mismeasuring our lives: Why GDP doesn't add up*. The New Press.

Tong, Susanna TY and Wenli Chen (2002). “Modeling the relationship between land use and surface water quality”. In: *Journal of environmental management* 66.4, pp. 377–393.

UN (2005). *Ecosystems and human well-being*. Island Press, Washington, DC.

Vincent, Jeffrey R et al. (2015). “Valuing Water Purification by Forests: An Analysis of Malaysian Panel Data”. In: *Environmental and Resource Economics*, pp. 1–22.

Widerström, Micael et al. (2014). “Large outbreak of Cryptosporidium hominis infection transmitted through the public water supply, Sweden”. In: *Emerging infectious diseases* 20.4, pp. 581–589.

Wooldridge, Jeffrey M. (2002). *Econometric analysis of cross section and panel data*. 1st ed. Cambridge, Mass.: MIT Press.

– (2008). *Introductory econometrics : a modern approach*. 4. ed. Mason, Ohio: South-Western Cengage Learning.

A GIS Data collection

In this appendix, the land-use data extraction process is described with simple screenshots and captions. The extraction was done by the author, using the software QGIS, and the data was then converted to spreadsheets. All figures are screenshots from QGIS (version 2.10.1) and they include the following layers: catchment areas from SMHI, the BY-layer (buildings) from the Swedish mapping, cadastral and land registration authority and the so-called Change Layer 2006-2012 from CLC.[31]

Figure 3 **The small black dot at the southwestern end of the lake is the water treatment plant.**

[31]The catchment area layer is available at http://vattenwebb.smhi.se/ and the CLC change layer is available at http://land.copernicus.eu/pan-european/corine-land-cover/clc-2012/view. The BY layer is part of the product Fastighetskartan, which is not publicly available.

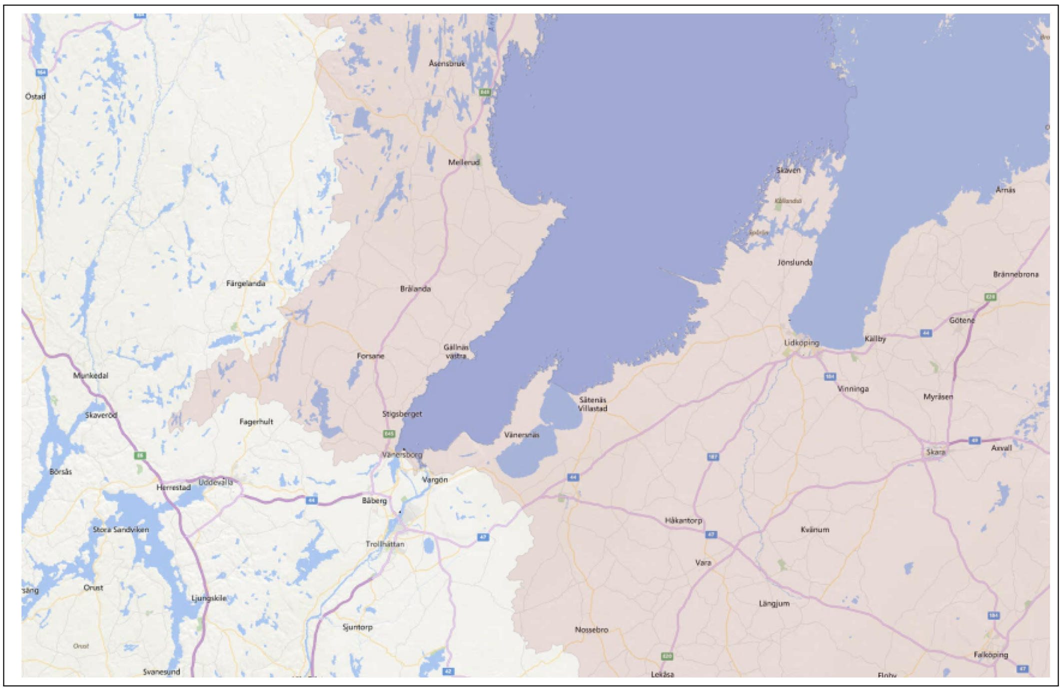

Figure 4 **The upstream area of the water treatment plant is identified.**

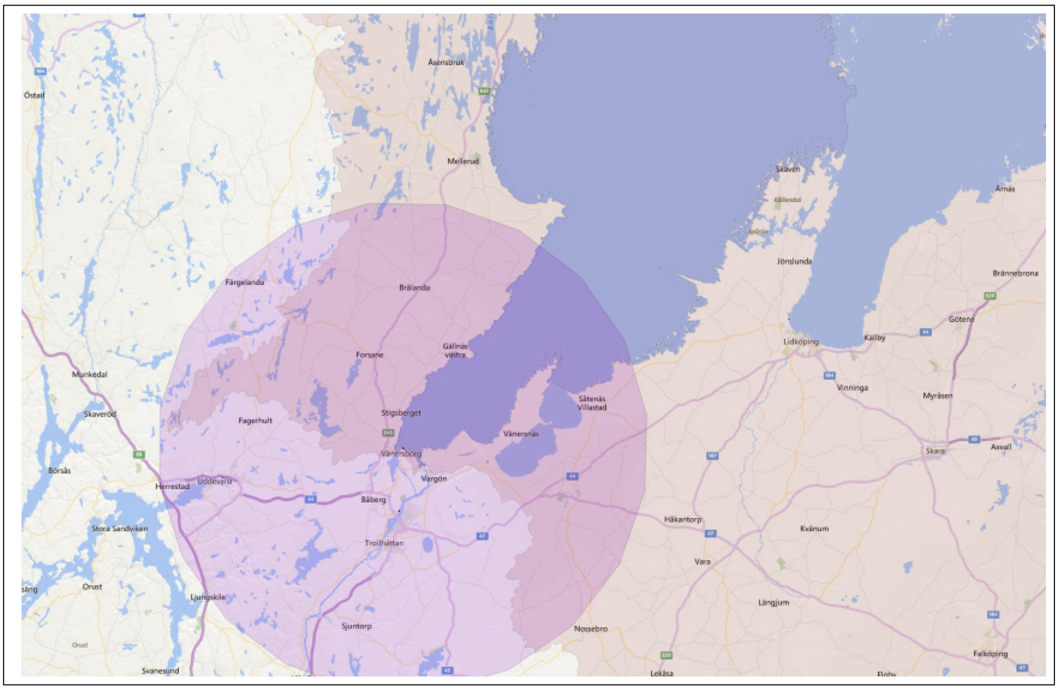

Figure 5 **A circle with radius 30 km around the water treatment plant is created.**

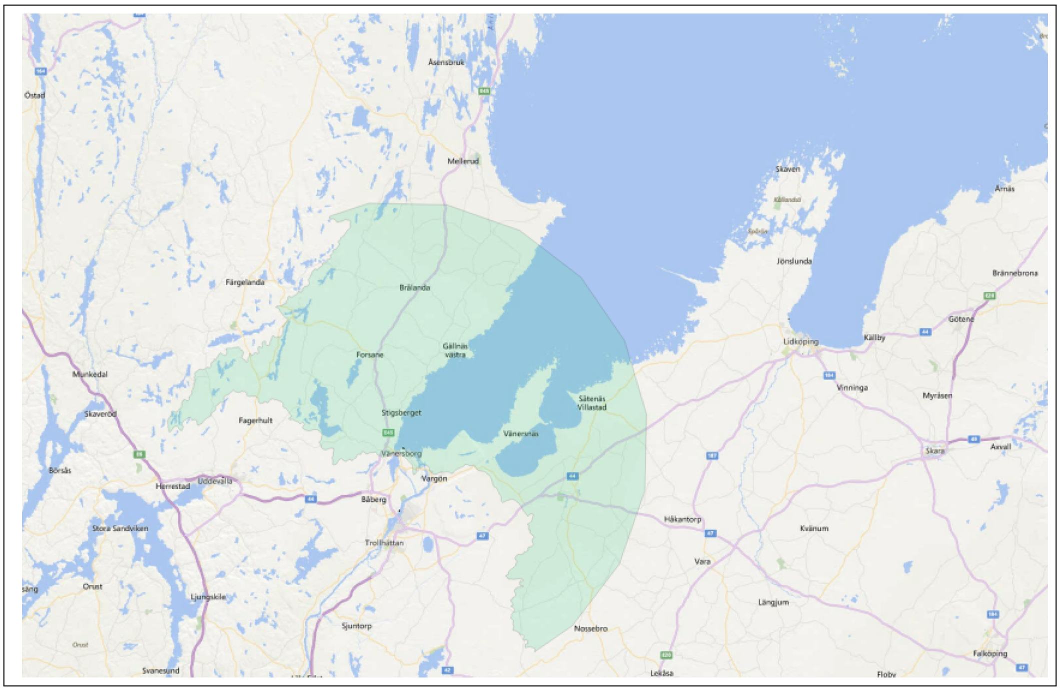

Figure 6 **The intersection of the two layers will be the area relevant to study.**

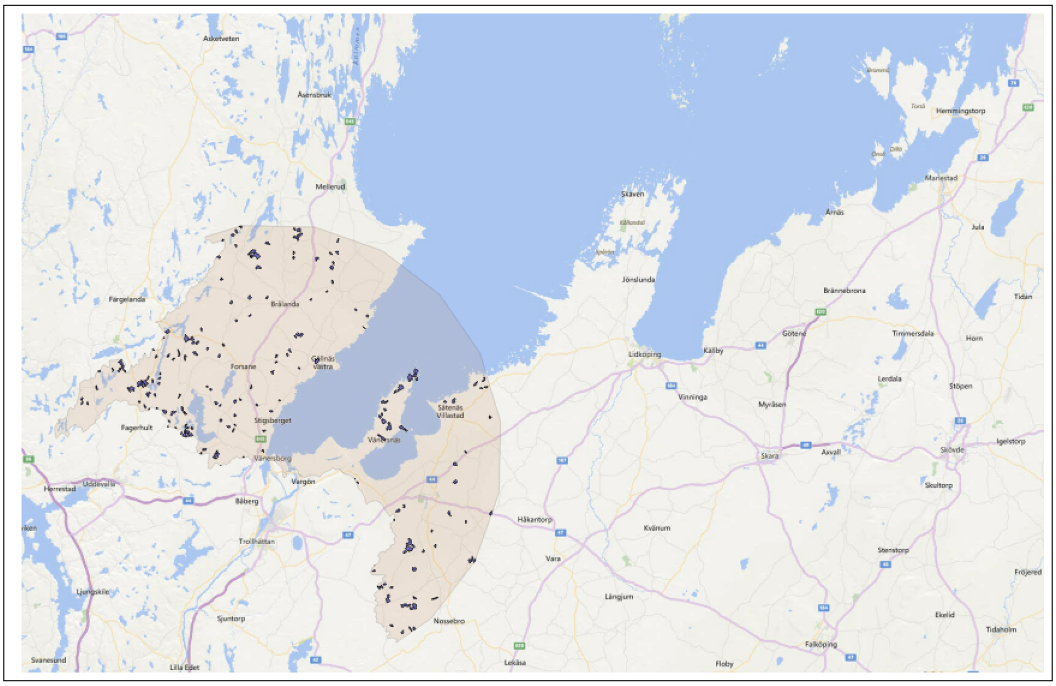

Figure 7 **A layer is added with all land-use changes between 2006 and 2012. Their meta-data, with land-use category before and after the change as well as number of hectares of the area, can be extracted to a spreadsheet.**

B Land-use categories

LEVEL 1	LEVEL 2	LEVEL 3
1. ARTIFICIAL SURFACES	1.1. Urban fabric	1.1.1. Continuous urban fabric 1.1.2. Discontinuous urban fabric
	1.2. Industrial, commercial and transport units	1.2.1. Industrial or commercial units 1.2.2. Road and rail networks and associated land 1.2.3. Port areas 1.2.4. Airports
	1.3. Mine, dump and construction sites	1.3.1. Mineral extraction sites 1.3.2. Dump sites 1.3.3. Construction sites
	1.4. Artificial, non-agricultural vegetated areas	1.4.1. Green urban areas 1.4.2. Sport and leisure facilities
2. AGRICULTURAL AREAS	2.1. Arable land	2.1.1. Non-irrigated arable land 2.1.2. Permanently irrigated land 2.1.3. Rice fields
	2.2. Permanent crops	2.2.1. Vineyards 2.2.2. Fruit trees and berry plantations 2.2.3. Olive groves
	2.3. Pastures	2.3.1. Pastures
	2.4. Heterogeneous agricultural areas	2.4.1. Annual crops associated with permanent crops 2.4.2. Complex cultivation patterns 2.4.3. Land principally occupied by agriculture, with significant areas of natural vegetation 2.4.4. Agro-forestry areas
3. FOREST AND SEMI-NATURAL AREAS	3.1. Forests	3.1.1. Broad-leaved forest 3.1.2. Coniferous forest 3.1.3. Mixed forest
	3.2. Scrub and/or herbaceous associations	3.2.1. Natural grassland 3.2.2. Moors and heathland 3.2.3. Sclerophyllous vegetation 3.2.4. Transitional woodland-scrub
	3.3. Open spaces with little or no vegetation	3.3.1. Beaches, dunes, sands 3.3.2. Bare rocks 3.3.3. Sparsely vegetated areas 3.3.4. Burnt areas 3.3.5. Glaciers and perpetual snow
4. WETLANDS	4.1.Inland wetlands	4.1.1. Inland marshes 4.1.2. Peat bogs
	4.2.Marine wetlands	4.2.1. Salt marshes 4.2.2. Salines 4.2.3. Intertidal flats
5. WATER BODIES	5.1. Inland waters	5.1.1. Water courses 5.1.2. Water bodies
	5.2. Marine waters	5.2.1. Coastal lagoons 5.2.2. Estuaries 5.2.3. Sea and ocean

Figure 8 **The different classes for land-use.**