WATER GLOBAL PRACTICE • WSS GSG UTILITY TURNAROUND SERIES

Statistical Analysis

Global Study on the Aggregation of Water Supply and Sanitation Utilities

AUGUST 2017

Michael Klien

WORLD BANK GROUP

About the Water Global Practice

Launched in 2014, the Word Bank Group's Water Global Practice brings together financing, knowledge, and implementation in one platform. By combining the Bank's global knowledge with country investments, this model generates more firepower for transformational solutions to help countries grow sustainably.

Please visit us at www.worldbank.org/water or follow us on Twitter at https://twitter.com/search?q=%40WorldBankWater&src=tyah.

Statistical Analysis

Global Study on the Aggregation of Water Supply and Sanitation Utilities

AUGUST 2017

Michael Klien

WORLD BANK GROUP

1818 H Street NW, Washington, DC 20433

Telephone: 202-473-1000; Internet: www.worldbank.org

Cover design: Jean Franz, Franz & Company, Inc.

Contents

Figures

Tables

Executive Summary

By definition, aggregation implies a change in the structure of involved utilities by combining several preexisting utilities into an integrated organization. To analyze and quantify the performance consequences of aggregations, one must have an understanding of the relationship between utility structure and performance and of the way aggregations change utility structure.

Structure is more than size. A one-dimensional description of utilities is too narrow to describe key aspects of utility structure. The data-driven framework in this report suggests that apart from pure size-related output indicators (such as volume or number of customers), utilities need to be differentiated according to density and the number of towns served.

Using the three dimensions—customers, density, and towns—to describe utility structure, the universe of utilities in the IBNET database can be grouped into a small number of clusters of similar utilities. Both the clustering and also the findings regarding the relationship of structure with performance and utility input structure show that there is a divide between utilities that serve a single town and utilities that serve several towns.

For utilities that serve a single town with water and wastewater, larger volumes and density are positively related to performance in terms of lower unit cost and higher quality of service. These results do not carry over as clearly to utilities that serve multiple towns. Because aggregations will tend to move utilities from serving a single town to serving several towns, the performance consequences are much less clear than in the simple case of growing a utility that serves a single town.

In addition, although utilities serving a single town experience reductions in the share of labor costs as they grow, utilities serving several towns exhibit less clear patterns and seem to incur additional transaction costs. This observation suggests that if aggregations are simply a process to transform utilities from one structural setup to another, the outcome will depend on not only (a) the initial structure before aggregation, but also (b) the way the aggregation changes the structure.

Given that utility structure is a key to understanding how utility performance might change because of aggregations, the longitudinal data in IBNET enables researchers to analyze how aggregations change the three dimensions of utility structure. At least for the sample of aggregations examined in IBNET, the number of towns increased (by definition of aggregation), but the aggregations added few customers and in many cases reduced density. This result corresponds to the findings in a number of previous studies that suggest that density losses prohibit economies of scale. Another contrast to the case of growing a utility that serves a single town is that aggregating utilities has not been found to decrease the share of labor cost over time. These findings raise the following question: Through which channels can the aggregations in principle and in practice improve the performance of utilities?

The causal analysis also confirms these impressions: On average, the analyzed aggregations have had no effect on cost and various other performance indicators when compared with similar utilities that did not aggregate. However, the analysis also indicates that the results often depend on the initial structure of the utility and the design of the aggregation. Regardless of the fact that the conditionalities may be very situation specific, the findings stress that the design and structure of the affected utilities may be more important than the question of whether or not to aggregate. And although it is difficult to directly deduce a recipe for successful aggregations from these results, there is little doubt that a careful analysis of the existing and targeted utility structure is a prerequisite to managing expectations and making the most of an aggregation reform.

Chapter 1
Introduction

The ultimate goal of this report is to empirically assess the performance consequences of aggregations. To answer this question in a meaningful way, it is crucial to understand what aggregations are and how they change the structure of water utilities. Although technically aggregations are loosely defined as "the process by which two or more WSS service providers consolidate some or all their activities under a shared organizational structure,"[1] from an organizational perspective, aggregations can also be seen as a transformation process that changes the structure of the involved utilities along various dimensions. As a consequence, the effect of an aggregation will depend largely on how it changes the structural characteristics of a utility.

The focus on utility structure as the main mechanism through which aggregations could affect performance is warranted by two facts. First, already the definition of aggregations as a merger of several organizational structures implies that the change in structure is a key component of the phenomenon. Second, arguments in favor of aggregation often relate to organizational design. For instance, aggregation reforms are often accompanied by the expectation of achieving economies of scale through increased utility size (see Abbott and Cohen 2009; Carvalho, Marques, and Berg 2012; González-Gómez and García-Rubio 2008; Saal and others 2013; Walter and others 2009). However, as the following sections will show, aggregations affect several dimensions of utility structure simultaneously. Utility size in the sense of volume or customers is too narrow a description of how aggregations actually change utility structure.

For this reason, the first part of this report proposes a framework to describe, classify, and analyze water utilities according to their structural characteristics. Applying a data-driven approach to the universe of IBNET data, we start by identifying a small number of key dimensions of utility structure. These indicators are then used to classify utilities into homogeneous clusters of specific utility types. Finally, the section analyzes the relationship between utility structure and (a) performance measures and (b) the cost structure.

In the second part of the report, the developed framework is used to compare the structure of aggregating utilities (before aggregation) with utilities that do not aggregate. This step is helpful to understanding if certain utility types are more frequently involved in aggregations than others are, a crucial factor in choosing a suitable control group in the ensuing empirical analysis. In addition, the section describes how aggregations change the structure of utilities on the basis of observed aggregations in IBNET. This step should convey a clearer picture of what aggregations mean in terms of utility structure (magnitude and direction of changes) and which aggregation designs appear frequently.

In the final section, we attempt to measure the causal effect of aggregations on utility performance. The approach following Klien and Michaud (2017) is to measure how utility performance evolved for utilities that grew through an aggregation compared with utilities that were not aggregated. Building on the insights from the previous discussion, the effect of the aggregations will be allowed to vary depending on the type of aggregation as well as on the initial structure of a utility. This step should help to explain whether and in which cases the reform design and the structure of the affected utilities matter.

Note

1. The definitions and conceptual basis for large parts of analysis can be found in Michaud and others (2017).

References

Abbott, Malcolm, and Bruce Cohen. 2009. "Productivity and Efficiency in the Water Industry." *Utilities Policy* 17 (3-4): 233-44.

Carvalho, Pedro, Rui Cunha Marques, and Sanford Berg. 2012. "A Meta-regression Analysis of Benchmarking Studies on Water Utilities Market Structure." *Utilities Policy* 21: 40-49.

González-Gómez, Francisco, and Miguel A. García-Rubio. 2008. "Efficiency in the Management of Urban Water Services. What Have We Learned after Four Decades of Research?" *Hacienda Pública Española/ Revista de Economía Pública* 185 (2): 39-67.

Klien, Michael, and David Michaud. 2017. "Diseconomies of Consolidation in Water Utility Mergers: When Economies of Scale Are Not Enough." Unpublished manuscript.

Michaud, David, Maria Salvetti, Michael Klien, Berenice Flores, Gustavo Ferro, and Stjepan Gabric. 2017. *Joining Forces for Better Services? When, Why, and How Water and Sanitation Utilities Can Benefit from Working Together.* Washington, DC: World Bank.

Saal, David S., Pablo Arocena, Alexandros Maziotis, and Thomas Triebs. 2013. "Scale and Scope Economies and the Efficient Vertical and Horizontal Configuration of the Water Industry: A Survey of the Literature." *Review of Network Economics* 12(1): 93-129.

Walter, Matthias, Astrid Cullmann, Christian von Hirschhausen, Robert Wand, and Michael Zschille. 2009. "Quo vadis Efficiency Analysis of Water Distribution? A Comparative Literature Review." *Utilities Policy* 17 (3-4): 225-32.

Chapter 2
A Framework for Water Utility Structure

2.1 Key Dimensions of Utility Structure

Given the vast number of dimensions of utility size and structure, it is hardly surprising that no existing framework is precise and operational enough to allow a classification of water utilities according to very stylized theoretical considerations. In the absence of such a clear theoretical concept, a data-driven approach is applied to identify a small number of distinct measures for utility structure. The focus is to reduce the large number of indicators by discarding indicators that measure similar underlying structural factors.

For the purpose of the underlying study, the definition of utility structure should capture those aspects of a utility's setup (in a very broad sense, and in contrast to the more narrow meaning of size) that could potentially change in an aggregation process. Apart from the increase in the number of towns served by the aggregated utility—which is the definition of an aggregation[1]—it should particularly include factors related to output and supply characteristics. Conversely, it does not include indicators for input choices (share of different cost components) or performance in the sense of economic outcomes (like cost or quality). This exclusion is deliberate and should avoid a conflation of (largely) exogenous structural features with highly endogenous process and managerial choice variables. Starting from a long list of possible indicators, correlation measures as well as principal component analysis (PCA) are used to identify the key structural indicators.

To avoid inconsistent comparisons between utilities offering only water or wastewater services and utilities offering both services, we restrict the data to the latter type of utilities. Utilities providing both water and wastewater services account for more than 80 percent of the observations in the underlying IBNET dataset. Moreover, wastewater-only utilities represent a negligible share (<2 percent), and the results for water-only companies appear very similar.

Using the indicators available in IBNET[2], the following potential measures are considered:

- Volume of water produced (m^3)
- Volume of wastewater collected (m^3)
- Population in the service area for water (#)
- Population in the service area for wastewater (#)
- Number of customers connected to water supply (#)
- Number of customers connected to wastewater services (#)
- Length of water network (km)
- Length of sewer network (km)
- Number of towns served with water (#)
- Number of towns served with wastewater (#)
- Density of water system (equals population connected to water supply/length of water network)
- Density of wastewater system (equals population connected to wastewater services/length of sewer network)

To narrow down the number of indicators, redundant indicators that measure the same underlying structural characteristic are removed step by step. First, the correlations in table 2.1 show that water and wastewater characteristics are generally very highly correlated.[3] Utilities with a large volume of water tend to exhibit a large volume of wastewater. In addition, a PCA on all these variables suggests that water and wastewater characteristics of a particular measure capture the same underlying characteristics (see table 2.2). For instance, the variables on the number of towns served with both water and wastewater load on the same component.

TABLE 2.1. **Pairwise Correlations**

	vol_w	vol_ww	cus_w	cus_ww	popsa_w	popsa_ww	len_w	len_ww	dens_w	dens_ww	towns_w	towns_ww
vol_w	1											
vol_ww	0.887	1										
cus_w	0.956	0.861	1									
cus_ww	0.890	0.917	0.915	1								
popsa_w	0.941	0.845	0.987	0.897	1							
popsa_ww	0.925	0.842	0.968	0.905	0.978	1						
len_w	0.891	0.807	0.920	0.837	0.911	0.879	1					
len_ww	0.825	0.867	0.832	0.892	0.812	0.808	0.855	1				
dens_w	0.428	0.377	0.476	0.447	0.462	0.488	0.0935	0.196	1			
dens_ww	0.366	0.345	0.408	0.479	0.409	0.432	0.191	0.0315	0.607	1		
towns_w	0.250	0.235	0.275	0.223	0.282	0.225	0.311	0.222	0.00209	0.0622	1	
towns_ww	0.255	0.279	0.285	0.275	0.290	0.273	0.310	0.273	0.0282	0.0768	0.848	1

Note: The variables have been transformed by taking the natural log of the original value and then standardizing the variables. vol_w = volume of water produced; vol_ww = volume of wastewater collected; cus_w = customers connected to water supply; cus_ww = customers connected to wastewater services; popsa_w = population of service area for water; popsa_ww = population of service area for wastewater; len_w = length of water network; len_ ww = length of service network; dens_w = density of water system; dens_ww = density of wastewater system; towns_w_= number of towns served with water; towns_ww = number of towns served with wastewater.

TABLE 2.2. **Principal Component Analysis Output**

	Component 1	Component 2	Component 3
vol_w	0.3446793	−0.0449649	−0.0680859
cus_w	0.3529888	−0.04525	−0.0241244
popsa_w	0.3499888	−0.038884	−0.0177919
len_w	0.3261232	0.1137127	−0.2510409
dens_w	0.1650122	−0.3699714	0.5017993
towns_w	0.11942	0.6087636	0.3438152
vol_ww	0.3301522	−0.0223877	−0.0866724
cus_ww	0.3441992	−0.0712156	−0.017246
popsa_ww	0.3469444	−0.0760358	−0.0084673
len_ww	0.3133416	0.081476	−0.317169
dens_ww	0.1529108	−0.3161935	0.5786726
towns_ww	0.129799	0.5966951	0.3436472

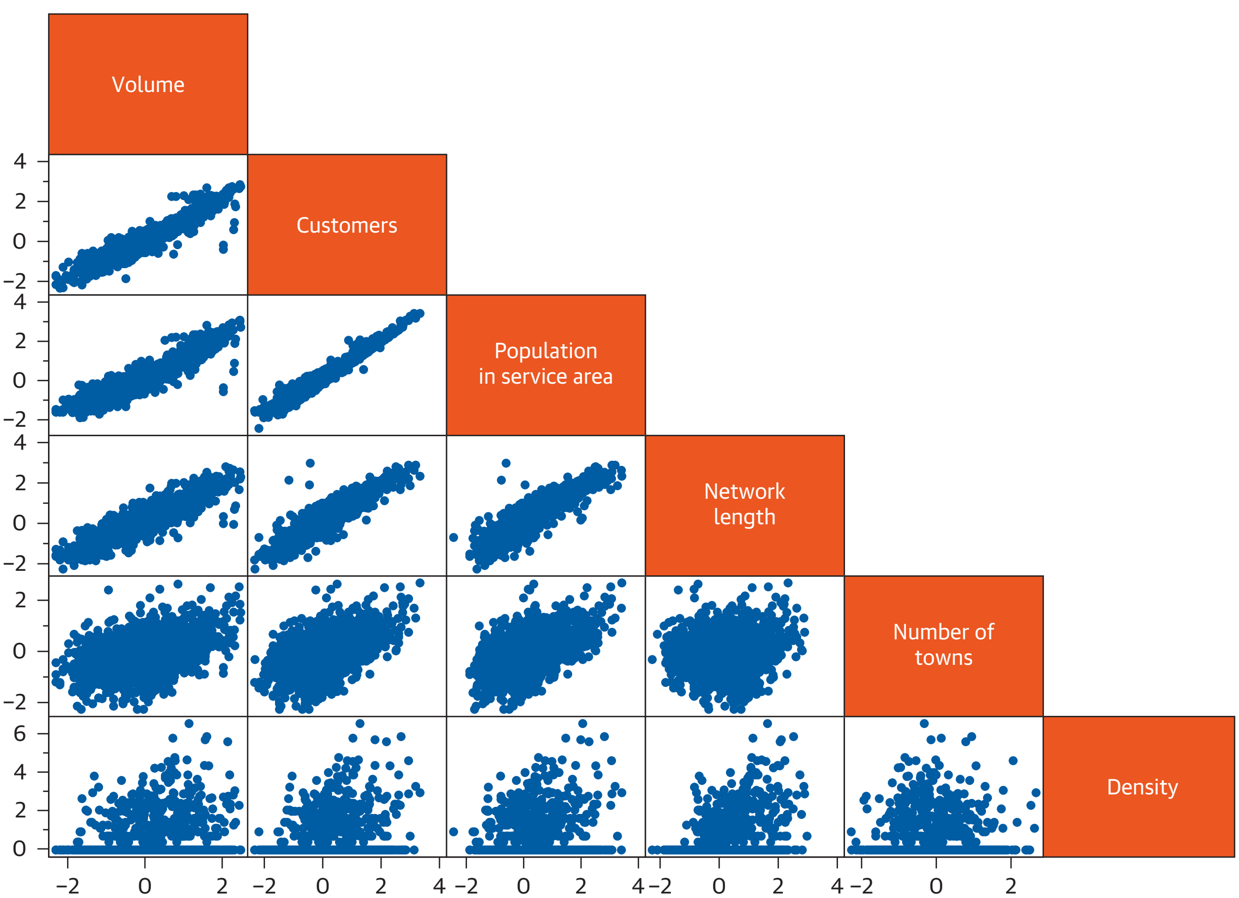
Note: The variables have been transformed by taking the natural log of the original value and then standardizing the variables. vol_ww = volume of wastewater collected; cus_w = customers connected to water supply; cus_ww = customers connected to wastewater services; popsa_w = population of service area for water; popsa_ww = population of service area for wastewater; len_w = length of water network; len_ww = length of service network; dens_w = density of water system; dens_ww = density of wastewater system; towns_w_= number of towns served with water; towns_ww = number of towns served with wastewater.

The components represent the underlying factors. Following a rule of thumb, we choose the number of components depending on eigenvectors being close to or above 1, yielding three components. As a result, the distinction between water and wastewater indicators is dropped, and instead an integrated measure representing the sum of each water and wastewater indicator is used henceforth. This reduces the number of indicators needed to measure utility structure by half.

From the remaining six indicators, four appear to measure a similar structural characteristic that could loosely be interpreted as "size." As the correlations in the scatter plots in figure 2.1 show, the indicators volume, customers, population in service area, and length of the network are very highly correlated—the correlations of these variables vary between 0.90 and 0.98. Bearing in mind that 0 implies no correlation and 1 represents a perfect correlation, the observed correlations are extremely high. Utilities that serve a large volume have many customers, a large population in the service area, and a large network. Also, the previous PCA indicated that these variables load on the same component—that is, they seem to represent similar structural characteristics. Consequently, to further reduce the number of indicators of utility structure, of the four size indicators, only customers is retained for the further analysis.

What remains are three structural indicators of utility structure: the number of customers, density, and the

FIGURE 2.1. Scatter Plot for Structural Characteristics



Note: The variables have been transformed by taking the natural log of the original value and then standardizing the variables.

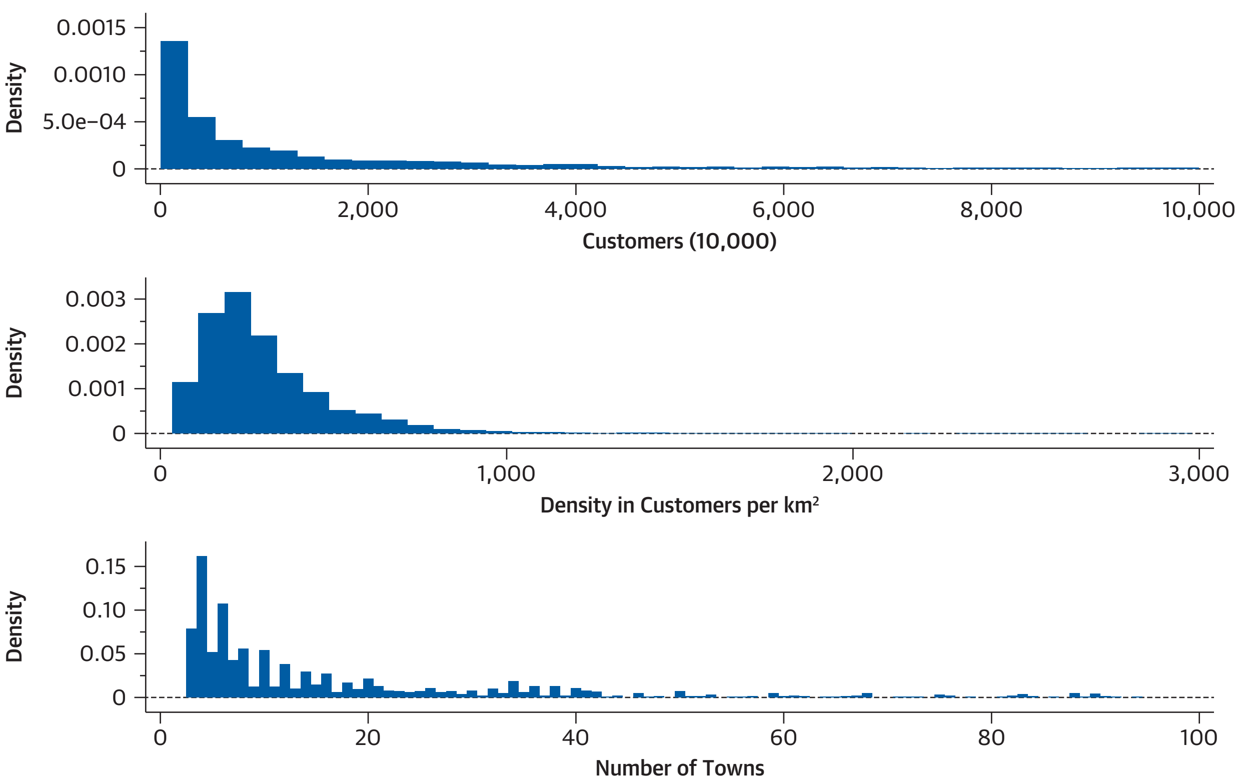
number of towns.[4] It is important to stress that these indicators measure different aspects of utility structure. For example, for a similar number of customers, we observe utilities with a large variation in density and the number of served towns. It also means that with these 3 indicators we are able to describe utilities of widely varying structure without using all 12 initial indicators.

Before we go on to identify utility clusters with similar structural characteristics, it is useful to look at some descriptive statistics for the three chosen structural indicators.

First, for all three indicators, the distribution is (heavily) right skewed, meaning that the number of observations steadily decreases for increasing values of the indicators. In other words, there are many observations with relatively low and moderate numbers of customers, density, or number of towns and only a few observations with very high values.

As shown in the upper and middle panels in figure 2.2, because of the very long tail of the distribution, the median values of customers and density are considerably lower than the average. In the case of density, the median is 252 compared with the mean of 308; for volume the median is 69,000 customers compared with an average of over 383,000. The distribution is even more extreme for the number of towns: although the average is roughly seven towns, more than 80 percent of all observations in the sample serve a single town for water and wastewater (and are thus

FIGURE 2.2. Histograms for Customers, Density, and Number of Towns



Note: In the lowest panel, utilities that serve 2 and more than 100 towns are excluded.

counted as two towns: one for water and one for wastewater). For the remaining utilities that serve more towns, the distribution is also very heavily right skewed. As seen in the lowest panel of figure 2.2, excluding utilities that serve one town for water and wastewater, there are fewer and fewer observations as the number of towns increases.

This demonstrates that although IBNET may not be representative of the whole population of utilities—it likely oversamples larger utilities—there is still considerable variation in the structural dimensions. For example, more than 10 percent of all observations are utilities with a combined number of customers for water and wastewater below 10,000. This fact should ensure that the ensuing description and classification of utility types may be applied beyond the data sample in IBNET.

2.2 Clustering Utilities According to Structural Dimensions

In this section the chosen dimensions of utility structure are used to classify utilities into homogeneous clusters of specific utility types. The first goal is to identify frequently appearing configurations of utility structure. Clustering utilities can be seen as a middle ground between a hard-to-interpret multidimensional representation and an overly simplified one-dimensional description—for instance, a small-versus-large dichotomy. The choice of the number of clusters considers the trade-off between few but very heterogeneous clusters and many not easily distinguishable clusters.[5] Although the final number of six clusters is somewhat arbitrary, using a relatively small number of clusters appears to give an appropriate but meaningful representation of the utilities in IBNET.

The results of the clustering—a combination of hierarchical clustering followed by k-means clustering—are shown in figure 2.3. The box plots are particularly useful because they simultaneously display information about the shape and dispersion of the structural dimensions across the six clusters. Each box itself comprises the middle 50 percent of observations. The line within the box is the median. The lower end of the box signifies the first quartile, whereas the upper end of the box corresponds to the third quartile. In addition, the lowest and the highest lines outside the box indicate the minimum and maximum values.

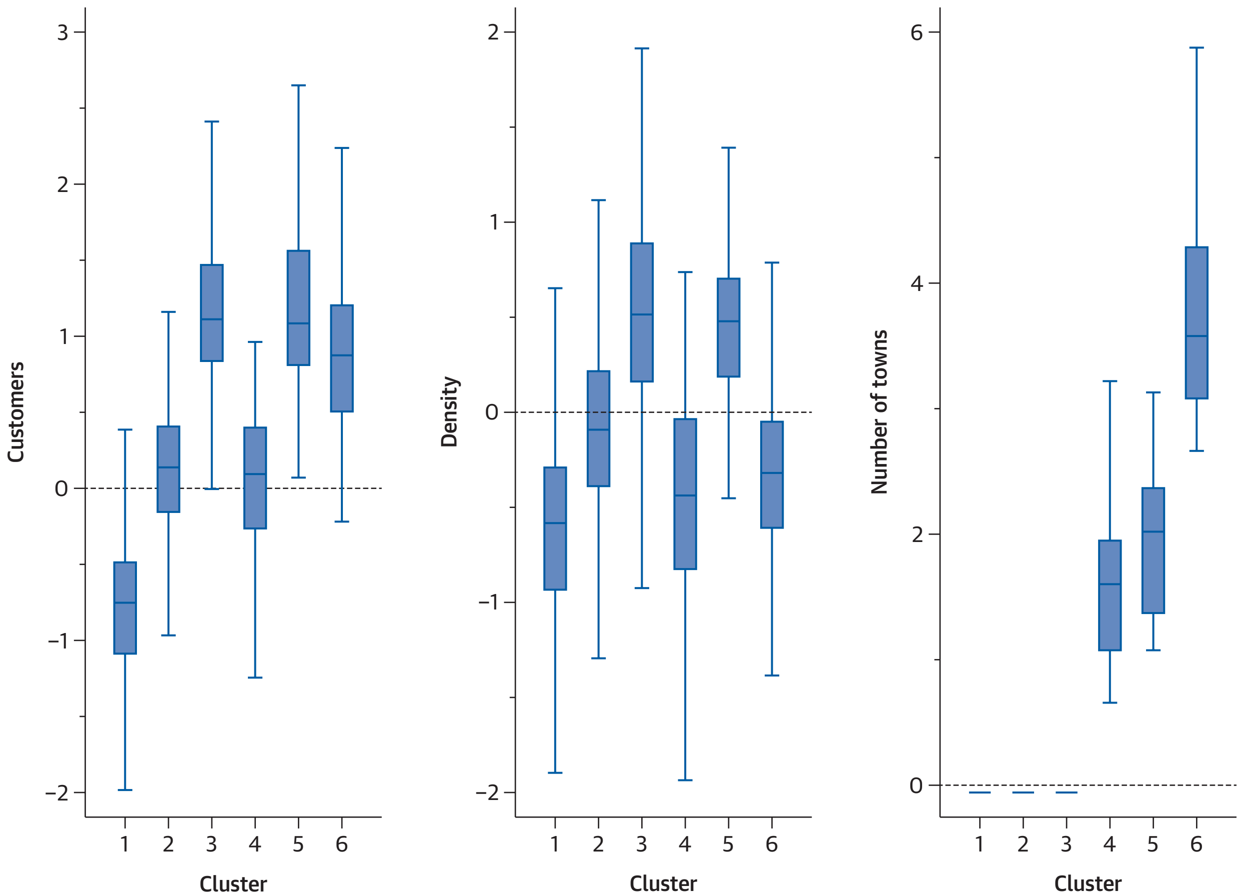
The clustering appears to strongly distinguish between utilities that serve a single town with water and sewerage (clusters 1 to 3), and those that serve several towns (clusters 4 to 6). In the latter group, the clusters are further distinguished into utilities that tend to serve a limited number of towns (clusters 4 and 5) and a group of utilities that serve a large number of towns (cluster 6).

For customers and even more so for density, the distinctions between the clusters are not as clear as for the dimension of towns. As can be seen, the dispersion of customers is nonnegligible for most clusters. For customers, cluster 1 is clearly the cluster with the lowest number, followed by clusters 2 and 4, which serve an intermediate category. The remaining three clusters (cluster 3, 5, and 6) serve a high or medium-high number of customers. In the case of density, the clusters could also be roughly described as exhibiting low, medium, or high densities. Clusters 2 and 5 show high densities; clusters 3 and 4 show low densities. For clusters 1 and 6, the results are less clear with a wide dispersion of densities, possibly indicating an average density.

Although the overlaps seem considerable for some clusters in terms of customers and density, it should be noted, however, that in combination with the number of towns the final cluster grouping is quite distinctive. To illustrate this observation, figures 2.4 to 2.6 show the scatter plots of each dimension with each other, differentiating clusters by color. As shown in the panels, already using two combinations of the structural dimensions clearly separates most clusters.

Finally, the combination of all three dimensions gives a quite clear-cut distinction, which can be qualitatively described as in table 2.3. Cluster 1, for example, is what

FIGURE 2.3. **Box plots for Customers, Density, and Number of Towns, by Cluster**

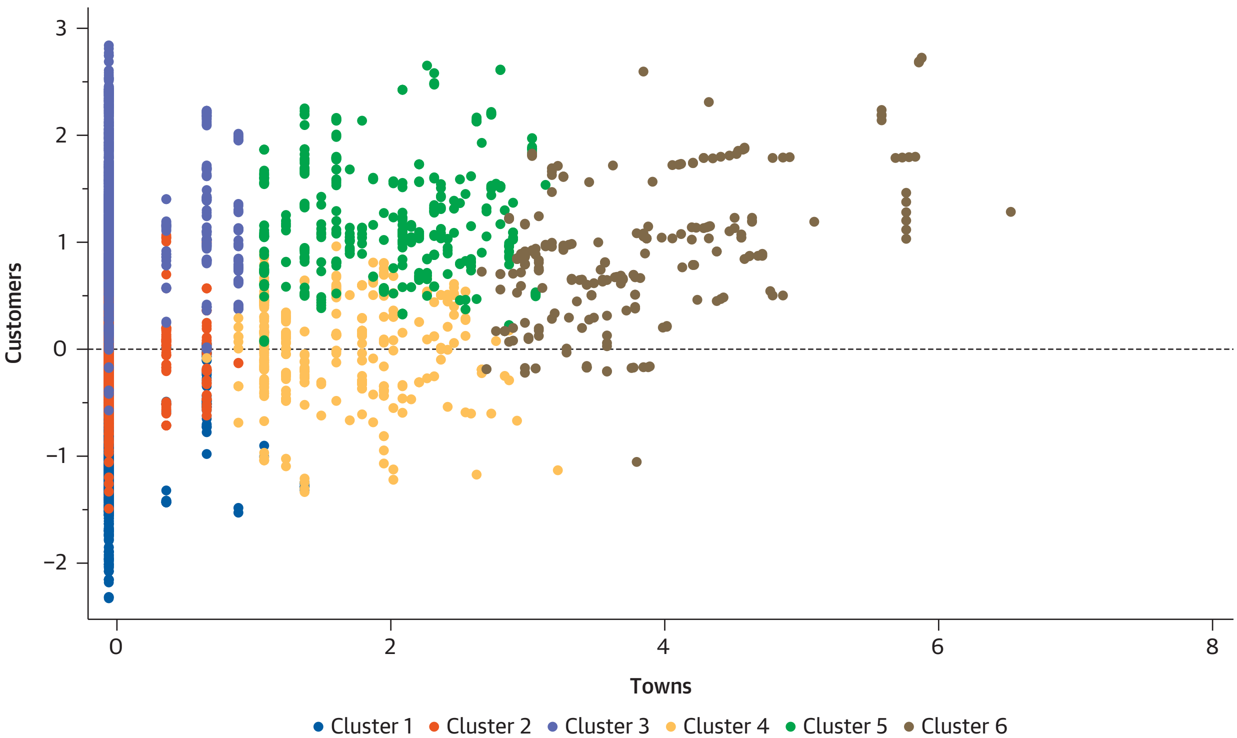


Note: The variables have been transformed by taking the natural log of the original value and then standardizing the variables. Outliers have been omitted from the graph for presentation purposes.

might typically be understood as a "small" utility, in all respects: it serves few customers, with low density and only in a single town. The other two clusters that serve only a single town are cluster 2 and 3, and they distinguish themselves through customers and density. Cluster 2 exhibits both a medium number of customers and medium density. Cluster 3 has a high number of customers and high density. Hence, utilities that serve a single town could broadly be distinguished as small, medium, and large because customers and density seem to be highly correlated inside this subgroup. It should be added that the distinction by number of towns is also critical for the observations of each cluster: clusters with utilities that serve only a single town are much more common than are utilities that serve several towns, by a multiple.

Also, the utilities that serve more than several towns are separated rather distinctly through the clustering—however, the relation between customers and density is much less clear-cut. Cluster 6 represents utilities that serve a large number of towns and typically many customers, albeit at a lower density. Cluster 5 also serves many customers, but with a higher density and fewer towns. Finally, cluster 4 serves an intermediate number of towns, a medium number of customers, and low density. Compared with clusters that serve only a single town (clusters 1 to 3), utilities in the

FIGURE 2.4. **Scatter Plot for Customers and Number of Towns**



Note: The variables have been transformed by taking the natural log of the original value and then standardizing the variables.

clusters that serve more towns tend to exhibit lower densities for similar numbers of customers. This finding suggests that serving more towns will often go hand in hand with reduced supply densities.

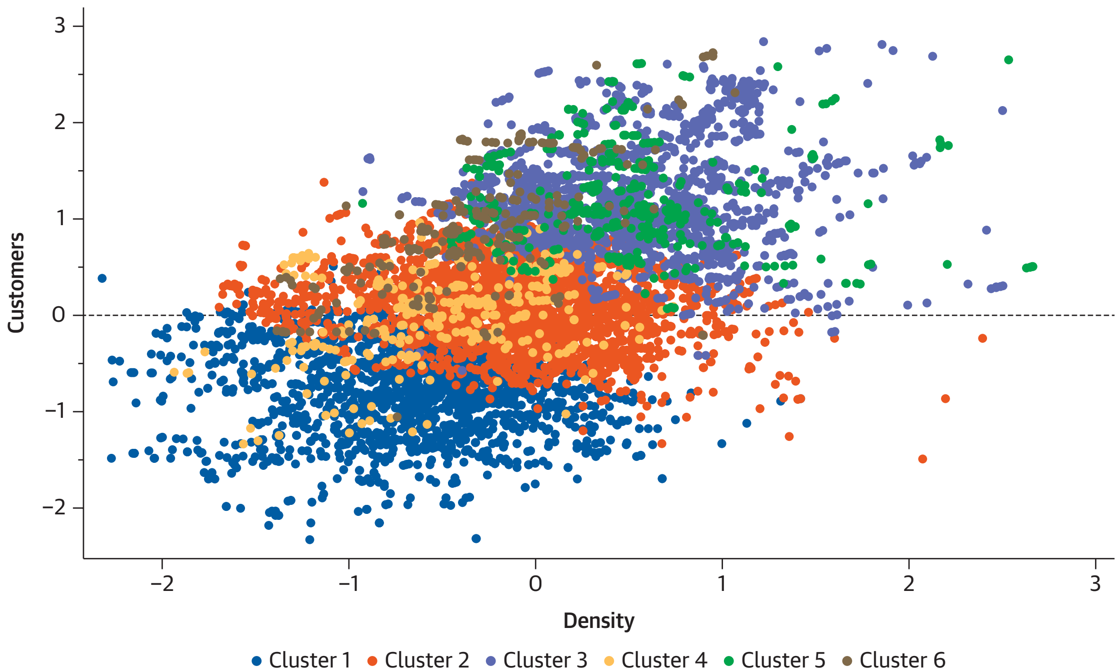
Overall, the results of the clustering suggest that the number of towns, particularly whether a municipality serves a single town or several towns, is a key variable to distinguish utilities. By definition, aggregations will tend to move utilities from clusters 1, 2, and 3 to cluster 4 or 5, or even to cluster 6.

2.3 Relationship to Performance and Input Structure

Even if the clustering is able to identify homogeneous but distinct clusters of utilities, the question arises whether the observed differences in utility structure are also meaningful for input decisions and output/outcomes. Put differently, do utilities of varying structure exhibit systematic differences with respect to (a) performance indicators and (b) their input structures? Although this is no causal analysis, sustained differentials in production decisions (input mix), outcomes, or both could indicate how changing structure—through aggregations— will ultimately affect a utility.

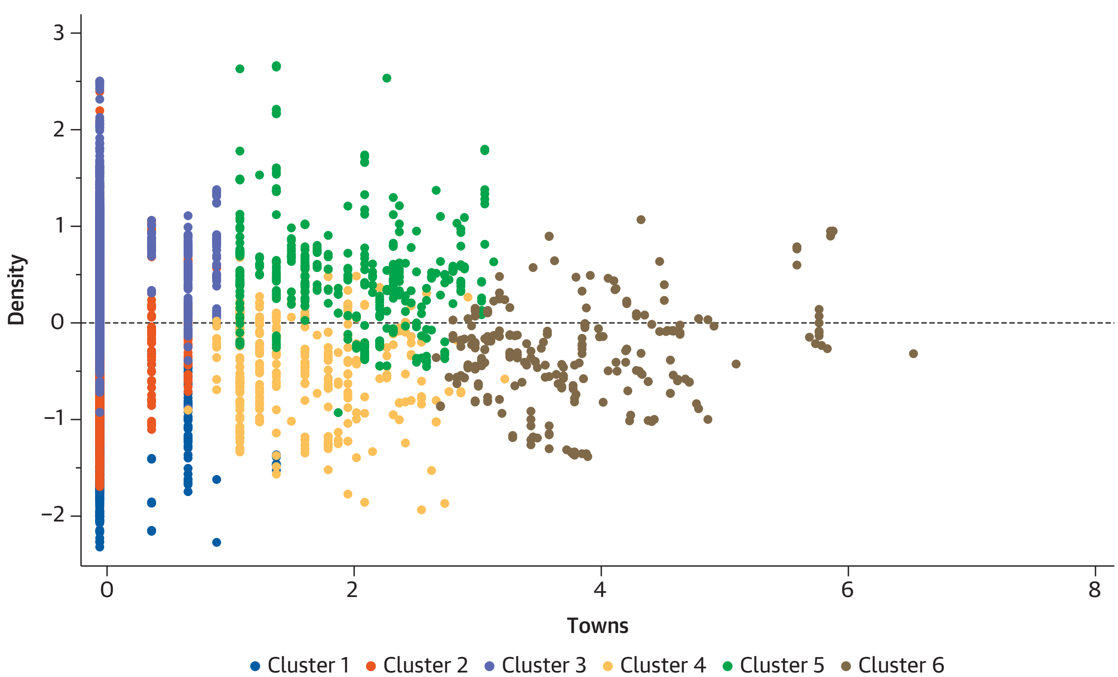
To start with, figure 2.7 exhibits box plots for cost per m^3 and the composite performance indicator *water utility performance indicator* (WUPI).[6] What is striking is the quite large dispersion both for costs and for WUPI. Cost and utility performance are affected by a multitude of other factors apart from structural characteristics. Thus not only is this a correlation exercise and not causal relationship, but also the differences between clusters are not extremely clear-cut in the sense that some clusters always exhibit better performance indicators than other clusters do.

FIGURE 2.5. Scatter Plot for Customers and Density



Note: The variables have been transformed by taking the natural log of the original value and then standardizing it.

FIGURE 2.6. Scatter Plot for Density and Number of Towns



Note: The variables have been transformed by taking the natural log of the original value and then standardizing the variables.

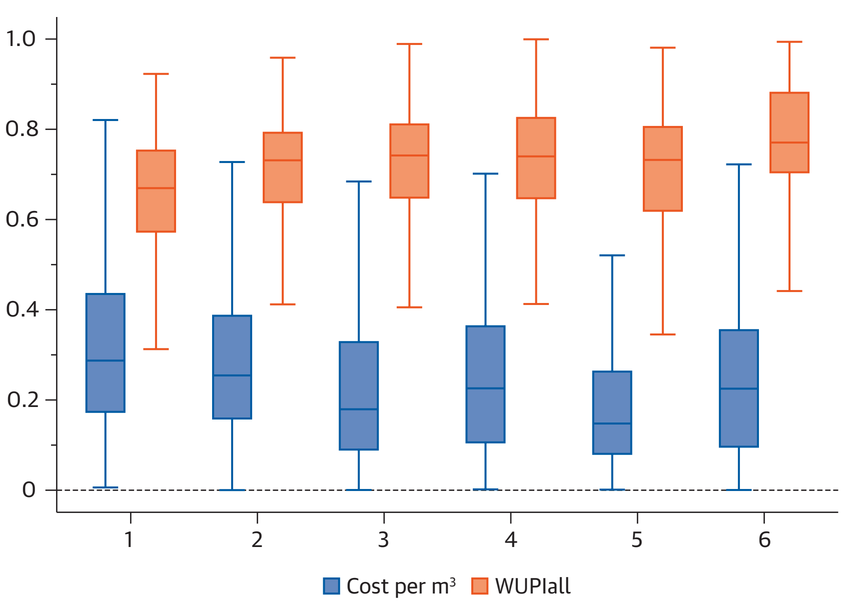
Moreover, there is a continuum of utilities—ranging from well performing to potentially troubled, unsustainable providers—in every cluster. Nevertheless, a number of regularities seem to arise: in the case of the clusters for utilities that serve a single town, supplying more customers with higher density seems to be positively correlated with performance. Both WUPI and cost per m³ improve when a utility moves from cluster 1 to cluster 2 and further to cluster 3. The picture is particularly clear for the "small" utilities in cluster 1, whose WUPI scores (25th, median, and 75th percentile) are below and whose unit costs (25th, median, and 75th percentile) are above any other cluster. The results are more mixed for utilities that serve several towns. For instance, although the cost distribution of cluster 5 tends to be systematically lower than for utilities in cluster 4, performance in terms of WUPI scores is comparable or even slightly lower. Moreover, although the average number of customers in cluster 6 is much larger than in cluster 4, the unit cost distributions look very similar. Conversely, the WUPI score distribution of cluster 6 appears to be the highest of all clusters. Without trying to interpret these correlations too much, one can glean an important insight that more customers and higher density clearly seem to be positively related to performance in utilities that serve a single town. The relationships, however, are more complicated when looking at clusters of utilities that serve several towns.

TABLE 2.3. Clusters According to Structural Dimensions

	Towns	Customers	Density	Observations
Cluster 1	Low	Low		2,148
Cluster 2	Low	Medium		2,995
Cluster 3	Low	High		1,834
Cluster 4	Medium	Medium	Low	380
Cluster 5	Medium	High	High	400
Cluster 6	High	High	Medium	251

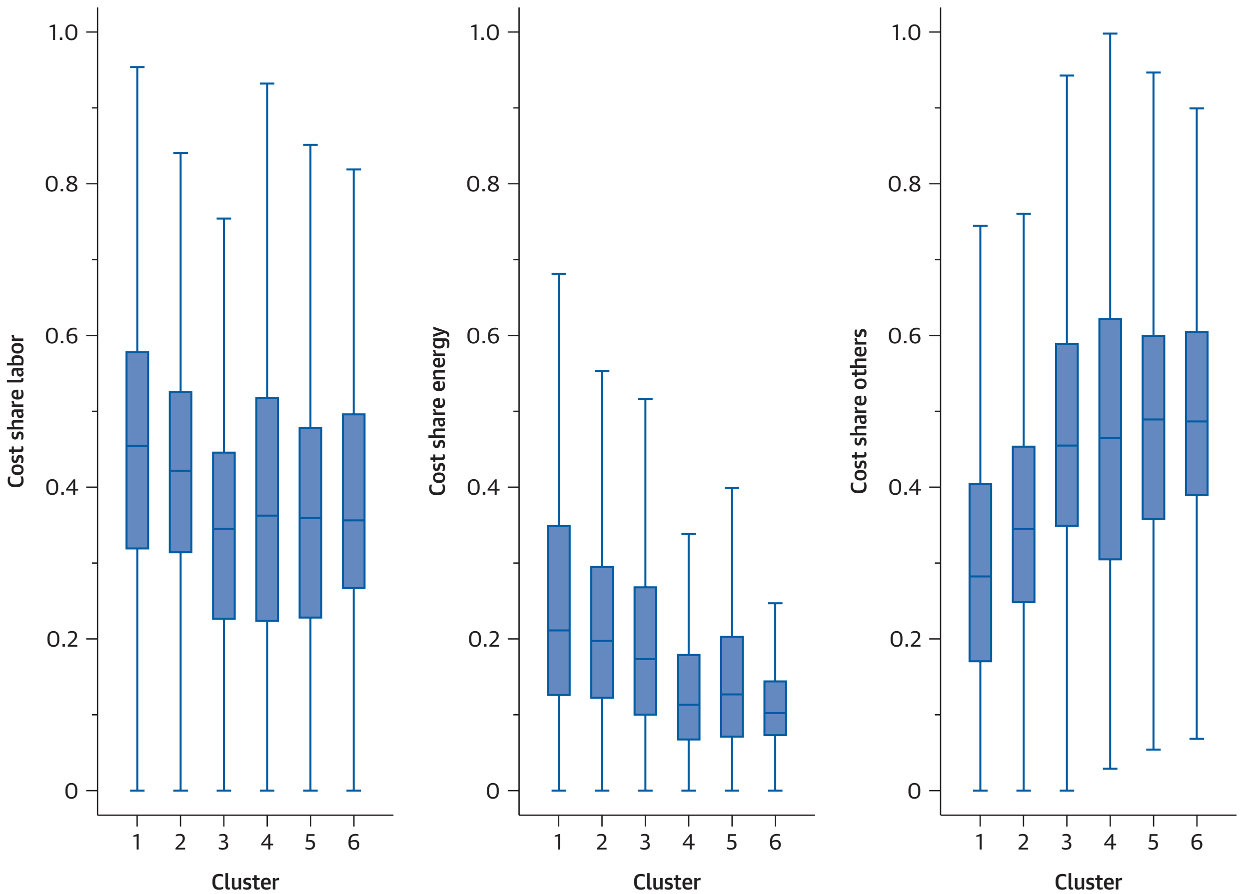
In order to understand the performance differences between clusters of different structures, considering the role that differences in the cost structure may play is important. To help with analyzing this relationship, figure 2.8 displays the cost shares for (a) labor, (b) energy, and (c) other costs for the six clusters. Despite considerable variation in each cluster—for example, each cluster contains utilities with very high but also very low labor cost shares—a few striking patterns emerge. Again, it is helpful to first concentrate on clusters 1 to 3, which serve only a single town. The left panel shows that for utilities that serve a single town, cost shares (median as well as 25th and 75th percentiles) spent on labor decrease from cluster 1

FIGURE 2.7. Box Plots for Unit Costs and WUPIall over Clusters



Note: Unit costs are in converted local currency unit (LCU)-dollars per m³; water utility performance indicator (WUPI) scores range between 0 and 1, with 1 indicating the highest score.

FIGURE 2.8. Box Plots for Cost Shares



to cluster 2 and further to cluster 3. Given the underlying characteristics of these clusters, this observation can be interpreted as a negative correlation between size and density and the share of labor cost. Larger, denser utilities spend a lower proportion on labor. A similar pattern applies for energy costs, which also decrease from cluster 1 to 2 and 3. Because the three cost shares add up to 1, it is little surprising that the converse holds for other costs (such as consulting costs or various procured goods).

The idea that larger utilities spend less on labor and energy not only as a share but also in absolute values is confirmed by figure 2.9, in which as we move from cluster 1 to cluster 2 and 3 we observe falling labor and energy cost per m^3. Zooming in further on labor cost and taking the size of the staff into consideration, figure 2.10 shows that the cost reductions related to labor are due to having fewer workers (right panel). Conversely, as the left panel in figure 2.10 shows, cost per worker as an indicator of wage level seems higher in larger utilities.

Switching to the component for "other cost," as shown in the rightmost panel of figure 2.9, this cost component does not appear to decrease with m^3 as labor and energy costs do. If anything, the clusters with larger and denser utilities exhibit higher cost per m^3. Taken together, the results for utilities serving a single town suggest that increasing customers and density are related to lower labor cost and lower energy cost—both in cost shares and in absolute terms.

FIGURE 2.9. Box Plots for Costs per m³

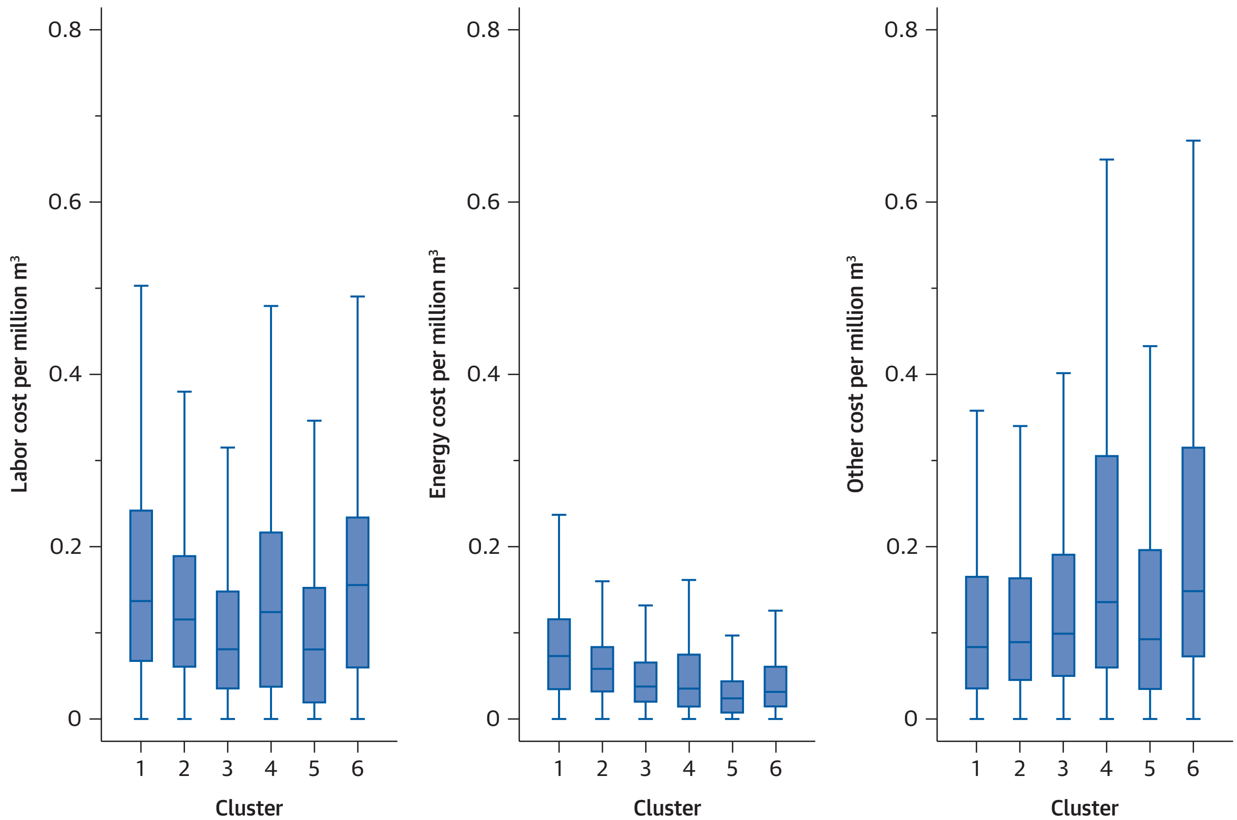
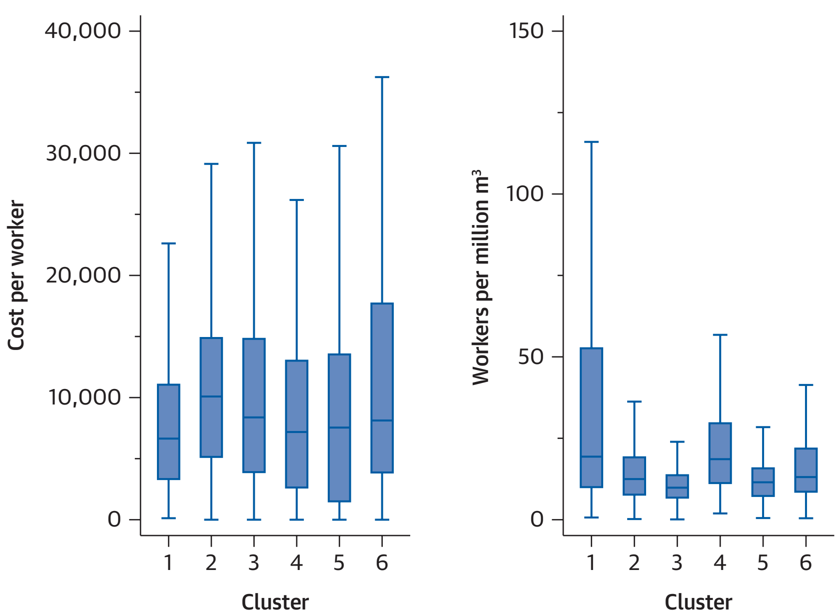


FIGURE 2.10. Box Plots for Labor Cost Components



For other cost components, rather the opposite relation seems to apply. Economies of scale, if any, therefore seem to originate from labor and energy, whereas other costs might even increase with customers and density. Switching to utilities that serve several towns, the results on the input structure are generally less clear. Despite differences in customers, density, and the number of towns, clusters 4 to 6 are not clearly distinguishable in terms of cost shares. For energy cost, all utilities that serve several towns exhibit comparatively low energy cost shares.

The difference is considerable compared with clusters of utilities that serve only a single town. Conversely, all three clusters have relatively high shares of "other" costs (right panel in figure 2.8) and also in absolute terms, clusters 4 to 6 exhibit higher "other" costs per unit than clusters 1 to 3 (right panel in figure 2.9). Because these other costs often account for more than 50 percent of total cost, the question arises whether those costs represent higher transaction costs in the case of utilities that serve several municipalities. However, without more detailed knowledge about the cost types in this residual category, it is difficult to speculate about the source of these cost differences.

Also for labor cost, the picture is very mixed. Both for labor share and the absolute labor cost per unit, the dispersion in clusters 4 to 6 is very large—suggesting a large amount of heterogeneity beyond utility structure. The median cost per worker is relatively low in these clusters, and staffing per m^3 tends to be comparable to that of utilities that serve a single town. A tentative appraisal is that utilities that serve several towns, therefore, seem to have larger staffs albeit at lower wages.

Notes

1. In this study, we follow World Bank (2005) and define aggregations as a situation in which previously separate utilities are integrated into a single utility. This definition is general enough to comprise both purely managerial aggregations and cases of asset bundling or even physical connection of networks and infrastructure
2. IBNET is a data repository initiated and maintained by the World Bank with the objective of improving the service delivery of water supply and sewerage utilities through the provision of international comparative benchmark performance information. For more information on IBNET, see the appendix and Van den Berg and Danilenko (2011).
3. Because the dispersion of the indicators is often considerable, due to few very large values, the indicators are in natural logs and are standardized.
4. Although it would also be possible to use the principal components directly, the raw indicators sort sufficiently clear into the components—indicating that they measure different structural aspects. Moreover, the interpretation of the indicators is much more straightforward than of the principal components.
5. See appendix A for a description of the methodology for the clustering procedures.
6. WUPIall indicates the aggregate/composite indicator from subcomponents WUPIcoverage, WUPIquality, and WUPImgmt, which are used later on. More information about the index and its construction is given in World Bank/IAWD (2015).

References

Van den Berg, Caroline, and Alexander Danilenko. 2011. *The IBNET Water Supply and Sanitation Performance Blue Book: The International Benchmarking Network of Water and Sanitation Utilities Databook*. Washington, DC: World Bank.

World Bank. 2005. *Models of Aggregation for Water and Sanitation Provision*. Water Supply and Sanitation Working Notes, note 1. Washington DC: World Bank. http://documents.worldbank.org/curated/en/2005/ 01/5731013/models-aggregation-water-sanitation-provision.

World Bank/IAWD (International Association of Water Supply Companies in the Danube River Catchment Area). 2015. Water and Wastewater Services in the Danube Region: A State of the Sector. Regional Report. Washington, DC: International Bank for Reconstruction and Development /The World Bank. http://sos.danubis.org/files/File/SoS_Report.pdf.

Chapter 3
The Empirics of Aggregation

Although the relationship between utility structure and performance is interesting and helpful to understanding how aggregations could possibly affect a utility, looking at the relationship alone ignores that aggregations are an inherently dynamic process. As a result, here the focus is to describe the process of aggregation in terms of utility structure. The starting point is to look at the structure of aggregating utilities before the aggregation and compare them with the utilities not aggregating. The second subsection deals with the question of how aggregations, as observed in IBNET, change the structure of a utility. Thus, the movement between utility types and clusters is analyzed.

3.1 Are Aggregating Utilities Different?

The question of whether aggregating utilities are different from the "average" utility is important for any evaluation measuring the success of aggregations. Although a pure before-and-after comparison is interesting, a real test of whether a reform was successful is by comparing the change in performance with that of other similar utilities. Hence the question of choosing an appropriate counterfactual is a key step in any evaluation process. To understand if more involved statistical tools are necessary to choose meaningful comparison utilities, table 3.1 shows the structural indicators for aggregating and not aggregating utilities at various percentiles and the mean.

The units are identified as aggregating utilities when the number of towns served increases over time. These utilities could also be interpreted as the acquiring firms. As table 3.1 illustrates, even before aggregating, the utilities that would later take over more towns were different from utilities that do not aggregate in all three measured structural dimensions. Looking at the average, the number of customers served by aggregating utilities is roughly two times larger, their density is more than 10 percent higher, and the number of towns they serve even before the aggregation is already higher in most cases. This comparison suggests that the aggregating utilities that are observed in IBNET (those that increase the number of towns served) are larger, denser, and serve more towns from the start.

This finding has two main implications. First, the causal analysis in the following section will have to incorporate the fact that aggregating utilities are considerably different from nonaggregating utilities in terms of structural features. Choosing appropriately similar comparison utilities will be important. Second, the following results have to be interpreted carefully in the sense that they do

TABLE 3.1. Comparison of Aggregating and Not Aggregating Utilities in IBNET Sample

	Indicator	p10	p25	p50	mean	p75	p90
Aggregating utilities	Customers	34.1	64.7	167.7	593.5	349.0	1,021.0
	Density	144.5	226.5	299.0	380.1	478.0	628.2
	Towns	2.0	2.0	4.5	17.2	10.0	24.0
Not aggregating utilities	Customers	8.0	18.9	57.0	281.4	198.4	542.5
	Density	130.6	180.9	254.5	306.9	372.0	553.0
	Towns	2.0	2.0	2.0	3.5	2.0	2.0

Note: Customers are in 1,000s; p = percentile.

not measure how aggregations affected the average utility. Rather, the results measure the effect of aggregations on utilities that were already larger in many dimensions before the aggregation. Although it is quite common in aggregation reforms to have large utilities take over many small providers, we can only speculate about the performance effects for very small utilities.

3.2 How Aggregations Change Utility Structure

Aggregations involve the expectation that utilities will grow in size. By definition, the acquiring utility will grow in the number of served towns. By how many towns depends on the aggregation design, as do the changes in customers and density. The clustering results suggested that utilities with more towns tend to have more customers, but those conditions are often coupled with a lower density. Using the data in IBNET it is possible to go a step further and look directly at how aggregations change all three structural dimensions.

The way that aggregations affect utilities in the sample of IBNET data is displayed in figures 3.1 and 3.2. The scatterplots show nonaggregating utilities once (averaged over all observation periods) and aggregating utilities twice, once before and once after the aggregation (again averaged over the preaggregation and postaggregation periods). In both figures, the arrows indicate how the aggregation changed the structural characteristics. In figure 3.1, the arrows represent the change in customers and the number of towns served. In figure 3.2, the arrows represent the change in density and the number of towns served. As before, the variables were first logged and then standardized. What the graphs have in common is the fact that the change in number of towns is larger than the change in customers or density. In the case of customers, the arrows appear almost horizontal, in most instances showing little increase in customers.[1]

Even more striking is the decrease in density through aggregations, shown in figure 3.2. Except in a small minority of cases, aggregations seem to lower density, sometimes very strongly. A likely explanation for this finding is that many of the aggregations involved a large number of small utilities, hence decreasing density.

FIGURE 3.1. Aggregations and Change in Number of Customers

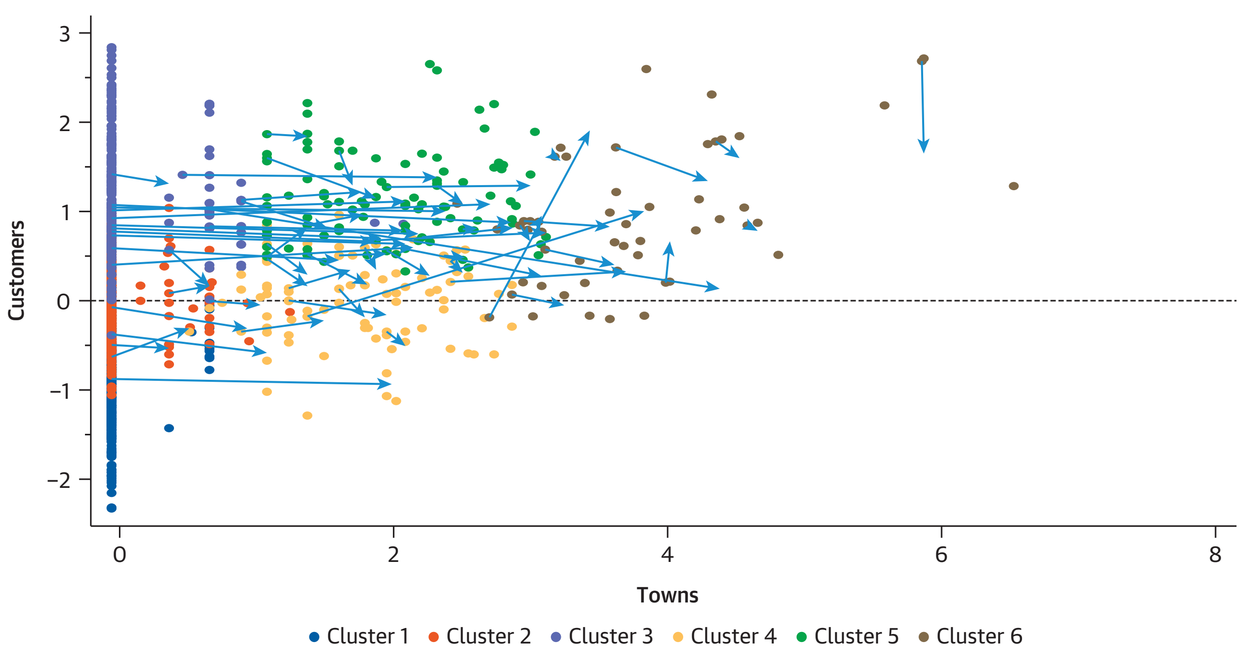
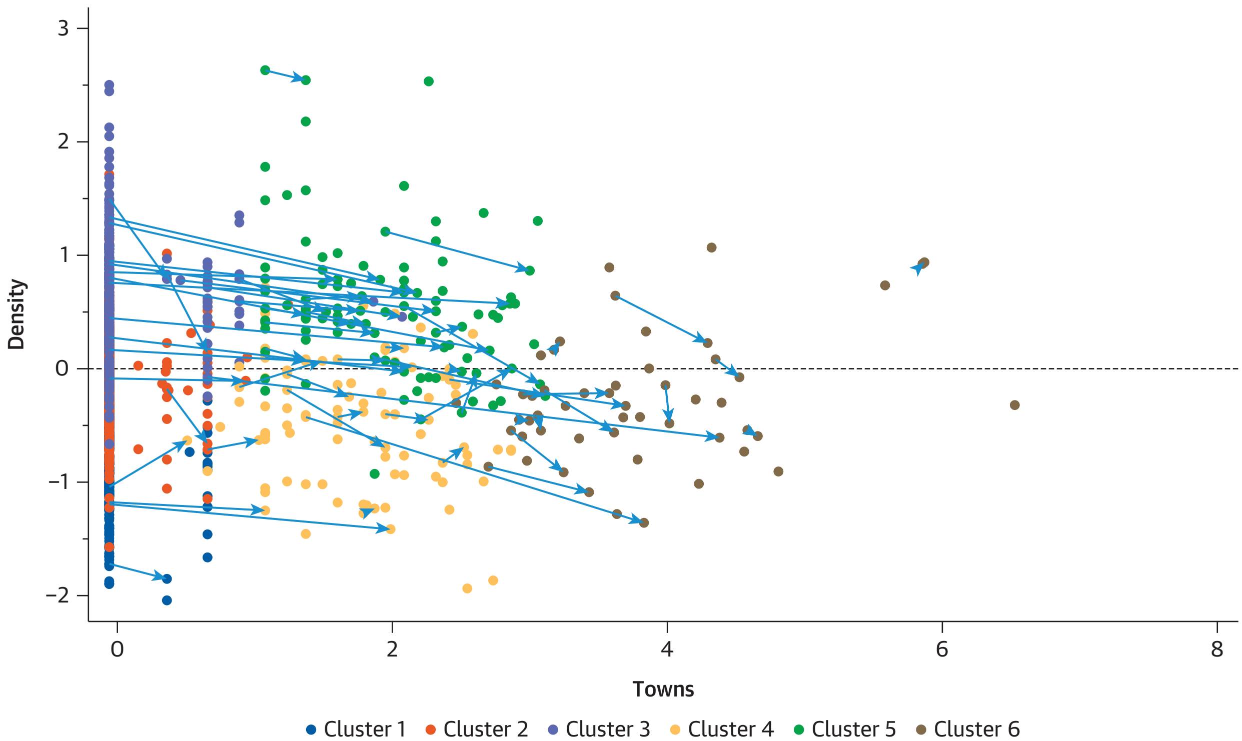


FIGURE 3.2. Aggregations and Change in Density



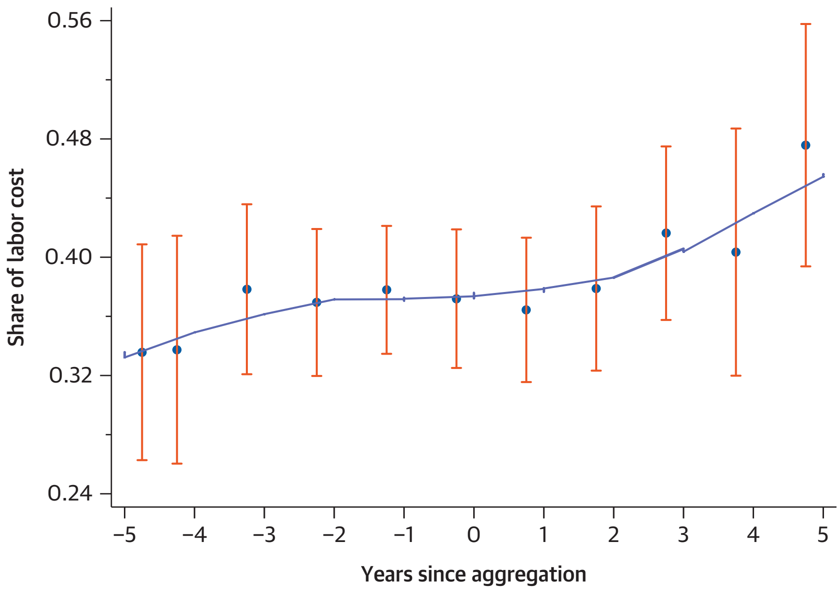
Also, the movement between clusters is telling in this respect. Because the movement is mostly right and down, the utilities are moving from clusters of higher density to clusters of lower density (such as from cluster 3 to cluster 4).

The picture that develops of small gain in the number of customers and loss of density might be an explanation why not all consolidations decrease cost through economies of scale. Importantly, in most empirical studies, economies of scale are defined as a proportional increase in outputs—that is, in customers, volume, and the number of towns (see, for example, Garcia and Thomas 2001). The assumption of proportionality seems clearly violated in the sample of IBNET utilities. In this case, it seems that the design of the aggregations might have been unfavorable to achieving cost savings from the start.

Another issue that calls into question the potential cost savings in the observed aggregations is the evolution of the cost structure. The clustering suggested that cost savings due to having more customers and higher density are often related, with lower labor shares, especially a reduced workforce. The data do not suggest any reduction in the labor share for aggregating utilities. Figure 3.3 shows the evolution of the labor share from five years before the aggregation to five years after the aggregation in a local linear smooth plot: the orange bars mark the upper and lower 95 percent confidence intervals; the blue dots show the mean per year, which are smoothed in the purple line by a local linear smoother (lowess). The graphs show that on average, labor cost shares do not seem to decrease—rather the contrary. This finding is also consistent with some of the case studies in Michaud and others (2017), in which an upward wage harmonization occurred or the aggregated utility was forced to take over the staff of the previous utilities.

Labor costs appear to play a key role in this setting, not only because they are frequently the largest single

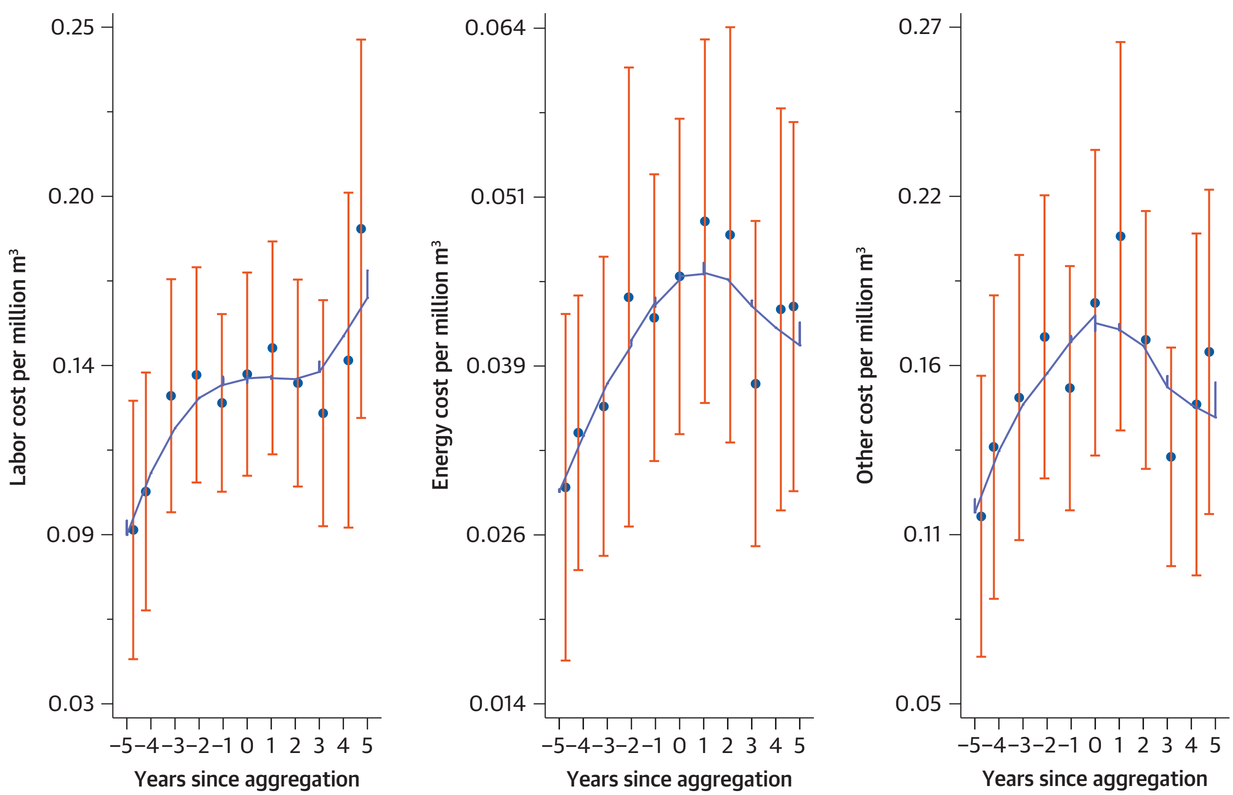
FIGURE 3.3. Labor Share before and after Aggregations



cost component, but also because they are the only cost component that appears to exhibit downward rigidity. Similar to the macroeconomic phenomenon that wages rarely decrease in nominal terms, utility labor costs do not seem to decrease even after aggregations (see figure 3.4).

Although all cost components increase before aggregations—caused by inflation and possibly also some short-run transaction cost of the aggregation reform—energy and other costs come to a halt and even decrease after the

FIGURE 3.4. Cost Components before and after Aggregations



aggregations while labor costs continue to increase. Although this is no causal analysis, it alludes to the critical role of labor costs to achieving cost savings through an aggregation reform. A more coherent analytical framework to analyze the consequences of aggregations is considered in the next section.

Note

1. Some utilities even observe decreases, which are likely the result of depopulation. Similarly, a significant number of utilities reduce produced water over time, a change that manifests as a secular trend in the data over time.

References

Garcia, Serge, and Alban Thomas. 2001. "The Structure of Municipal Water Supply Costs: Application to a Panel of French Local Communities." *Journal of Productivity Analysis* 16 (1): 5-29.

Michaud, David, Maria Salvetti, Michael Klien, Berenice Flores, Gustavo Ferro, and Stjepan Gabric. 2017. *Joining Forces for Better Services? When, Why, and How Water and Sanitation Utilities Can Benefit from Working Together.* Washington, DC: World Bank.

Chapter 4
The Performance Consequences of Aggregation

4.1 Empirical Strategy

Although IBNET covers several thousand utilities all over the globe, the number of aggregations in the database is substantially lower: after cleaning the data and restricting the analysis to utilities suitable for an evaluation, 79 aggregation cases remained. Most of those cases occurred in Europe or Central Asia (table 4.1). Although IBNET is not representative in terms of country coverage, the database suggests that most of the aggregation reforms occurred in Central and Eastern Europe and, to some extent, in South America and Central Asia. Virtually all aggregations are in the time period 2000 to 2010, with a scant few before and after these dates. On a country level, the following countries exhibit most aggregation cases: Romania (15); Poland (12); Kazakhstan (7); Hungary (6); Serbia (5); the Czech Republic (4); and the former Yugoslav Republic of Macedonia (4). It should be noted, however, that although the bulk of cases are located in this region, overall 25 countries exhibit cases of aggregations that feed into the analysis.

Related to the country distribution, most aggregations occur in upper-middle-income and high-income countries (table 4.2). Some cases are located in lower-middle-income countries, but none are from low-income countries.

The quantitative analysis is limited to the data and information available in IBNET and therefore focuses on the outcome of aggregation processes in terms of economic efficiency and performance improvements. Because of data limitations, the impact on externalities such as equity or environmental factors is therefore excluded. Likewise, the dataset does not allow an in-depth investigation of the influence of utility governance or aggregation process design on overall outcomes. Those issues are investigated in greater detail in qualitative case studies in Michaud and others (2017). In this statistical analysis, the focus is on a general appraisal of whether aggregations generated the expected cost savings or performance improvements. Regarding utility performance, this report uses a set of quantitative indicators to capture the various purposes of aggregations. Most important, these indicators are coverage, quality of service, and management efficiency. In addition, these subindicators are also used as an aggregate in the form of a composite performance indicator (WUPI). It should again be noted, however, that although these indicators capture some important aggregation purposes, goals that go beyond these dimensions are outside the analysis of this report.

The previous sections relied heavily on cross-sectional comparisons of utility structure and its connection to performance. Do systems with many customers, for example, exhibit lower unit cost than systems with few customers? When considering an aggregation reform, the relevant policy question, however, is whether the

TABLE 4.1. Distribution of Aggregations by Region

Region	Number of cases	Number of countries
East Asia and Pacific	3	3
Europe and Central Asia	69	17
Latin America and Caribbean	5	3
Middle East and North Africa	1	1
Sub-Saharan Africa	1	1

TABLE 4.2. Distribution of Aggregations by Income Level of Countries

Region	Number of cases
High	income 27
Upper	middle income 43
Lower	middle income 9

Note: Income levels based on World Bank's *World Development Indicators* definitions.

utility did improve compared with a situation in which a utility did not aggregate. Particularly, the previous section has shown that aggregations often add few customers and tend to decrease density—hence a comparison of utilities with many and few customers could be very misleading.

In this section, utility performance is monitored before and after consolidations for aggregating utilities and is compared with nonaggregating utilities. To this end, we run regressions including utility-fixed effects to compare the performance change of consolidating firms with that of nonconsolidating firms.[1] As the previous sections have shown, aggregating utilities are different from the average utility in IBNET, suggesting that the choice of the control group—that is, the group of utilities without aggregation that is used as a comparison—might be important for the obtained results. With the overall goal of a counterfactual scenario—of what the average cost of a utility would be in the absence of a consolidation—not all utilities are suitable for comparison.

For this reason, different matching techniques are used to select suitable comparison utilities. In each case, a large set of pretreatment characteristics to estimate the probability that a utility experiences a consolidation (see Rosenbaum and Rubin 1985) is used to identify the final sample. Depending on the matching algorithm, one or several utilities with similar treatment probability are then chosen as the control group. While the combined analysis of water and wastewater is continued (volume is the sum of water produced and wastewater collected), for the choice of comparison units the separate indicators are used.[2] Hence the variables $x_{k,it}$ to estimate the probability of an aggregation include important utility characteristics such as the population in the service area and the number of towns served, separate for water and wastewater. In addition, the pretreatment performance of a utility in terms of managerial and operating efficiency (WUPI) is also added. Finally, country as well as year dummies enter the specification to capture heterogeneity across countries and time. The former is particularly relevant because some countries do not experience any aggregation while others exhibit a considerable number.

Apart from the statistical need to balance utility characteristics between treatment and control groups, this approach also ensures that the consolidation effects are evaluated in comparison with utilities of similar initial size, and that utilities do exhibit a similar share of water and wastewater services. Because the existing empirical literature has stressed decreasing economies of scale and even diseconomies of scale, matching utilities according to their production structure in size and scope seems imperative. The production characteristics were first added linearly, before adding squared terms where necessary to achieve balancing.

Because the choice of the matching algorithm is somewhat arbitrary, we use three different matching approaches and also the full sample of utilities, which result in using four different control groups. We use (a) nearest-neighbor propensity score matching, (b) four-nearest neighbor propensity score matching, (c) radius matching, and (d) all utilities in the sample. The different algorithms (a) to (c) represent difference choices in the trade-off between bias and variance (see Caliendo and Kopeinig 2008). All three algorithms are limited to the utilities on common support. The full sample, (d), is displayed for comparison reasons but should be interpreted with care because the compared utilities differ substantially.

These different subsamples of comparable treatment and control utilities are then used in the generalized difference-in-differences specification:[3]

$$Perf_{it} = \beta_0 + \beta_1 {*}\, aggregation_{it} + \gamma_i + \eta_t + u_{it} \qquad (1)$$

where $Perf_{it}$ refers to a performance indicator for utility i in year t. In addition to variable cost per m^3 (in natural logs of dollar-converted local currency), the composite performance indicator WUPI as well as its subcomponents are used. Regarding the subcomponents, *WUPIcoverage*, *WUPIquality,* and *WUPImgmt* are distinguished:[4]

- *WUPIcoverage* is basically an indicator for the share of population connected to water and wastewater services and the extent of wastewater treatment. Higher values indicate a higher share of population connected and a higher extent of wastewater treatment.
- *WUPIquality* represents the performance of a utility with respect to the number of hours of service as well as the frequency of sewerage blockages. Higher values indicate more hours of service and fewer blockages.
- *WUPImgmt* is less an indicator on the customer side than it is related to the managerial efficiency. It is based on a number of subindicators such as the extent of staffing, cost recovery, the share of metered connections, revenue collection, and nonrevenue water. Higher values indicate higher cost recovery and recovery collection, more metered connections, lower staffing, and lower nonrevenue water.

As does the aggregate indicator WUPIall, the subindicators range from 0 to 100, with higher values indicating better performance. Looking at various performance indicators is necessary because aggregations can follow various purposes, and achieving scale economies may not be a goal at all. The regressions include utility and time-fixed effects, thus the effect of $aggregation_{it}$ is identified by comparing unit costs over time and between treated and control utilities.

It should be noted that the use of variable cost gives the estimates a short- term interpretation. Capital-stock in terms of the network infrastructure is certainly fixed, a modification infeasible or prohibitively costly. (See Garcia and Thomas 2001.) Water pipes typically last a long time—up to 50 years, depending on the situation and the chosen material. Such durability would indicate that the system configuration is fixed for a very long time horizon. Although a comprehensive analysis of short-run and long-run costs would still be desirable, that is not feasible with the data at hand.

Given the discussions in the previous sections, the effect of aggregations might depend both on the initial structure of the utility and on how the aggregation changes a utility's structure. To allow for the latter possibility that the effect of the aggregations is not independent of the size of the change, the model in equation (1) is rerun with the indicator variable $aggregation_{it}$ (a) replaced by dummy variables distinguishing small aggregations (*aggregation_size* < 20 percent increase in the number of towns), medium aggregations (*aggregation_size*, between 20 percent and 100 percent change in the number of towns), and large aggregations (*aggregation_size* more than 100 percent change in the number of towns).

$$Perf_{it} = \beta_0 + \sum_{k=1}^{4} \beta_k * k.aggregation\_size_{it} + \gamma_i + \eta_t + u_{it} \qquad (2)$$

Similar specifications are run for small, medium, and large changes in density and volume.[5] Moreover, to make the aggregation effect conditional on the initial structure of the utility, the simple treatment dummy is replaced by adding dummy variables that distinguish utilities with few towns (*initial_level*, 2 towns), utilities with an intermediate number of towns (*initial_level*, between 4 and 14 towns), and utilities with many towns (*initial_level*, more than 14 towns).

$$Perf_{it} = \beta_0 + \sum_{k=1}^{4} \beta_k * k.initial\_level_{it} + \gamma_i + \eta_t + u_{it} \qquad (3)$$

Again, the same estimations are repeated with dummies indicating utilities of small, medium, and large density and volume. In all specifications, we cluster standard errors at the utility level and robustify for heteroscedasticity.

4.2 Matching Results

Before we present the regression results for the effect of aggregations on various performance indicators, this section addresses the results from the matching algorithms that are used to identify useful control utilities. The probit regression to obtain the propensity score is exhibited in table 4.3.

It should be noted that for aggregating utilities, the period t-1 with t indicating the aggregation year is used in the regression. The pseudo-R-squared of the regression is 0.44, indicating that the chosen variables can help determine the probability that a utility consolidates. Apart from the country and year fixed effects, the indicators for population in the service area water and wastewater utilities seems to enter the regression significantly. Because the goal of the matching is not exactly to explain the determinants of aggregation but rather to evaluate utilities similarity using the estimated propensity score, the success of the matching is judged by comparing the treatment and control groups. Moreover, explaining the determinants of aggregation this way would be very difficult given the high collinearity in many included regressors.

A more substantive measure in this respect is to evaluate if the matching procedures decreased the observed differences between treatment and control group. This is displayed in table 4.4. The first column of the table shows the initial bias between treated and the full control sample. The measure standardized bias is calculated as the difference in means between the two groups, divided by the standard deviation of the variable in the treated group: $(X_{\text{treated}} - X_{\text{control}})/s_{\text{treated}}$. As can be seen from the first column in the table, these differences are large for a number of variables in the initial sample.[6] The treatment group is systematically different from the nontreated group. Columns 2 to 4 show the remaining bias after the matching procedures. As a rule of thumb, the absolute values of the remaining bias should be statistically insignificant and below 25 (see Rubin 2001). Except in the case of radius matching, in which the number of towns is slightly unbalanced, the applied matching techniques seem capable of choosing appropriate control units. None of the remaining biases are statistically significant at a 10 percent level.

TABLE 4.3. Propensity Score Estimation

	(1) aggregation
WUPIall	0.103 (0.0759)
popsa_w	0.0000123*** (0.00000324)
popsa_ww	−0.0000115*** (0.00000300)
vol_w	−3.45e-09 (9.18e-09)
vol_ww	−3.18e-11 (4.48e-10)
cus_w	0.000000995 (0.00000183)
cus_ww	−0.00000208** (0.00000105)
towns_w	0.0108 (0.00731)
towns_ww	0.0184 (0.0165)
dens_w	0.0000309 (0.000257)
dens_ww	−0.000432 (0.000317)
WUPIall^2	−0.000712 (0.000542)
popsa_w^2	6.15e-14 (3.98e-13)
vol_w^2	−2.18e-18 (1.57e-17)
_cons	−5.407* (2.825)
N	3897

Note: WUPIall is a performance indicator. popsa_w = population of service area for water; popsa_ww = population of service area for wastewater; vol_ww = volume of wastewater collected; cus_w = customers connected to water supply; cus_ww = customers connected to wastewater services; towns_w = number of towns served with water; towns_ww = number of towns served with wastewater; dens_w = density of water system; dens_ww = density of wastewater system; ^2 = squared variable; _cons is the constant.+
Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE 4.4. Bias before and after Matching

	Initial	NN-PSM	4NN-PSM	Radius
WUPIall	36.383186	15.185448	3.892014	7.6758876
popsa_w	21.7962	0.22726597	1.6039469	0.48275611
popsa_ww	19.818401	0.35545123	1.6005663	0.25640118
vol_w	19.591293	2.8924422	1.8626966	0.77712399
vol_ww	11.313143	1.9680327	4.1508121	1.5142661
cus_w	22.180407	1.0372882	1.2646612	0.73440069
cus_ww	21.990953	0.6844226	0.18139637	0.27736446
towns_w	67.926651	1.222091	1.9422516	35.645447
towns_ww	52.364017	4.9559183	8.1277056	35.711647
dens_w	14.555485	13.110049	8.5666676	6.4670525
dens_ww	14.837564	6.7616673	1.1273751	3.0515025
WUPIall^2	38.948284	14.194642	3.6725125	8.1848831
popsa_w^2	17.278624	0.65376735	0.55176806	0.75902593
vol_w^2	17.133339	2.1188266	0.53734493	0.40599945

Note: WUPIall is a performance indicator. popsa_w = population of service area for water; popsa_ww = population of service area for wastewater; vol_ww = volume of wastewater collected; cus_w = customers connected to water supply; cus_ww = customers connected to wastewater services; towns_w = number of towns served with water; towns_ww = number of towns served with wastewater; dens_w = density of water system; dens_ww = density of wastewater system; ^2 = squared variable; NN-PSM = nearest neighbor propensity score matching; 4NN-PSM = four-nearest neighbors propensity score matching.

Table 4.4 suggests that at least on observables, treated and nontreated utilities do not differ systematically after the matching approaches.

4.3 Difference-in-Differences Results

The results from estimating the model in equation (1) for the different performance indicators are shown in table 4.5. Except in the estimations in which the full sample is used (column 4), aggregations do not seem to matter for firm performance, for unit cost, nor for WUPI and its subcomponents. In other words, when similar utilities are used as comparison, no evidence suggests that aggregation affects a utility's performance, positively or negatively.

Apart from the possibility that aggregations have a very limited impact on performance, the previous discussions have suggested a heterogeneous effect, depending on the magnitude of the reform as well as on the initial structure of a utility. For the magnitude of the reform—which is measured as change relative to the initial value—there is little evidence that aggregation matters for the analyzed performance indicators. The only case with a statistically significant effect is the effect of aggregations that add only a small number of towns (less than 10 percent). Only for this type of aggregations is there a clearly negative effect on unit cost.[7]

The second conditionality—that the impact depends on the initial configuration—is at least partially supported by the data. From a unit-cost perspective, utilities that are initially large in the number of towns and in volume appear to be able to profit from economies of scale (see the upper panels in tables 4.6 and 4.7). Conversely, utilities with initially low densities and volumes seem to experience improving WUPIall scores after aggregation (see table 4.8). Negative effects in terms of WUPIcoverage appear for utilities of

TABLE 4.5. Difference-in-Differences

	(1)	(2)	(3)	(4)
	AVC2	AVC2	AVC2	AVC2
1.after	−0.00666	−0.0103	−0.0153	−0.0512**
	(0.0221)	(0.0217)	(0.0220)	(0.0202)
N	865	1,159	5,721	7,621
	(1)	(2)	(3)	(4)
	WUPIall	WUPIall	WUPIall	WUPIall
1.after	−0.0506	0.248	0.280	0.426
	(0.975)	(0.850)	(0.813)	(0.903)
N	936	1,244	5,487	7,014
	(1)	(2)	(3)	(4)
	WUPIcoverage	WUPIcoverage	WUPIcoverage	WUPIcoverage
1.after	−1.159	−1.473	−1.245	−3.109*
	(1.988)	(1.864)	(1.877)	(1.816)
N	936	1,244	5,487	7,014
	(1)	(2)	(3)	(4)
	WUPIquality	WUPIquality	WUPIquality	WUPIquality
1.after	−0.795	0.556	0.245	0.689
	(1.519)	(1.154)	(1.233)	(1.490)
N	915	1,223	2,718	4,209
	(1)	(2)	(3)	(4)
	WUPImgmt	WUPImgmt	WUPImgmt	WUPImgmt
1.after	0.605	1.089	1.031	2.315**
	(1.081)	(1.016)	(0.961)	(1.066)
N	936	1,244	5,487	7,014

Note: Standard errors are in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE 4.6. Difference-in-Differences: Conditional on Initial Number of Systems

	(1)	(2)	(3)	(4)
	AVC2	AVC2	AVC2	AVC2
1.treatedsize	0.00794	0.00287	−0.00344	−0.0569*
	(0.0337)	(0.0340)	(0.0348)	(0.0321)
2.treatedsize	0.00827	0.00474	−0.000235	−0.0361
	(0.0344)	(0.0339)	(0.0336)	(0.0360)
3.treatedsize	−0.0458**	−0.0474**	−0.0511**	−0.0712**
	(0.0215)	(0.0213)	(0.0216)	(0.0281)
N	865	1,159	5,721	7,621

table continues next page

TABLE 4.6. Continued

	(1)	(2)	(3)	(4)
	WUPIall	**WUPIall**	**WUPIall**	**WUPIall**
1.treatedsize	2.416	2.809	2.854	2.924
	(2.275)	(2.195)	(2.171)	(2.245)
2.treatedsize	−0.425	−0.109	−0.0696	−0.0603
	(0.999)	(0.890)	(0.868)	(0.849)
3.treatedsize	−1.433	−1.182	−1.144	−1.001
	(1.528)	(1.434)	(1.413)	(1.672)
N	936	1,244	5,487	7,014
	(1)	**(2)**	**(3)**	**(4)**
	WUPIcoverage	**WUPIcoverage**	**WUPIcoverage**	**WUPIcoverage**
1.treatedsize	0.203	−0.128	0.188	−2.307
	(6.023)	(6.006)	(5.999)	(5.906)
2.treatedsize	−1.333	−1.657	−1.441	−3.219*
	(1.941)	(1.797)	(1.797)	(1.858)
3.treatedsize	−1.973	−2.228	−2.034	−3.639**
	(1.597)	(1.434)	(1.456)	(1.501)
N	936	1,244	5,487	7,014
	(1)	**(2)**	**(3)**	**(4)**
	WUPIquality	**WUPIquality**	**WUPIquality**	**WUPIquality**
1.treatedsize	−0.978	0.578	0.176	0.695
	(3.872)	(3.597)	(3.825)	(3.981)
2.treatedsize	−1.744	−0.398	−0.679	−0.387
	(1.379)	(0.975)	(1.065)	(1.214)
3.treatedsize	0.851	2.022	1.738	2.377
	(1.829)	(1.566)	(1.679)	(1.888)
N	915	1,223	2,718	4,209
	(1)	**(2)**	**(3)**	**(4)**
	WUPImgmt	**WUPImgmt**	**WUPImgmt**	**WUPImgmt**
1.treatedsize	2.662	3.300	3.196	4.785**
	(2.217)	(2.158)	(2.110)	(2.055)
2.treatedsize	0.931	1.451	1.404	2.510**
	(1.254)	(1.231)	(1.204)	(1.096)
3.treatedsize	−1.525	−1.157	−1.173	−0.141
	(2.268)	(2.176)	(2.158)	(2.341)
N	936	1,244	5,487	7,014

Note: Standard errors are in parentheses; 1. indicates the smallest initial structure, 2. indicates intermediate initial values, and 3. indicates the largest values of initial structure.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE 4.7. Difference-in-Differences: Conditional on Initial Volume

	(1) AVC2	(2) AVC2	(3) AVC2	(4) AVC2
1.treatedsize	−0.00165	−0.00603	−0.0119	−0.0366
	(0.0306)	(0.0304)	(0.0307)	(0.0321)
2.treatedsize	−0.00219	−0.00618	−0.0123	−0.0482
	(0.0356)	(0.0353)	(0.0354)	(0.0360)
3.treatedsize	−0.0427*	−0.0481*	−0.0542**	−0.0832***
	(0.0249)	(0.0252)	(0.0264)	(0.0240)
N	813	1107	5669	7569
	(1) WUPIall	**(2) WUPIall**	**(3) WUPIall**	**(4) WUPIall**
1.treatedsize	2.898*	3.163*	3.228**	3.278*
	(1.711)	(1.663)	(1.632)	(1.821)
2.treatedsize	−0.633	−0.313	−0.263	−0.575
	(0.967)	(0.833)	(0.810)	(0.792)
3.treatedsize	−2.329	−1.881	−1.873	−1.217
	(2.049)	(1.938)	(1.907)	(2.173)
N	883	1,191	5,434	6,961
	(1) WUPIcoverage	**(2) WUPIcoverage**	**(3) WUPIcoverage**	**(4) WUPIcoverage**
1.treatedsize	5.654	5.250	5.467	3.960
	(4.312)	(4.253)	(4.228)	(4.304)
2.treatedsize	−4.942***	−5.286***	−4.985***	−7.327***
	(1.848)	(1.675)	(1.693)	(1.661)
3.treatedsize	−1.160	−1.392	−1.277	−2.928
	(2.568)	(2.450)	(2.497)	(2.343)
N	883	1,191	5,434	6,961
	(1) WUPIquality	**(2) WUPIquality**	**(3) WUPIquality**	**(4) WUPIquality**
1.treatedsize	−0.969	0.364	0.197	−0.301
	(2.632)	(2.342)	(2.531)	(2.469)
2.treatedsize	0.277	1.722	1.367	1.935
	(1.771)	(1.434)	(1.558)	(1.781)
3.treatedsize	−2.351	−0.624	−0.922	−0.689
	(2.735)	(2.339)	(2.448)	(2.835)
N	867	1,175	2,670	4,161

table continues next page

TABLE 4.7. Continued

	(1) WUPImgmt	(2) WUPImgmt	(3) WUPImgmt	(4) WUPImgmt
1.treatedsize	1.728	2.213	2.163	3.136*
	(1.697)	(1.690)	(1.646)	(1.767)
2.treatedsize	1.727	2.258*	2.207*	3.173***
	(1.349)	(1.305)	(1.278)	(1.195)
3.treatedsize	−3.376	−2.772	−2.806	−1.004
	(2.676)	(2.564)	(2.535)	(2.860)
N	883	1,191	5,434	6,961

*Note:*Standard errors in parentheses, 1. indicates the smallest initial structure, 2. indicates intermediate initial values; and 3. Indicates the largest values; of initial structure.

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

TABLE 4.8. Difference-in-Differences: Conditional on Initial Density

	(1) AVC2	(2) AVC2	(3) AVC2	(4) AVC2
1.treatedsize	−0.0258	−0.0278	−0.0318	−0.0520
	(0.0266)	(0.0263)	(0.0266)	(0.0316)
2.treatedsize	0.00724	0.00342	−0.00190	−0.0436
	(0.0363)	(0.0361)	(0.0360)	(0.0372)
3.treatedsize	−0.00505	−0.0104	−0.0162	−0.0647***
	(0.0261)	(0.0259)	(0.0263)	(0.0228)
N	864	1,158	5,714	7,614
	(1) WUPIall	**(2) WUPIall**	**(3) WUPIall**	**(4) WUPIall**
1.treatedsize	2.299	2.541*	2.564*	2.764*
	(1.488)	(1.433)	(1.410)	(1.670)
2.treatedsize	−2.512**	−2.191**	−2.168**	−1.738
	(1.204)	(1.071)	(1.032)	(1.080)
3.treatedsize	0.743	1.036	1.077	0.869
	(1.010)	(0.911)	(0.896)	(0.919)
N	934	1,242	5,479	7,006
	(1) WUPIcoverage	**(2) WUPIcoverage**	**(3) WUPIcoverage**	**(4) WUPIcoverage**
1.treatedsize	4.995	4.665	4.846	3.399
	(3.671)	(3.583)	(3.560)	(3.666)
2.treatedsize	−4.650**	−5.062***	−4.858***	−6.562***
	(1.818)	(1.614)	(1.634)	(1.589)
3.treatedsize	−4.945**	−5.235**	−4.950**	−7.090***
	(2.485)	(2.467)	(2.477)	(2.353)
N	934	1,242	5,479	7,006

table continues next page

TABLE 4.8. Continued

	(1) WUPIquality	(2) WUPIquality	(3) WUPIquality	(4) WUPIquality
1.treatedsize	−1.620	−0.347	−0.604	−0.219
	(2.084)	(1.813)	(1.945)	(2.062)
2.treatedsize	−2.047	−0.537	−0.818	−0.527
	(1.746)	(1.333)	(1.463)	(1.494)
3.treatedsize	3.074	4.304	3.854	4.678
	(3.273)	(3.072)	(3.225)	(3.608)
N	913	1,221	2,710	4,201
	(1) WUPImgmt	**(2) WUPImgmt**	**(3) WUPImgmt**	**(4) WUPImgmt**
1.treatedsize	1.447	1.839	1.787	2.855*
	(1.577)	(1.550)	(1.518)	(1.677)
2.treatedsize	−1.475	−0.924	−1.013	0.559
	(1.659)	(1.602)	(1.547)	(1.675)
3.treatedsize	3.266***	3.778***	3.763***	4.964***
	(1.472)	(1.421)	(1.383)	(1.372)
N	934	1,242	5,479	7,006

Note: Standard errors are in parentheses; 1. indicates the smallest initial structure; 2. Indicates intermediate initial values; and 3. indicates the largest values of initial structure.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

medium and large density upon aggregation. Large-density utilities, however, can compensate for this decrease with improvements in WUPImgmt, yielding no decrease in the aggregate WUPIall.

It is important to note that the obtained results are sample specific and that the conditionalities are not easy to interpret: there may well be a correlation between the magnitude of aggregation and the initial structure, which could confound both sets of results. For instance, it is unlikely in aggregation reforms that utilities with low volume and density would be tasked to take over multiple other possibly large-volume utilities. For this reason, an observation of an improvement in unit cost for large-volume utilities should not be interpreted too narrowly because the types of aggregations probably vary with the initial structure. Moreover, the effects are identified on a subsample of the initial 75 aggregations, a matter that makes the findings very sample specific. Nevertheless, the effects of the aggregations seem to depend to some extent on the initial structure of the utilities before the aggregation. Positive effects are not limited to small utilities but can also accrue to large ones, albeit showing in different performance indicators.

4.4 Postaggregation Performance

In addition to comparing the performance of utilities before and after the aggregation, one gains insightful by looking at the postmerger evolution of performance in particular. An aggregation might entail a change in performance when utilities are aggregated simply because the new integrated utility experiences the (weighted) average of the previous performance. For example, adding many rural utilities with low degrees of connection might decrease the indicator for WUPIcoverage despite no actual change. Because data on the integrated utilities are not available, a second-best strategy is to look at the performance evolution after the aggregation. The idea is that discarding initial performance and therefore any detrimental shocks through the aggregation could enable advantages through the aggregation to then materialize over the years after the aggregation.[8]

The results are displayed in table 4.9 Similar to the difference-in-differences estimates, the postaggregation estimations show little effect by aggregations on average. The panel at the bottom of table 4.9 indicates a slightly positive effect on WUPImgmt, which is, however, too weak to show up in the overall WUPIall score. That the aggregations can have a small but positive effect on WUPImgmt in the postaggregation years is supported by a number of specifications that differentiate the magnitude of the reform: first, moderate and large increases in the number of towns are associated with higher growth in WUPImgmt. Second, large relative increases in density also seem to lead to more improvement in WUPImgmt. The latter effect is sufficiently large to show up in an improvement in WUPIall. To summarize, although the impact is generally neither positive nor negative, in some cases aggregations helped to improve the growth in WUPImgmt.

TABLE 4.9. Postaggregation Phase

	(1) D.AVC2	(2) D.AVC2	(3) D.AVC2	(4) D.AVC2
1.after	−0.0000858	0.00115	0.00123	−0.00124
	(0.00572)	(0.00568)	(0.00581)	(0.00523)
N	639	759	1,848	5,700
	(1)	**(2)**	**(3)**	**(4)**
	D.WUPIall	**D.WUPIall**	**D.WUPIall**	**D.WUPIall**
1.after	0.246	0.323	0.270	0.427
	(0.314)	(0.308)	(0.312)	(0.300)
N	685	808	1,749	4,973
	(1)	**(2)**	**(3)**	**(4)**
	D.WUPIcoverage	**D.WUPIcoverage**	**D.WUPIcoverage**	**D.WUPIcoverage**
1.after	0.0288	0.0353	−0.0348	−0.0477
	(0.613)	(0.610)	(0.612)	(0.549)
N	685	808	1,749	4,973
	(1)	**(2)**	**(3)**	**(4)**
	D.WUPIquality	**D.WUPIquality**	**D.WUPIquality**	**D.WUPIquality**
1.after	−0.342	−0.180	−0.149	−0.287
	(0.475)	(0.421)	(0.465)	(0.406)
N	667	790	885	2,946
	(1)	**(2)**	**(3)**	**(4)**
	D.WUPImgmt	**D.WUPImgmt**	**D.WUPImgmt**	**D.WUPImgmt**
1.after	0.561	0.662*	0.579	0.626*
	(0.378)	(0.374)	(0.373)	(0.360)
N	685	808	1,749	4,973

Note: Standard errors are in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

With respect to the initial structure, the postaggregation performance seems to vary little. Among the few results, utilities with only two towns before the aggregation can improve their WUPImgmt and through that improvement also raise their aggregate WUPIall score. Utilities with large densities are also able to improve their WUPImgmt. Overall, it might not be surprising that many of the postmerger results are related to WUPImgmt, simply because, for example, decisions related to metering and collection are somewhat more easily altered than reducing the labor force or improving quality and coverage.[9] The analysis of the postmerger phase also shows that some of the previous aggregation effects seem to be driven by immediate changes in the wake of the aggregation.

4.5 Alternative Aggregation Measures

In this section of the empirical analysis, an alternative measure for aggregations is used. The main motivation for choosing an alternative lies in the question of how closely the indicator "number of towns" in IBNET measures whether a utility has increased its service area. Although a change in number of towns might capture pure mergers well, other reorganizations in which a utility takes over a service area without taking over a town, at least by definition, may be missed. For this reason, a change in the population in the service area is considered as an alternative here. Although changes in population are quite frequent, in the absence of data errors, a large change in the service area population can be explained only by an enlargement or aggregation of the service area. The year-by-year change considered here is 20 percent—that might seem excessive, but the results are not sensitive to this choice because 10 percent and 15 percent lead to similar conclusions.

As the treatment indicator changes, also the control group has to be adapted. Accordingly, the matching procedures are repeated but with a different dependent variable: a change in the service area population exceeding 20 percent instead of a change in the number of towns. As before, utilities were eliminated beforehand if the population in the service area decreased in the sample period to ensure a meaningful comparison.

The results are exhibited in table 4.10. In contrast to before, on average the aggregations here appear to

TABLE 4.10. Alternative Merger Indicator

	(1)	(2)	(3)	(4)
	AVC2	**AVC2**	**AVC2**	**AVC2**
1.after	−0.0536**	−0.0371**	−0.0303*	−0.0300
	(0.0237)	(0.0178)	(0.0170)	(0.0205)
N	566	846	1,602	2,624
	(1)	**(2)**	**(3)**	**(4)**
	WUPIall	**WUPIall**	**WUPIall**	**WUPIall**
1.after	0.108	−0.351	−1.132	−1.464
	(1.066)	(1.038)	(1.005)	(1.180)
N	553	833	1,557	2,394
	(1)	**(2)**	**(3)**	**(4)**
	WUPIcoverage	**WUPIcoverage**	**WUPIcoverage**	**WUPIcoverage**
1.after	−3.234**	−3.126**	−4.348***	−4.086**
	(1.405)	(1.495)	(1.375)	(1.780)
N	553	833	1,557	2,394

table continues next page

TABLE 4.10. Continued

	(1) WUPIquality	(2) WUPIquality	(3) WUPIquality	(4) WUPIquality
1.after	0.0214	−0.918	−1.065	−4.589
	(3.651)	(3.353)	(3.164)	(3.715)
N	334	474	884	1,632
	(1) WUPImgmt	**(2) WUPImgmt**	**(3) WUPImgmt**	**(4) WUPImgmt**
1.after	2.023*	1.158	0.694	0.407
	(1.199)	(1.063)	(1.091)	(1.311)
N	553	833	1,557	2,394

Note: Standard errors are in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

affect utility performance. First, there is a sizeable negative effect on unit cost. Although the number of aggregations to identify the conditional effects is partly very small, the results suggest that aggregations with medium to large increases in volume exhibit the largest unit cost reductions. This finding is probably of little surprise, because volume enters the denominator of unit cost. Still it confirms that unit cost reductions depend on a sizeable increase in volume. Moreover, the unit cost decreases also seem to be associated with larger utilities in terms of towns, volume, and density.

There are also some indications that WUPIcoverage decreases because of an increase in the population in the service area. Most commonly the coverage decreases when density decreases most or with utilities that already serve many towns. Finally, WUPImgmt can rise through aggregations when utilities have a high density and high volume.

4.6 Distinguishing Strong and Weak Utilities

This final subsection of the empirical analysis addresses the question of whether apart from structure, the initial performance of a utility matters. If aggregations are considered a general reform strategy to improve quality and to leave a low-level equilibrium, one would expect different results for utilities with different starting positions in terms of service quality. The previous results showed that smaller utilities in terms of customers tend to benefit through aggregations by improving WUPI. Because small utilities are also typically those that have a lower initial WUPI score, this result could also indicate that aggregations are particularly beneficial to utilities with low performance.

This intuition is confirmed by statistical tests that differentiate utilities depending on the initial level of quality (see table 4.11).[10] Utilities with the lowest initial WUPI compared with utilities in the same country exhibit larger improvements in the postaggregation phase than do nonmerging utilities. This observation is true for both managerial efficiency and the overall performance indicator (WUPI). In contrast, utilities with higher initial WUPI do not exhibit such an improvement through aggregations. There is even some evidence that utilities with high initial WUPI experience lower improvement in coverage—and through this also lower overall WUPI—in the postaggregation phase.

The fact that no results are found with respect to cost suggests that utilities with initially low quality (measured by WUPI) can benefit from aggregations by improving quality, though the aggregation does not

TABLE 4.11. Difference-in-Differences: Conditional on Initial Performance (WUPI)

	(1)	(2)	(3)	(4)
	AVC2	**AVC2**	**AVC2**	**AVC2**
1.treatedweak	−0.0115*	−0.0101	−0.0103	−0.0116*
	(0.00664)	(0.00655)	(0.00667)	(0.00655)
2.treatednormal	0.0102	0.0114	0.0112	0.00560
	(0.00977)	(0.00982)	(0.00988)	(0.00927)
3.treatedstrong	−0.00197	−0.000868	−0.000838	−0.00250
	(0.00504)	(0.00487)	(0.00496)	(0.00570)
N	608	728	1,817	5,472
	(1)	**(2)**	**(3)**	**(4)**
	WUPIall	**WUPIall**	**WUPIall**	**WUPIall**
1.treatedweak	0.638*	0.726**	0.665**	0.856**
	(0.346)	(0.334)	(0.337)	(0.336)
2.treatednormal	0.123	0.216	0.165	0.332
	(0.483)	(0.475)	(0.483)	(0.485)
3.treatedstrong	−0.567**	−0.502*	−0.562**	−0.409
	(0.276)	(0.269)	(0.272)	(0.271)
N	659	782	1,723	4,947
	(1)	**(2)**	**(3)**	**(4)**
	WUPIcoverage	**WUPIcoverage**	**WUPIcoverage**	**WUPIcoverage**
1.treatedweak	0.801	0.797	0.711	0.736
	(0.589)	(0.583)	(0.578)	(0.546)
2.treatednormal	−0.207	−0.206	−0.284	−0.342
	(0.896)	(0.894)	(0.898)	(0.812)
3.treatedstrong	−1.420*	−1.407*	−1.505*	−1.440*
	(0.827)	(0.817)	(0.810)	(0.788)
N	659	782	1,723	4,947
	(1)	**(2)**	**(3)**	**(4)**
	WUPIquality	**WUPIquality**	**WUPIquality**	**WUPIquality**
1.treatedweak	−0.834	−0.649	−0.634	−0.674
	(0.621)	(0.566)	(0.627)	(0.485)
2.treatednormal	−0.239	−0.0317	−0.0284	−0.0794
	(0.597)	(0.533)	(0.590)	(0.493)
3.treatedstrong	−0.251	−0.136	−0.0950	0.00460
	(0.614)	(0.567)	(0.584)	(0.535)
N	641	764	859	2,920

table continues next page

TABLE 4.11. Continued

	(1) WUPImgmt	(2) WUPImgmt	(3) WUPImgmt	(4) WUPImgmt
1.treatedweak	0.878*	1.003**	0.913**	0.952**
	(0.471)	(0.461)	(0.459)	(0.473)
2.treatednormal	0.557	0.678	0.612	0.666
	(0.606)	(0.599)	(0.604)	(0.583)
3.treatedstrong	−0.180	−0.0976	−0.190	−0.238
	(0.329)	(0.329)	(0.319)	(0.313)
N	659	782	1,723	4,947

Notes: Standard errors are in parentheses; 1. indicates the smallest initial structure; 2. indicates intermediate initial values; and 3. Indicates the largest values of initial structure.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

lead to cost savings. This observation is consistent with the view of aggregations as a reform option to enable utilities to leave a low-level equilibrium. The finding that utilities with low initial WUPI improve performance is related to the previous finding that small utilities seem to be able to improve WUPI through aggregations: utilities with low WUPI are more frequent in the group of small utilities (few customers), and therefore it is not surprising that the empirical results show that small utilities and those with low WUPI can improve WUPI scores through aggregations.

Notes

1. See Angrist and Pischke (2009) or Wooldridge (2010) for comprehensive approaches to the treatment effect literature.
2. Unlike before, the number of customers is replaced by the volume produced by a utility. Although the overall results are unaffected by this choice, the goal is to be comparable to the bulk of existing research, which has focused on volume
3. In the case of multiple comparison utilities, the weights are adjusted accordingly.
4. For more detailed information on the construction and background, see World Bank/IAWD (2015).
5. In both cases, the groups are less than −5 percent, between −5 percent and 5 percent, and above 5 percent.
6. In most cases the biases are also statistically significant (not shown in the table).
7. The results are not shown but are available upon request.
8. This means to drop all preaggregation observations. Because with utility fixed effects the aggregation effect is no longer estimable, the dependent variable is transformed to first differences. The observations are now pooled and aggregating utilities identified by a dummy variable.
9. Results for the conditional effects are omitted but are available upon request.
10. Given the previous arguments that the aggregation may initially lower WUPI, the focus of this analysis is again on the postaggregation period.

References

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Caliendo, Marco, and Sabine Kopeinig. 2008. “Some Practical Guidance for the Implementation of Propensity Score Matching.” *Journal of Economic Surveys* 22 (1): 31–72.

Garcia, Serge, and Alban Thomas. 2001. “The Structure of Municipal Water Supply Costs: Application to a Panel of French Local Communities.” *Journal of Productivity Analysis* 16 (1): 5–29.

Michaud, David, Maria Salvetti, Michael Klien, Berenice Flores, Gustavo Ferro, and Stjepan Gabric. 2017. *Joining Forces for Better Services? When, Why, and How Water and Sanitation Utilities Can Benefit from Working Together.* Washington, DC: World Bank.

Rosenbaum, Paul R., and Donald B. Rubin. 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score." *American Statistician* 39(1): 33-38.

Rubin, Donald B. 2001. "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation." *Health Services and Outcomes Research Methodology* 2 (3-4): 169-88.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. Cambridge, MA: MIT Press.

World Bank/IAWD (International Association of Water Supply Companies in the Danube River Catchment Area). 2015. Water and Wastewater Services in the Danube Region: A State of the Sector. Regional Report. Washington, DC: International Bank for Reconstruction and Development/ The World Bank. http://sos.danubis.org/files/File/SoS_Report.pdf.

Chapter 5
Discussion and Conclusion

The results of the causal analysis suggest that on average, the analyzed aggregations have had no systematic effect on cost or other performance measures. Nevertheless, more nuanced additional tests, conditioning the aggregation effect on either (a) the magnitude of the reform or (b) the initial structure of the utility, suggest that the effect varies considerably across aggregations.

Some of the findings correspond with the preceding analysis of utility structure. For instance, utilities with initially low densities and volume can benefit from aggregations by improving performance. Moreover, reforms that add only a few towns were found more helpful cost-wise than others. On the other hand, utilities that were already large to begin with, both in terms of volume and number of towns, still appear to be able to benefit from cost savings due to economies of scale. Finally, a number of results found that WUPImgmt, which measures the financial and managerial performance, also experienced positive developments in the postaggregation period. Although the original conditions can only be speculated, the improvements in WUPImgmt were frequently related to utilities that were initially high in density and volume. Considering that this indicator might be changed more easily and quickly than coverage or quality—which relate to the whole network and to changes in the infrastructure—the finding highlights that aggregations can benefit already-large utilities but possibly more in areas other than cost. On top of that, these results indicate the potential of aggregations to achieve improvements in the short term.

The results are quite clear in the sense that the finding of no average effect seems to be driven by a strongly heterogeneous treatment effect. In some cases, the effects of the aggregations were positive in the sense of cost or performance; in other cases, the effects were negative. Therefore, the findings also call into question the simple logic of economies of scale, which assumes that small utilities would always benefit from aggregations and that large utilities would experience diseconomies. The fact that small utilities did not consistently show more favorable results from aggregations than did large utilities is particularly striking and suggests that the process and type of reform may also matter greatly. Therefore, although clear-cut predictions on which type of aggregation or which initial structure is most appropriate are beyond the possibilities of this analysis, the results strongly suggest that there are factors related to utility structure that can aid or hamper the success of aggregations.

The empirical analysis is, however, subject to a number of limitations. Most prominently, there are a number of points that cannot be tackled with the underlying data. First, the analysis is rather short term (in the sense of observation period after the aggregation). It should still be noted that in the long term, the overall cost effects could be different from what is measured here by looking at variable cost, because the structure of the supply system might be adapted to the larger network after an aggregation. Second, there is no information on the "acquired" utilities. Depending on the initial state of these utilities, the aggregation results might differ considerably. For example, some of the results on reduced coverage could possibly be explained by the fact that the merged utilities exhibited a lower coverage than the acquiring utility.

Related, the measured average effects after the matching are sometimes called "average treatment effect on the treated," which suggests that it measures the effect if the aggregating utilities would not have aggregated. Because aggregating utilities were often larger in many dimensions, this effect is a poor picture of what would happen if the reforms were targeted to more average or smaller utilities. Although differentiating the treatment effects with respect to initial utility size tries to deal with this issue, the problem may still prevail.

Finally, the design and process of an aggregation reform not only determines how the structure of a utility changes (for example, to which cluster a utility moves), but also affects a large number of other factors such as the allocation of control and decision rights. Many such factors are not measured in IBNET and therefore cannot be analyzed in this section. In particular, the various purposes of the analyzed aggregations are missing. While the focus on structural characteristics is certainly warranted, the sole focus on that dimension is a clear limitation of this statistical analysis.

To conclude, the analysis stresses the importance of the reform design. More important than whether to aggregate or not seems the way a utility is transformed through the reform, and which utilities are affected. The case-studies featured in Michaud and others (2017) can help to fill this gap by shedding light on the way aggregations are implemented. This work will go beyond the purely structural analysis in this report and can also pinpoint less visible but nevertheless decisive changes. Moreover, whereas this report could only allude to the mechanism through which aggregations affect performance (such as the input mix and cost shares), the case studies should be able to highlight more in detail the channels that were responsible for success or failure of aggregations.

Reference

Michaud, David, Maria Salvetti, Michael Klien, Berenice Flores, Gustavo Ferro, and Stjepan Gabric. 2017. *Joining Forces for Better Services? When, Why, and How Water and Sanitation Utilities Can Benefit from Working Together.* Washington, DC: World Bank.

Appendix A
IBNET Data

The main data for this analysis stem from the International Benchmarking Network (IBNET) database. IBNET is a data repository initiated and maintained by the World Bank with the objective to improve the service delivery of water supply and sewerage utilities through the provision of international comparative benchmark performance information.

The utility coverage by IBNET varies strongly among countries, both in terms of the number of utilities and in the population living in the service area of the utilities. Because the main objective of this study is to measure the effect of consolidations, some particular utilities were excluded. The main idea was to restrict the comparisons to cases in which utilities consolidate—increase the number of served towns—and utilities that keep the number of served towns stable. For this reason, utilities that experienced a reduction in the number of served towns were excluded, even if they followed or were preceded by an increase. Those utilities or part of them might be integrated into other firms in our sample and could therefore blur the effect we try to identify.

The eventual number of covered utilities by country is exhibited in table A.1. The 1,306 utilities span an unbalanced panel of 8,059 utility-year observations from 1995 to 2015. Summary statistics of the used variables are displayed in table A.2.

TABLE A.1. Utilities per Country, by Treatment Status

Country	Not aggregated	Aggregated	Total
Albania	27	0	27
American Samoa	0	1	1
Argentina	5	0	5
Armenia	1	2	3
Belarus	27	2	29
Bolivia	2	0	2
Bosnia and Herzegovina	32	2	34
Brazil	629	2	631
Chile	4	1	5
Côte d'Ivoire	0	1	1
Croatia	11	1	12
Czech Republic	3	4	7
Ecuador	1	0	1
Egypt, Arab Rep.	18	0	18
Fiji	1	0	1
Georgia	17	2	19
Guam	1	0	1

table continues next page

TABLE A.1. Continued

Country	Not aggregated	Aggregated	Total
Honduras	1	0	1
Hungary	11	6	17
Jordan	2	0	2
Kazakhstan	22	5	27
Kiribati	1	0	1
Korea, Rep	160	0	160
Kosovo	6	0	6
Kyrgyz Republic	11	2	13
Lithuania	35	2	37
Macedonia, FYR	10	4	14
Marshall Islands	1	0	1
Mexico	7	2	9
Micronesia, Fed. Sts.	2	0	2
Moldova	29	0	29
Mongolia	1	0	1
Montenegro	5	0	5
Namibia	1	0	1
Norway	1	0	1
Pakistan	2	0	2
Panama	1	0	1
Papua New Guinea	1	1	2
Poland	3	12	15
Romania	8	15	23
Russian Federation	38	3	41
Samoa	0	1	1
Serbia	16	5	21
Singapore	1	0	1
Slovak Republic	7	1	8
Solomon Islands	1	0	1
South Africa	8	0	8
Swaziland	1	0	1
Tajikistan	8	0	8
Turkey	5	0	5
Uganda	1	0	1

table continues next page

TABLE A.1. Continued

Country		Not aggregated	Aggregated	Total
Ukraine		26	0	26
Uzbekistan		6	1	7
West Bank and Gaza		1	1	2
Yemen, Rep.		6	0	6
Zambia		2	0	2
Total	49	1,227	79	1,306

TABLE A.2. Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
AVC	0.28	0.21	0	1.89	7,769
after	0.05	0.23	0	1	8,059
WUPIall	70.44	12.21	20.16	99.95	7,152
WUPIcoverage	60.82	18.91	3.92	100	7,152
WUPIquality	86.52	20.42	0	100	4,336
WUPImgmt	74.05	14.91	8.52	100	7,152
vol_both	30,201,859.37	76,735,798.81	40,000	996,000,000	8,008
dens_both	307.53	225.62	34.33	2,977.94	8,059
towns_both	7.34	35.11	2	1,187	8,059

Note: AVC = average variable cost; after = utilities after aggregation; WUPIall = performance indicator including all subindicators; WUPIcoverage = performance indicator for coverage; WUPIquality = performance indicator for quality; WUPImgmt = performance indicator for management; vol_both = volume of both water and wastewater; dens_both = density of both water and wastewaster; towns_both = number of towns for both water and wastewater.

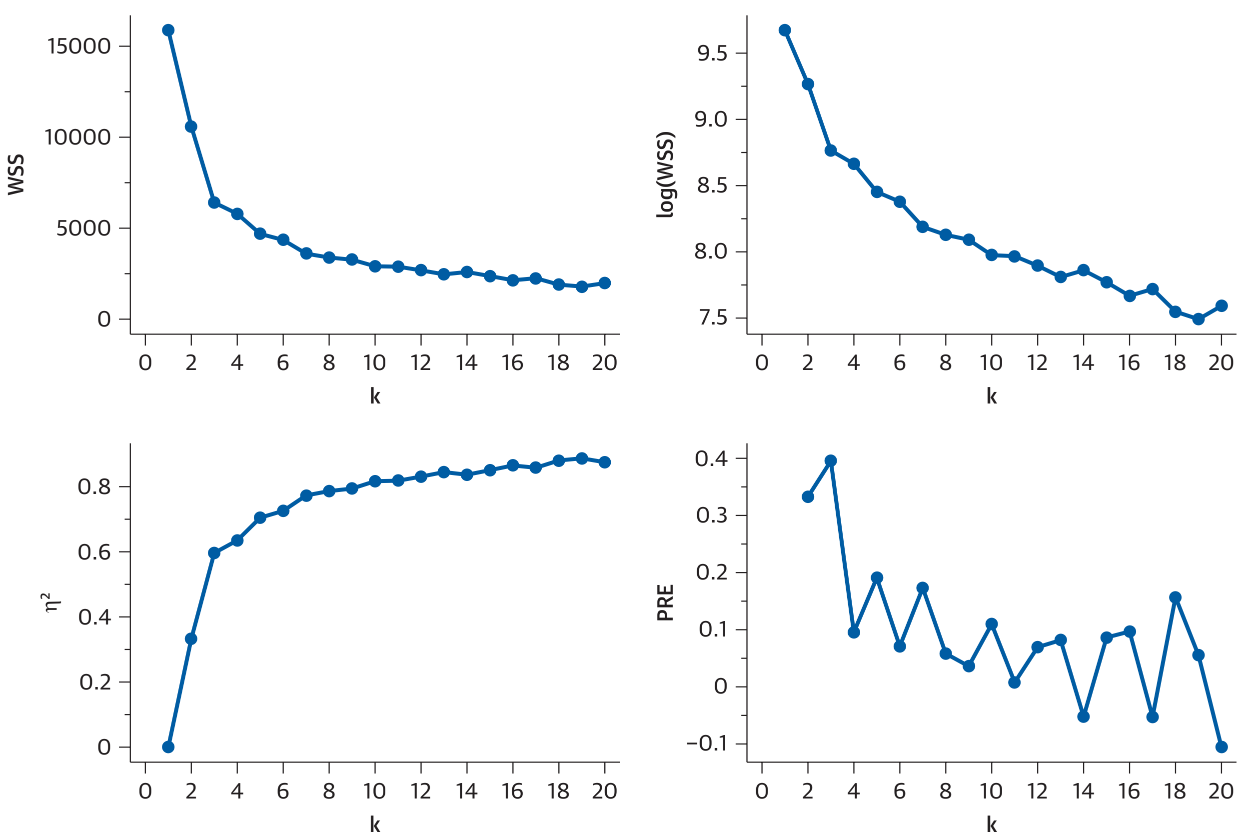
Appendix B Methodological Details of Clustering

In this section the approach used for the cluster analysis is briefly described. First, it should be noted that the three input variables—volume, density, and number of towns—are transformed by taking the natural log and then standardizing the variable. The resulting variable has mean 0 and a standard deviation of 1.

This last step is taken to ensure that the different measurements do not affect the cluster results. Without the transformation, variables with higher variation would tend to have a higher influence on the clustering. Taking the natural log is motivated by the fact that the distribution is highly right skewed. After taking the natural log, the dispersion is much reduced.

Second, because the number of clusters in k-means clustering is somewhat ad hoc (similar to the cut-off in hierarchical clustering), we compute a number of test statistics for the choice of the number of clusters. Following Makles (2012), we compute within cluster sum of squares (WSS), its logarithm (ln(WSS)), as well two measures of fit (η^2 and PRE, proportional reduction of error). The results are displayed in figure B.1. What is apparent in the graph is that a minimum of three clusters is necessary to capture a large share of

FIGURE B.1. Tests Statistics for K-Means Cluster Choice



Note: The top panels exhibit within sum of squares and the log of within sum of squares for different numbers of k. The lower panels show η^2 and PRE.

the cluster variation. Moreover, after seven clusters, the gains in fit by adding more clusters are very small. After experimenting with the results and the distinct clusters generated by different k's, the number of six final clusters was chosen. Less than six would mean obtaining considerably more heterogeneous clusters, whereas having more than six does not add to the interpretation and distinctness of the clusters. Adding a seventh cluster would further differentiate the utilities serving a single town.

Third, because k-means clustering is sensitive to the initial cluster allocation of individual utilities, first a hierarchical clustering is run and the associated results are used as a starting classification for the k-means clustering. For the hierarchical clustering, Euclidean distance and an average-linkage algorithm is used.

Reference

Makles, Anna. 2012. "Stata Tip 110: How to get the Optimal K-means Cluster Solution." *Stata Journal* 12 (2) :347–51.

SKU W17061

WORLD BANK GROUP