Controlling omitted variables and measurement errors by means of constrained autoregression and structural equation modeling
Suparman, Yusep

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 1  Problem Statement

1. Introduction

Initiated by the Declaration of the International Clean Drinking Water Supply Decade (ICDWS) 1981-1990 and continued by the United Nations’ Millennium Development Goals (MDG), parties responsible for promoting health in developing countries have focused on improving domestic water supply and basic sanitation. Particularly, the MDG drinking water target is to halve the proportion of people without sustainable access to safe drinking water by 2015. This target is translated into the International Decade for Action, Water for Life, 2005-2015. The action is designed for combating two major challenges of meeting the MDG target for drinking water. The first challenge is to keep up the current coverage levels of safe drinking water in urban areas with the pace of urbanization. The second is to reduce the vast coverage gap between urban and rural populations with a huge backlog of rural people without safe drinking water (WHO-UNICEF, 2006).

Indonesia, as one of the targeted countries, has actively participated in improving its domestic water supply. In the line with ICDWS, during the 1980s the Government of Indonesia (GoI) used a technology oriented approach for improving access to safe water supply. In this period, low income communities were considered to be suitably served by low-cost affordable technologies such as hand pumps or public taps. In-house piped water was considered to be unaffordable technology. Unfortunately, this approach led to disappointment. Many of these water facilities were under-used and poorly maintained (DFID, 1999). Next, in the line with the Dublin Principle: “water has an economic value in all its competing uses and should be recognized as an economic good”, the GoI introduced an economic driven approach to improving access to safe water. That is,
domestic water provisions were designed according to communities’ preferences. Furthermore, the GoI adopted the full cost recovery principle for piped water provision (Ministry of Home Affair Decree No. 23/2005).

Currently, 82% of the Indonesian population is served by improved water supply. The following types are common in Indonesia (WHO-UNICEF 2015):

- Piped water. This type encompasses the following three categories:
  - Piped water into dwelling via a water service pipe connected with in-house plumbing to one or more taps (e.g. in the kitchen or bathroom).
  - Piped water to yard or plot connected to a tap placed in the yard or plot outside the house.
- Public tap or standpipe is a public water point from which people can collect water.
- Tube-well or borehole is a deep hole that has been bored to reach groundwater supplies. The well is constructed with casing or pipes which prevent the hole from caving in and protect the water source from infiltration by run-off water.
- Protected dug well or spring is a well or spring that is protected from run-off, bird droppings and animal contact. Wells are constructed with a lining that is raised from the ground and is covered. Springs are protected by a spring box which is constructed of brick, masonry or concrete around the spring. The water from the spring flows directly out of the box into a pipe.
- Rain water collected from roof or ground catchment and stored in a container.

However, the prevalence of water borne diseases is still high in Indonesia. For instance, annually about 30.3 thousand deaths are estimated to be associated with diarrhea incidence due to the consumption of contaminated water (WHO, 2012). In addition to diarrhea, there are various other kinds of illnesses related to water quality such as
schistosomiasis, trachoma, ascariasis, trichuriasis and hookworm (Prüss et al. 2002). In total, poor sanitation and hygiene contribute to 120 million disease incidents and 50 thousand premature deaths annually in Indonesia (Pinto 2013).

The high prevalence of waterborne diseases relates to the quality of water extracted from lakes and rivers, but also from wells and springs. Particularly, current rapid environmental degradation has deteriorated the quality, as well as decreased the availability, of natural clean water. For instance, insufficient sanitation has contaminated surface water, particularly river water. A typical example is the Citarum River in West Java. In the upstream river basin, about 40% of the six million citizens use septic tanks and only about 500,000 people (about 8%) are served by integrated wastewater treatment installations. Consequently, about 40% of the Bio-Oxygen Demand burden produced in the region every day is discharged into the river. This condition has very serious health and economic impacts (BPLHD Jabar, 2001). Water from wells and springs has deteriorated due to infiltration of contaminated surface water. Under these conditions of deteriorated water quality, extension of piped water into dwellings or yards (denoted in-house piped water in the sequel) is the most effective mode to improve health in developing countries like Indonesia (Hutton and Haller 2004).

Currently, only 20% of the Indonesian population is served by in-house piped water service. The distribution between rural and urban Indonesia, however, is skewed. Of the rural population only 8% is served while the service covers 36% of the urban population (WHO-UNICEF 2013). A similar gap is found in other developing countries. For example, in other South Eastern Asian countries, on average 53% of the urban and 13% of the rural population is served by in-house piped water (WHO-UNICEF 2013).

The gap between urban and rural water provision in Indonesia and other developing countries is related to the prevalent policy of domestic water supply.
Particularly, the GoI has concentrated in-house piped water investments mainly in urban areas. In rural areas, it has promoted alternatives, such as public wells, taps or hydrants. For instance, the rural water and sanitation project, ‘PAMSIMAS\(^1\), focuses on public taps and hydrants in 5000 villages only, which is a small fraction of the 79075 villages in rural Indonesia. Furthermore, of the 4,217 piped-water connections in 2012, only 218 were in-house connections. Most of the project was dedicated to the provision of public taps (3467 units) and public hydrants (532 units) (PAMSIMAS 2013).

The prevalent policy is supported by some studies which generally report that urban households have stronger preferences for in-house piped water than rural households. In the case of urban areas, McPhail (1993) found that households in several Moroccan cities are willing to pay 7 to 10% of their expenditure for in-house water connection. Nauges et al. (2009) found that households in three cities in El Salvador and four cities in Guatemala are willing to pay 1 to 5% of their monthly real income for tap water connection. De Oca et al. (2003) reported that the willingness to pay for in-house piped water connection of households in Mexico City is about 5.8 to 8.2% of their monthly income. Whittington (1991) reported that households in the city of Onitsha spent at least 5% of their income on water.

In the case of rural areas, North and Griffin (1993) found that households in the Bicol Region–Philippines were willing to pay at most 2.4% of their income for in-house piped water connection. Mbata (2006) reported that the willingness to pay for in-house piped water connection of households in Kanye-Botswana was statistically insignificant from zero. For Indonesia, Yusuf and Koundouri (2004) found that urban households

\(^1\) PAMSIMAS (2013) is an Indonesian Government project aimed at halving the number of households without access to drinking water and basic sanitation, as determined by the UN MDGs minimum standards. The focus of the project is on villages in rural areas, particularly those with a high poverty rate. The project finances both physical (e.g. public tap and hydrant facilities) and non-physical (e.g. community empowerment and local institutional development) investments.
were willingness to pay 3.6% of their expenditure for piped water connection while the rural household WTP was insignificant from zero.

There is growing (anecdotal) evidence that the willingness to pay for piped water in rural areas in Indonesia is increasing. Particularly, rural households in Indonesia increasingly choose vended water for consumption. For example, the Indonesia Family Life Survey (IFLS) (Straus et al. 2009) report an increase of purchases of vended water in rural areas from 0.8% in 1997 to 5.8% in 2007. In addition, WHO-UNICEF (2013) reports that more rural households use bottled water as drinking water (i.e. 0.3% in 2002 to 8.8% in 2012). Compared to in-house piped water, vended water and bottled water are more expensive, its quality is worse and availability is less.\(^2\) Hence, the increase in purchases might indicate a change in the preference for in-house piped water in rural Indonesia. However, more research on possible differences in preferences between rural and urban areas is needed. The research results may be relevant for the GoI’s domestic water provision policy in that in-house piped water provision is to be expanded in both urban and rural areas.

2. Measuring preference for in-house piped water

Basically, individual or household preferences can be measured by two classes of valuation approaches i.e. indirect and direct methods (Haab and McConnell, 2002). The former infers an implicit value for an unpriced good from observable prices of market

\(^2\) Piped water meets quality standards set and controlled by the government. Although in some areas, piped water may not be fully available for 24 hours, in principle it is distributed such that it meets daily needs. In contrast, for vended water that is usually distributed by chart or truck, no quality standards apply. It is often directly extracted from wells and contaminated at delivery. Its distribution is dependent on vendors’ schedules. In some cases, it is to be ordered.
goods or services\(^3\) that are related to—notably consumed together with - the good whose value is to be estimated. The latter aims at revealing the value of an unpriced good without the intervention of related goods. Surveys are used, where individuals are asked to state their preference for or valuation of the good or service (Shechter, 2000).\(^4\) The approach invokes the framework of a hypothetical market. The most common direct approach is the contingent valuation (CV) method.

In the case of domestic water supply, house rent or price reflects differentials such as the availability and quality of water provision (North and Griffin, 1993; Yusuf and Koundouri, 2004). By using data on house rent or price for houses that differ in terms of water provision and controlling for differences in other housing attributes, the contribution of water supply to the rent or price of a house can be estimated. That is, one basically identifies a shadow price for the attribute of water supply. This approach is called the hedonic price method. In this study, I adopt this method to measure the willingness to pay for in-house piped water. The main reason for this choice is that for both urban and rural areas, panel data on housing rent price and housing characteristics is available in the Indonesia Family Life Survey (IFLS)\(^5\) data sets (Strauss et al. 2009)\(^6\).

---

\(^3\) Valuation methods have been developed to measure preferences for non-market public good in the first place. However, they can also be applied to infer the value of publicly provided market goods like in-house piped water that do not carry a price.

\(^4\) For a comparison of pros and cons of direct and indirect methods, see inter alia Shechter (2000), Smith et al. (1986) and Bijlenga et al. (2011).

\(^5\) The IFLS reports imputed rent rather than observed rent. See chapter 2 for a discussion of the use of imputed rent.

\(^6\) Another approach used to value the availability and quality of water supply based on market behavior is the Cost of Illness (COI) method. It estimates the economic loss caused by water related illnesses by assessing the total medical costs and productivity losses associated with those illnesses (Shechter, 2000). Because water availability and quality are risk factors of waterborne illness and households spend resources on treating the illnesses (Roberts, 1999; Shechter, 2000), estimates of the benefits from water availability and quality improvement can be obtained from the amount that could potentially be saved or gained from the risk reduction of the related illnesses. Another indirect approach is the household production function (HPF). It is based on household behavior (“production”) in preventing health damage due to pollution. It is considered to be superior to COI but in practice more difficult to apply than COI (Shechter, 2000).
The hedonic price (HP) model is based on the notion that a market good is composed of a bundle of attributes. For instance, a house is made up of structural attributes like size, number of rooms, wall and floor material, availability of facilities like toilets and electricity, and neighborhood characteristics like vicinity of schools, parks, shopping centers, safety, neighborhood wealth, and so on. The price of a good is a function of the prices of the attributes. Based on this notion, the HP approach decomposes the price of a good like a house into the prices of its attributes.

The foundation of HP modeling was developed by Rosen (1974) as follows. Let \( \mathbf{z} = (z_1, z_2, \ldots, z_n) \) be a vector of characteristics of the good of interest, say a house, \( n \) the number of characteristics and \( p(\mathbf{z}) \) the hedonic price function under equilibrium conditions. In the (housing) market, the hedonic price function emerges from competitive bidding among (home) owners and buyers or renters. For a household with preference function \( u(x, \mathbf{z}) \) and budget constraint \( y = p(\mathbf{z}) + x \), with \( x \) is a composite bundle with price 1, the optimality condition for each attribute is:

\[
\frac{\partial u(x, \mathbf{z})}{\partial z_i} = \lambda \frac{\partial p(\mathbf{z})}{\partial z_i}, \quad i = 1, 2, \ldots, n,
\]

where \( \lambda \) is the marginal utility of income. From (1) the marginal willingness to pay for the \( i^{th} \) characteristic is:

\[
\lambda^{-1} \frac{\partial u(x, \mathbf{z})}{\partial z_i}.
\]

From (1) and (2) it follows that estimation of the hedonic price function yields the implicit prices of housing characteristics (e.g. of piped water) from which the marginal WTPs for those characteristic can be derived.

HP models have been widely used to value both public and market goods. The first application of HP modeling dates back to Waugh (1926) who analyzed the impacts
of quality factors on vegetable price. Since the mid 1960’s application of HP modeling started increasing. For instance, Ridker and Henning (1967), Nourse (1967), and Barrett and Waddell (1970) estimated the effect of air pollution on housing prices while Dornbusch and Barrager (1973), and Gamble et al. (1973) analyzed the impacts of water pollution and highway noise on property values. Application of housing HP models has recently grown in popularity. For example, Ioannides (2002) evaluated the demand for houses in relation to neighborhood characteristics. Rehdanz and Maddison (2005) valued climate in Germany whereas Skinner (1989) and Engelhardt (1996) analyzed the relationship between saving behavior and house price. In the United States, Housing Price Indices are often obtained from HP models (Keil and Zabel, 1997). Finally, three previous studies on domestic water supply in developing countries (North and Griffin, 1993; Yusuf and Koundouri, 2004 and Anselin et al., 2008) also used the HP method.

Although housing HP studies have become popular and are routinely applied, they frequently encounter under-specification problems, i.e. the variables that actually belong to the true or population model are missing. Specifically, let the price of house $i$ at time $t$ ($p_{it}$) be determined by $a + b$ systematic house characteristics $q_{it1} \ldots q_{it(a+b)}$ according to the linear function:

$$p_{it} = \pi_{0i} + \sum_{j=1}^{a+b} \pi_{jit} q_{jit} + \omega_{it}, \quad (3)$$

with $\pi_{0i}$ the intercept, $\pi_{jit}$ the marginal price for the $j^{th}$ characteristic, and $\omega_{it}$ an independent-identically-distributed (iid) error term for which the zero conditional mean assumption holds, i.e. the expected value of the error term does not depend on the characteristics. If $b$ characteristics which are correlated with the $a$ characteristics are omitted, model (3) reduces to

$$p_{it} = \pi_{0i}^{*} + \sum_{j=1}^{a} \pi_{jit}^{*} q_{jit} + \omega_{it}^{*} \quad (4)$$

which is underspecified.
Under-specification may be due to data collection problems, such as time or resource constraints (Clarke 2005) or to methodological considerations. For instance, the inclusion of all relevant characteristics in a housing HP model increases multicollinearity. Accordingly, in practice several of the systematic house characteristics are often omitted from the model. This practice leads to omitted variable bias (Greene 2003; Butler 1982; Ozanne and Malpezzi 1985). Particularly, the estimators of the parameters of the included variables are biased in a complicated fashion. In addition, the directions of the biases are difficult to assess, since they depend on the correlations among the included and omitted variables, among the explanatory variables and on the signs of the impacts of the explanatory variables on the dependent variable (Greene 2003). Note that although researchers have started considering spatial spill-over and heterogeneity in HP models to increase prediction accuracy (Páez 2007; Pace and Lesage 2004), little attention is still being paid to omitted variables.

3. Objectives

As explained in section 1, the main overall purpose of this thesis is to shed further light on the preferences for in-house piped water in rural and urban Indonesia. In addition, in section 2, we addressed a major obstacle to obtain unbiased and efficient estimators of hedonic models to analyze preferences for piped water, viz. omitted variables. These two themes have led to the following 4 research objectives of this thesis.

1. Introduction of constrained autoregression (CAR) and structural equation modeling (SEM) as methods to control time-varying omitted variables and measurement error in a panel data model.

Omitted variables bias is corrected via the constrained autoregression option that panel data offers. To account for measurement error I propose structural equation...
models (SEM). A SEM estimated on the basis of panel data makes it possible to repeatedly measure a variable and thus to decompose its variance into a component related to the “true” latent variable on the one hand and measurement error variance on the other. I denote the combined model as the constrained autoregression-structural equation model (ASEM). The performance of ASEM is illustrated and compared to the fixed effect panel data model, SEM only and the autoregressive panel model on the basis of the IFLS urban data set.

2. Monte-Carlo simulation to compare the performance of CAR relative to the commonly used alternative methods to control time varying omitted variables, viz. the latent fixed effect model, first order differencing, demeaning regression, and the autoregressive model.

The simulation is restricted to a large cross sectional sample for the following reasons. First, the focus of the analysis is on bias reduction. A large sample size reduces the standard error and thus provides better insight into each method’s bias reduction potential. Secondly, since maximum likelihood (ML) is used to estimate the models, a large sample is required to achieve its consistency and efficiency properties (Casella and Berger, 2002). Thirdly, many micro panel data sets like the IFLS are based on large cross sectional samples.

3. Monte Carlo analysis of likelihood ratio (LR) test of CAR in the presence/absence of omitted variables.

Specifically, the probability of Type I error and the power of the LR test of CAR are analyzed. This objective complements objective 2.

4. Analysis of the preferences for in-house piped water in rural and urban Indonesia.
Hedonic price (HP) models for rural and urban Indonesia will be estimated by means of ASEM on the basis of the IFLS panel data set. Endogeneity of the presence of in-house piped water will be considered. Equality of the WTP for in-house piped water in rural and urban Indonesia will be tested. The implications of the results for the water provision program of the GoI (PAMSIMAS) will be discussed.

4. Outline of the thesis

The thesis is made up of the following chapters.

Chapter 2 entitled

**Hedonic price models with omitted variables and measurement errors: a constrained autoregression–structural equation modeling approach with application to urban Indonesia**

In this chapter, I develop a constrained autoregression - structural equation model (ASEM) to control both measurement errors in explanatory variables and omitted variables. The first issue is frequently ignored in standard models including panel data models while standard approaches to control it require additional external information which is usually difficult to obtain. The latter problem, if considered, is usually handled by taking the omitted variables as time-invariant. In the chapter, I show that ignoring measurement error and inappropriate handling of omitted variables lead to biased estimators. I furthermore show that ASEM allows handling of both time-varying and time-invariant omitted variables by constrained autoregression. Regarding measurement error, ASEM exploits the fact that panel data are repeatedly measured which allows decomposing the variance of a variable into the true variance and the variance due to measurement error. I apply ASEM to estimate a
hedonic housing model for urban Indonesia. To get insight into the consequences of measurement error and omitted variables, I compare the ASEM estimates with the outcomes of (i) a standard SEM, which does not account for omitted variables, (ii) a constrained autoregression model, which does not account for measurement error, and (iii) a fixed effects hedonic model which ignores measurement error and time-varying omitted variables.

Chapter 3 entitled

Controlling for time-varying omitted variables in panel data models: evidence from a Monte-Carlo simulation

This chapter presents evidence from a Monte-Carlo simulation study to control time-varying omitted variables in panel data models by means of latent fixed effects regression, demeaning, first order differencing, autoregression and constrained autoregression. The data are generated from a standard regression model with three explanatory variables, three time periods and one thousand cross sectional units. The bias, the standard error and the mean squared error of the estimators of the regression coefficients of the included variables, are analyzed to compare the performance of constrained autoregression compared to the alternative omitted variables correction procedures. By means of regression of the log-transformed absolute bias on the log-transformed simulation parameters I identify the main determinants of the bias of each correction procedure.

Chapter 4 entitled

Controlling for omitted variables by means of constrained autoregression: testing the constraints
This chapter presents evidence from Monte-Carlo simulation on Type I error probability and the power of a likelihood ratio test of constrained autoregression (CAR) as vehicle to control a time-varying omitted variables.

Chapter 5 entitled

The willingness to pay for in-house piped water in urban and rural Indonesia

This chapter analyzes household preferences for in-house piped water in urban and rural Indonesia on the basis of the Indonesia Family Life Survey data set by means of a hedonic price model, specified as a constrained autoregression-structural equation model (ASEM). Based on the analysis policy recommendations for further investment in in-house piped water will be formulated.

Chapter 6 entitled

Conclusions, policy recommendations and suggestions for further research

This chapter summarizes the main results of the thesis and formulates several policy conclusions. It also presents some suggestions for further research.

References


BPLHD Jabar (Regional Agency for Environmental Control, West Java), 2001, Strategi Pengendalian Pencemaran Air Melalui Prokasih di 14 Kab/Kota (Strategy of Water Pollution Control by Prokasih in 14 Regions, BPLHD Jawa Barat, Bandung.)


De Oca GSM, Bateman IJ, Tinch R, Moffatt PG (2003) Assign the willingness to pay for maintained and improved water supply in Mexico City. Working paper No ECM 03-11. CSERGE.


Dornbusch DM, Barrager SM (1973) *Benefit of water pollution control on property values*. Report, EPA-600/5-73-005, US Environmental Protection Agency


Pinto R (2013) Results, Impacts, and Learning from Improving Sanitation at Scale in East Java, Indonesia. Water and Sanitation Program: Field Note, World Bank


Waugh F (1926) Quality factor influencing vegetable price. *J Farm Econ* 10:185-196