Journal of Environmental Management xxx (2016) 1–9

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: www.elsevier.com/locate/jenvman

Research article

Forecasting urban water demand: A meta-regression analysis

Maamar Sebri

Faculty of Economics and Management of Sousse & LaREMFiQ Research Laboratory, University of Sousse, Tunisia

ARTICLE INFO

Article history:
Received 21 May 2016
Received in revised form
7 September 2016
Accepted 10 September 2016
Available online xxx

Keywords:
Meta-analysis
Forecasting
Urban water demand
Accuracy
Mean absolute percentage error (MAPE)
Box-Cox transformation

ABSTRACT

Water managers and planners require accurate water demand forecasts over the short-, medium- and long-term for many purposes. These range from assessing water supply needs over spatial and temporal patterns to optimizing future investments and planning future allocations across competing sectors. This study surveys the empirical literature on the urban water demand forecasting using the meta-analytical approach. Specifically, using more than 600 estimates, a meta-regression analysis is conducted to identify explanations of cross-studies variation in accuracy of urban water demand forecasting. Our study finds that accuracy depends significantly on study characteristics, including demand periodicity, modeling method, forecasting horizon, model specification and sample size. The meta-regression results remain robust to different estimators employed as well as to a series of sensitivity checks performed. The importance of these findings lies in the conclusions and implications drawn out for regulators and policymakers and for academics alike.

1. Introduction

Concerns about water scarcity have placed the integrated water resources management on the global agenda. In many regions over the world, satisfying the growing demand for water is encountered by a number of difficulties. Therefore, an increasingly attention is being paid to identifying the best practices towards an overall and sustainable strategy for its use. This necessarily requires, among others, a reliable water demand forecast for the short-, medium- and long-run for different sectors, especially the residential one as it is straightforwardly related to human existence (Sebri, 2013). According to Da Cunha (1983), a water demand forecast is not an end in itself, but an input for decision making. While the short-run water forecasting is useful in operations and management, the long-run forecasting is mandatory for planning and design of water supply (Herrera et al., 2010). Hall et al. (1989, p.3) have pointed out that: “the success of any water resource development is critically dependent upon the reliability of the forecasts of future water demands that are employed in its design”. Thus, the forecasting of urban water demand has been the object of an increasing literature and dates back to the 1960s with the studies of Linaweaver et al. (1967), Howe and Linaweaver (1967), Howe (1968) and White (1969). Since that, a growing sophistication as regards to methods and evaluation tools has been raised in an attempt to improve forecast accuracy and reliability.

More recently, Donkor et al. (2014) have performed a qualitative literature review on the urban water demand forecasting. They have considered studies published between 2000 and 2010 and have reported that some methodological differences, such as forecasting models, explanatory variables included, and forecasting horizon are likely to affect urban water demand forecast. Though this survey constitutes a nice descriptive contribution on summarizing the existing literature, it does not explore the outcomes of empirical studies in a systematic way. That is, the present paper attempts to complement this survey by conducting a quantitative literature review using the meta-analysis approach. Called also quantitative research synthesis and the analysis of the analyses, the term meta-analysis was first introduced by Glass (1976) in the social science literature. The author defined it as “the statistical analysis of a large collection of results from individual studies for the purpose of integrating the findings. It connotes a rigorous alternative to the casual, narrative discussions of research studies which typify our attempt to make sense of the rapidly expanding research literature.” (Glass, 1976, p. 3).

The advantages of the meta-analysis are numerous and mostly come as a remedy to the criticism addressed to the ordinary narrative literature review. Most importantly, it has an exhaustive character, i.e., in the process of collecting studies the meta-analyst tries to include all the empirical studies that present sufficient data to perform a statistical analysis. On the contrary, the sample of

E-mail address: maamer.sebri@gmail.com.

http://dx.doi.org/10.1016/j.jenvman.2016.09.032
0301-4797/© 2016 Published by Elsevier Ltd.

studies selected for a traditional literature review is generally based on the author's whim, which bears an important bias of omitting some core studies and generates a high risk of subjectivity. A second important advantage is the statistical efficiency gain from pooling primary studies. In fact, the meta-analysis approach may yield conclusive and meaningful results by combining inconclusive results (Poot, 2014). Third, while in the traditional literature review authors encounter some difficulties in comparing and drawing concise conclusions because of the use of different specifications, data and methodologies, the meta-analysis is based on a statistical procedure called meta-regression that facilitates the synthesis of empirical studies and explains the reasons behind estimates variation of outcomes (Florax et al., 2002).

These advantages and others make the meta-analysis approach a popular, versatile and powerful tool of quantitative literature surveys. It has been picked up in various research fields, including economics where it has been successfully applied in various economic areas. In particular, a large number of meta-analyses have been conducted in the environmental and natural resources economics.[1] In the case of water resources, meta-analyses range from studying the water demand elasticities (Dalhuisen et al., 2003; Espey et al., 1997; Sebri, 2014) to willingness to pay for water quality (van Houtven et al., 2007). On the other hand, to the best of our knowledge, the current study is the first meta-analysis that aims to quantitatively identify explanations of the urban water demand forecasting accuracy.

This paper is set out as follows. Section 2 details the modeling approach, moderator variables definition and their descriptive statistics. The next section presents the data collection procedure and descriptive statistics for the effect size per individual study. Section 4 analyses the results of the meta-regression, while Section 5 offers a series of sensitivity checks. The study ends with conclusions delivering some policy implications.

2. Meta-regression methodology

Following the meta-analysis approach, the objective of the current study is to provide a comprehensive and systematic review of empirical studies dealing with the urban water forecasting subject. In particular, it seeks to explain the leading factors behind the variation in forecasting performance measures computed by authors in their studies. However, many statistical accuracy metrics exist. In particular, the mean error (ME), the mean square error (MSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean square error (RMSE) are commonly employed in the water demand forecasting literature. Each of these measures has its advantages over others. Nevertheless, only few metrics such as MAPE, root mean square percentage error (RMSPE) and Theil's *U*-statistic have the advantage of being used for accuracy comparison purposes across different locations and periods because they are scale-independent. In particular, MAPE has the advantage as an absolute measure of forecasting accuracy, so two forecasting methods can display similar performance with respect to other metrics but their MAPEs values can be significantly different (Homwongs et al., 1994). “By exploiting it, it is easy to compare the efficiency of forecasts independent of the absolute values of the considered time series” (Froelich, 2015a, p.336).

Therefore, in the current meta-regression analysis and based on the reasons mentioned above, only MAPE will serve as our effect size. In fact, there was also an intention to use other scale-independent metrics (e.g., RMSPE or Theil's *U*-statistics) but they had been very hardly used in the urban water forecasting literature and therefore the corresponding number of observations is very limited and does not allow performing any complete statistical analysis.

To fulfill the objective of this meta-analysis, the following equation is estimated:

$$\ln MAPE_{ij} = \alpha_0 + \sum_{k=1}^{K} \alpha_k X_{ijk} + \varepsilon_{ij} \quad (1)$$

where $MAPE_{ij}$ is the ith MAPE estimate from the jth study. α_0 is the intercept, while α_k denote the meta-regression coefficients that reflect the impact of moderator variables, X_{ijk}. ε_{ij} is the usual error term assumed to be normally distributed with zero mean and varianceσ_ε^2.

For the sake of synthesis, the moderator variables are grouped into five groups. Table 1 presents their definition and main descriptive statistics along with the number of studies fits into the different categories.[2] By employing these variables within the above meta-regression equation, this study seeks to respond to the following core questions:

i. Does the demand periodicity affect the magnitude of forecasting measures?
ii. Are differences in forecasting measures related to the forecasting method?
iii. Is heterogeneity in forecasting measures sensitive to the model specification? And if so, which factors matter?
iv. Does the forecasting horizon that ranges from short-run to long-run has a systematic impact on the reported magnitude of forecasting measures?
v. Are some study-specific characteristics, such as sample size, publication year and development level of the country on which the study was performed, explaining the variation in the estimated forecasting measures?

In estimating meta-regression models, many issues related to data structure have been revealed in the literature. This makes relying on the simple ordinary least squares (OLS) less efficient. According to Nelson and Kennedy (2009), three main issues may arise: heteroscedasticity, heterogeneity and non-independence of observations. First, the problem of heteroscedasticity is mainly attributed to the fact that effect size estimates employed in the meta-regression analysis are derived from empirical studies that are usually based on different sample sizes. Second, the risk of data heterogeneity comes essentially from the fact that primary studies include different explanatory variables, use various functional forms and rely on different estimation techniques. Third, non-independence or correlation of observations occurs because the meta-analyst generally use all the reported effect size estimates from each primary study. In addition, this issue may arise when two or more primary studies rely on the same sample size (Nelson and Kennedy, 2009).

In the current meta-analysis, to ensure the robustness of empirical results and to deal with issues discussed above, Equation (1) is estimated using three estimators, namely the robust OLS, the weighted least squares (WLS) and the random effect maximum likelihood (REML).[3] The robust OLS is advocated by to the results of Breush-Pagan heteroscedasticity test. In fact, findings presented in

[1] For further references, reader would refer to the survey of Nelson and Kennedy (2009) and the meta-analysis of Sebri (2015).

[2] The different forecasting methods, forecasting horizons, and demand periodicity are described with more details in Donkor et al. (2014).

[3] For a detailed discussion, reader may refer, for example, to Benos and Zotou (2014) and Sebri (2014).