

Economies of Scale and Technical Efficiency in Community Water Systems

Jhih-Shyang Shih, Winston Harrington,
William A. Pizer, and Kenneth Gillingham

February 2004 • Discussion Paper 04–15



RESOURCES
FOR THE FUTURE

Resources for the Future
1616 P Street, NW
Washington, D.C. 20036
Telephone: 202–328–5000
Fax: 202–939–3460
Internet: <http://www.rff.org>

© 2004 Resources for the Future. All rights reserved. No portion of this paper may be reproduced without permission of the authors.

Discussion papers are research materials circulated by their authors for purposes of information and discussion. They have not necessarily undergone formal peer review or editorial treatment.

Economies of Scale and Technical Efficiency in Community Water Systems

Jhih-Shyang Shih, Winston Harrington, William A. Pizer, and Kenneth Gillingham

Abstract

In this study we use datasets from the 1995 and 2000 Community Water Supply surveys to examine the production costs of water supply systems. We first estimate the economies of scale in water supply by estimating the total unit cost as well as individual component cost elasticities. For total unit cost elasticity, we find that a 1% increase in production reduces unit costs by a statistically significant 0.16%. For individual component cost elasticities, we find that higher economies of scale exist in capital costs, outside costs, other costs, and materials costs; labor costs and energy costs exhibit lower but still positive economies of scale. These economies of scale may reflect production economies or suggest that larger systems are better than smaller systems at bargaining and can obtain inputs at a lower unit cost. Importantly, bargaining gains and some production economies do not necessarily depend on water systems' becoming physically interconnected.

Key Words: small water systems; water supply; capacity development; economies of scale; community water systems

JEL Classification Numbers: Q25, Q28

Contents

1. Introduction.....	1
2. Economy of Scale, Productivity, and Technical Efficiency.....	3
2.1 Efficiency Measurement with Multiple Inputs/Outputs	5
2.2 Scale Economies	6
3. Datasets and Variables	7
3.1 The 1995 CWSS Dataset	10
3.2 The 2000 CWSS Dataset	11
4. Results	12
4.1 Economies of Scale in Water Supply.....	12
4.2 Technical Efficiency	18
4.3 Comparative Analysis of 1995 and 2000 CWSS	22
5. Caveats	23
6. Cost Savings from Consolidation	25
7. Conclusions.....	28
References.....	30
Appendix A: Data Envelopment Analysis (DEA) Model	32
Appendix B: Some Recommendations for CWSS.....	34

Economies of Scale and Technical Efficiency in Community Water Systems¹

Jhih-Shyang Shih, Winston Harrington, William A. Pizer, and Kenneth Gillingham²

1. Introduction

Small water systems are facing increasingly stringent regulatory requirements under the Safe Drinking Water Act as amended in 1996. According to the National Public Water Systems Compliance Report, there are more than 54,000 publicly and privately owned community water systems (CWSs) in the country, serving about 252 million people (EPA, 1997a). Approximately 93 percent of CWSs are categorized as “small” or “very small,” serving fewer than 3,300 customers (EPA, 2003). Although these systems serve only 20 percent of the total population served by all community water systems, they have received much attention from federal regulators and state and local health officials because they face particular difficulties in complying with federal and state water quality requirements. Their size makes it difficult for them to have the technical, managerial and financial capacity that modern water treatment systems require. For example, systems serving 25-500 persons have the most violations per 1,000 people served among all of the size categories.

Small systems are also thought by many to have high costs of supplying customers with water (See Beecher et al. 2002, for a review of likely mechanisms). There are at least two distinct kinds of scale economies in water supply systems. The most familiar are scale economies in capital equipment. If consolidation is to achieve these economies, the systems involved must be geographically close, so that they can be connected to the same water treatment plants. There are

¹ This publication was developed under Cooperative Agreement No. 82925801-0 awarded by the U.S. Environmental Protection Agency (EPA). EPA offered comments and suggestions to improve the scientific analysis and technical accuracy of this document. However, the views expressed here are those of authors and not of RFF. EPA does not endorse any products or commercial services mentioned in this publication.

² Fellow, Senior Fellow, Fellow, and Research Assistant, respectively, Quality of the Environment Division, Resources for the Future, 1616 P St. NW, Washington, DC 20036. Thanks to Ray Kopp, Julie Hewitt, and Kris Wernstedt for valuable suggestions at the beginning of this research. This research benefited from support by the EPA Office of Water and from discussions with John Bennet, Evyonne Harris, Deborah Mccray, Michael Osinski, Carl Reeverts, Brian Rourke, Peter Shanaghan, David Travers, and Kathleen Vokes. Thanks to Richard A. Krop for providing CWSS datasets. The authors are solely responsible for any errors.

also scale economies in many ordinary business operations, such as billing, purchasing, and computer systems, as well as in ancillary water treatment and testing operations. Achieving scale economies thus might not require systems to be physically connected. For example, material cost may show scale economies because large systems may be able to negotiate better long-term contract. On the other hand, capital and energy costs may be very sensitive to physical connection.

Consolidation of water systems—the merging of smaller systems or the absorption of one or more small systems by a larger one—may be a way to reduce the cost of water supply and to improve the ability of these systems to meet more stringent regulatory requirements cost-effectively. The benefits are potentially large (AWWA, 1997; Cadmus, 2002), yet of all local and municipal services, water supply remains almost uniquely *unconsolidated*. Why this is so is beyond the scope of this report, but the existence of consolidation benefits indicates that it may be time to reexamine the issue, in particular to determine if there are policies that could enhance the benefits of consolidation while reducing their costs.

In this report we use the 1995 and 2000 Community Water Supply Surveys (CWSSs) to examine the potential for achieving reductions in unit costs of water supply by increasing system size, and in particular by consolidating existing systems. To achieve this we need to distinguish system size from other causes of cost variation. System size is only one of a multitude of variables affecting water supply cost. There are differences in the cost of raw water supply, depending on climate, topography, and geology. The quality of the raw water may affect cost, as some raw water supplies will require more expensive treatment than others to meet acceptable health and portability standards. The spatial distribution of the final demand for water will also affect the costs in the distribution system, with higher population densities enabling the fixed costs of the distribution system to be spread over a greater number of accounts. In addition, there may be differences in the *efficiency* of water supply systems, with some systems getting more output than others from the same quantity of inputs. Other things being equal, more efficient systems will have lower costs. Our principal concern is to attempt to separate, to the extent possible, the cost elements attributable to scale from everything else, although data limitations make that difficult³.

³ We discuss some of the data limitations in Appendix B.

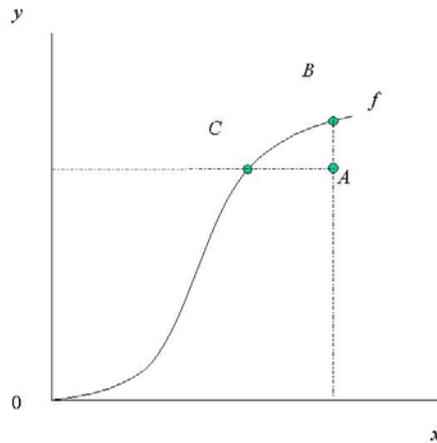
The structure of the paper is as follows. Section 2 reviews economy of scale, productivity, and technical efficiency theory. Section 3 discusses the CWSS 1995 and 2000 datasets. Section 4 presents 1995 and 2000 analysis results, respectively. Section 5 discusses some caveats. Section 6 reports our cost saving potential calculation. Section 7 gives the conclusions.

2. Economy of Scale, Productivity, and Technical Efficiency

In the production economics literature, there are two basic approaches to estimating production relations (Coelli et. al., 1998). The first approach assumes that all decisionmaking units (DMU, such as firms, plants, water systems) are technically efficient and use econometric estimation and index methods to study the aggregate technical change, return to scale, and optimization rules. That is, it assumes the firms are operating on the production possibility frontier and that no further output is technically possible with the given level of inputs. Observed variation along the frontier is assumed to be noise. The second approach does not assume that all decisionmaking units are technically efficient. The goal is to first identify the “technology frontier,” the maximum output achievable from a given set of inputs. DMUs on this frontier are said to be “technically efficient.” Then, efficiency analysis is concerned with the degree to which other DMUs lie inside the frontier and/or use the cost-minimizing set of inputs. In the empirical research below, we draw from both approaches. The current section provides a brief conceptual introduction.

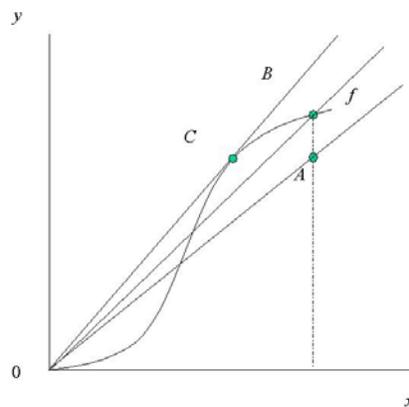
Figure 1 shows a simple production process in which a single input (x) is used to produce a single output (y). The curve Of represents the *production frontier*, which is the maximum output attainable from each input level. It reflects the current state of technology in the industry. All points between the production frontier and the x -axis form the feasible production set. Technically efficient DMUs operate on the frontier, and inefficient ones operate within it. For example, Point A represents an inefficient point whereas points B and C represent efficient points. A DMU operating at point A is inefficient because technically it could increase output to the level associated with the point B without requiring more input or it could reduce input to the level associated with the point C without reducing any output production. With more than one input the concept is the same, but the figure has three or more dimensions.

Figure 1
Production Frontier



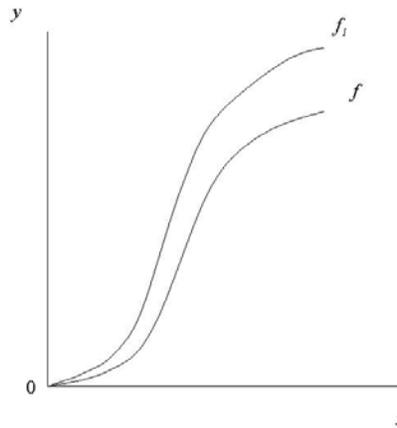
In Figure 2, we use a ray through the origin to measure productivity at a particular data point. The slope of this ray is y/x (output/input) and hence provides a measure of productivity. If the DMU operating at point A were to move to the technically efficient point B , the slope of the ray would be greater, implying higher productivity at point B . However, by moving to the point C , the ray from the origin is at a tangent to the production frontier and defines the point of maximum scale economies. From this figure we conclude that a DMU may be technically efficient but it may still be able to improve its productivity by exploiting scale economies.

Figure 2
Productivity, Efficiency, and Scale Economies



When one considers productivity comparison through time, an additional source of productivity change through technical change, is possible. As shown in the following Figure 3, the advances in technology which can be represented by an upward shift in the production frontier from Of in period 0 to Of_1 in period 1. In period 1, all DMUs can technically produce more output for each level of input, relative to what was possible in period 0.

Figure 3
Technical Change

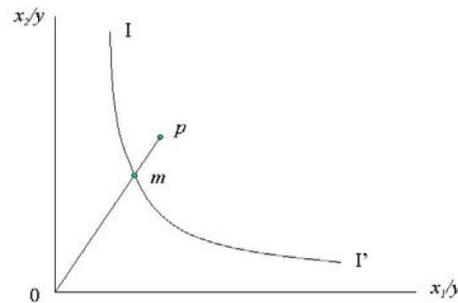


Increased productivity may be due to three factors: increased production efficiency, technical change or exploitation of scale economies.

2.1 Efficiency Measurement with Multiple Inputs/Outputs

So far we have discussed the efficiency concept using a single input and single output. For the multiple input case, Farrell (1957) proposed measuring efficiency through a set of assumptions that could be implemented by linear programming techniques. To illustrate the main concept underlying the proposed method, let us consider two inputs and one output and use water systems as an example. Given a sample of water systems, each characterized by the same production function of output, water quantity produced, y , which is assumed to be linearly homogeneous with respect to input x_1 (for example, energy) and x_2 (for example, labor), it is possible to define the unit-output isoquant shown in the following figure.

Figure 4
Technical and Allocative Efficiencies



The unit-output isoquant $I-I'$ is the set of input-output combinations characterized by the highest level of technical efficiency. All the water systems with $(x_1/y, x_2/y)$ above the $I-I'$ curve, are technically inefficient. For example at the point p , it uses relative high energy and labor to produce one unit of water, compare to the point m on the efficient frontier. Point p is therefore technically inefficient relative to point m . The index of technical efficiency (TE) is defined at the ratio between the distance from the origin of the unit-output isoquant and the distance from the origin of the given DMU's normalized input combination. The index of technical efficiency is defined as

$$TE = Om/Op$$

The above TE measures address by how much can input quantities be proportionally reduced without changing the output quantities produced. This is the approach in input-oriented DEA method, which we discuss in more detail in the Appendix.

2.2 Scale Economies

The previous discussion highlights the important role of scale economies in any type of production analysis. In Figures 1 through 3, we saw a production relation that exhibited initially increasing, and later decreasing returns to scale (noticeable by the initially convex and subsequently concave production function). By allowing for flexible scale economies, we can identify the efficient scale (e.g., point C in Figure 2). We could ignore scale economies, assume

constant returns to scale production, and consider unit production frontiers of the sort illustrated in Figure 4. In that case, scale economies will instead appear as a technical inefficiency—points lying off the technically efficient unit production function when production occurs somewhere other than the most scale efficient level. With this constant returns to scale assumption, one could examine the varying levels of technical efficiency over different scales of production to examine scale economies. However, this approach initially mixes scale inefficiency with other sources of technical inefficiency and requires an extra step to then examine scale economies. Given our focus in this study on economies of scale, a more sensible approach is to estimate flexible scale economies from the outset.

After using standard econometric techniques to estimate production functions assuming technical efficiency at all water systems, we use the data envelope analysis (DEA) technique to examine technical efficiency (Charnes et al., 1978; Bhattacharyya et al., 1995; Bardhan et al., 1998; Banker et al., 1984). DEA uses linear programming to trace out the most efficient production frontier, taking the data as given. We also fit our dataset using a stochastic frontier (SF) technique (Kopp et al., 1982; Kopp and Smith, 1980; Kopp and Mullahy, 1990), which assumes there are technical inefficiencies mixed with random observation errors. However, the distribution of our observations does not fit the basic assumptions of stochastic frontier model and as such, the SF technique does not generate a reliable efficiency score. Consequently, we do not report results using the SF technique. In addition, because our dataset does not have price information, we did not examine allocative efficiency, but focused solely on technical efficiency in this research. In the Appendix A, we more thoroughly discuss the DEA technique.

3. Datasets and Variables

About every five years the U.S. Environmental Protection Agency (EPA) conducts the Community Water System (CWS) Survey, a survey of community water supplies in the U.S., designed to obtain data to support the development and evaluation of drinking water regulations. In this project we utilize the 1995 and 2000 CWS Surveys (USEPA 1997b, 2002). The CWS Survey database contains detailed operating characteristics and financial information for a large random sample of community water supply systems. These questions include annual water production by water source, service population, characteristics of treatment facilities, characteristics of untreated sources, water sales revenues and deliveries by customer category,

water related revenues, water system expenses and water system assets. While much information is common to both the 1995 and 2000 surveys, each survey also has some unique elements. In 1995, for example, the survey contains more detailed information on the breakdown of water system finances, allowing us to classify inputs to production into six categories: capital, labor, materials, energy, outside services, and “other.” Unfortunately, only expenditures on each of these inputs are recorded, not the quantities and average prices, making economic analysis more difficult. This breakdown of cost information is not present in the 2000 survey, and categories of cost are defined differently, making cross-survey comparisons of water system finances somewhat problematic.⁴ The 2000 survey also contains some questions not present in the 1995 survey, including limited information on raw and finished water quality for systems serving more than 500,000.

In the CWS Survey water systems are classified primarily according to the number of people served. The full sample sizes for 1995 and 2000 are 1,980 and 1,246, respectively. However, some systems’ population served information is missing in the databases. Table 1 provides the definition of each category and gives the number of systems with complete population served information in each size category in the 1995 and 2000 surveys.

Water systems also vary in the sources of raw water. Table 2 below shows the percentage of plants in each size category using water from groundwater, surface water, and water purchased from other producers. The totals may exceed 100 percent, because some water systems have more than one source type.

⁴ For example, the past year’s capital improvements expenditure is included in the defined value of total water system expenses in the 1995 dataset. But in the 2000 dataset, capital improvements expenditures are defined as spending over the past five years and are included as a separate group outside of total expenses.

Table 1
CWS Survey:
Systems surveyed with population information available 1995 and 2000

Size category	Population served	Number of systems	
		1995	2000
Very small	25–500	574	336
Small	501–3300	511	207
Medium	3,301–10,000	270	168
Large	10,001–100,000	361	284
Very large	>100,000	121	251
Total		1,837	1,246

Table 2
Percentage of sources using surface water, groundwater, or purchased water

	Surface water	Groundwater	Purchased water
1995 survey			
Very small	30.8%	66.2%	42.0%
Small	23.9%	56.9%	37.2%
Medium	27.8%	60.7%	31.5%
Large	44.0%	61.8%	33.0%
Very large	55.4%	55.4%	43.8%
2000 survey			
Very small	38.7%	63.7%	21.4%
Small	35.7%	47.8%	26.6%
Medium	53.0%	46.4%	23.2%
Large	53.5%	45.8%	37.0%
Very large	67.3%	44.2%	37.1%

3.1 The 1995 CWSS Dataset

As noted above, the 1995 CWSS contains data on productive inputs to the output variable of water produced. These input variables include various costs such as capital, labor, material, energy, outside service, and other costs. We use depreciation expenses as a surrogate for capital costs. Labor costs include direct compensation in the form of wages, salaries, bonuses, and so on. Material costs include expenses for disinfectants, precipitant chemicals, other chemicals, materials and supplies, and purchased water (both raw and treated water). Energy costs include expenses for electricity and other energy, such as gas or oil. Outside service costs include expenses for outside analytical lab services and other outside contractor services. Other costs include other operating expenses, such as general and administrative expenses not reported elsewhere, payments in lieu of taxes, and other cash transfers out. Costs exclude tax payments, which means that the costs of public and private water supply systems are on the same footing. The output variable is total water produced in millions of gallons. We then derive total unit cost and factor unit costs using total operating expenses and individual factor costs divided by the total water produced.

The 1995 CWS Survey dataset begins with 1,980 observations. In order to properly analyze the dataset, we first had to account for missing values. We begin by eliminating 143 observations that had missing values for the population served, leaving us with 1,837 observations. We dropped 287 observations that have missing values for total expenses, bringing us to 1,559 observations. Next, we eliminated 86 systems that either have a missing value or a zero for total water production (output), leaving us with 1,473 observations. We then drop 96 observations that have a missing value for ownership class (public or private), resulting in 1,377 observations. Finally, we eliminate four systems that are not classified with a public water system (pws) id number, one system that has a missing value for the source of production, and five outlier systems with impossible values for costs per unit produced, resulting in a remainder of 1,367 systems.

For our analysis of the six individual input factors we also eliminate observations with zeroes for each of the input factors. We eliminate 553 observations with a zero value for capital costs, 37 with zero values for labor costs, 11 with zero values for materials costs, 41 with zero values for energy costs, 105 with zero values for outside costs and 55 with zero values for other costs. After these eliminations, the sample size drops to 565.

3.2 The 2000 CWSS Dataset

Community Water System Survey (CWSS) 2000 is the Fifth Edition of the CWS Survey. In the authors' opinion, the CWSS 2000 is more organized than 1995 survey and it contains more detail in certain categories. For example, detailed questions, such as treatment technologies, raw water quality and post treatment water quality were asked for very large systems and for special chemicals, such as Arsenic and MTBE. Similar to 1995 survey, the 2000 survey collects data on water system operations and finances that are critical to the preparation of regulatory, policy, implementation, and compliance analyses. Unfortunately, for the purposes of this project, the 2000 survey revealed less detailed financial information on drinking water expenses than the 1995 survey. Specifically, the detailed input cost information available in the 1995 survey is not available in the 2000 survey.

For the objective of this research, the most interesting aspect of the 2000 CWSS is the sub-sample of systems that were surveyed in both the 1995 and 2000 surveys. The number of overlapping samples is 132, which is not a huge number but adequate for a simple panel data⁵ analysis. In order to conduct a consistent comparison of the 1995 and 2000 CWSS data, we use total expenses, total finished water production, and population served, and water source information.

In the analysis that follows, we utilize three different sub-datasets of the 1995 and 2000 CWSS, depending on the model being estimated. We summarize these datasets as follows: First, for our analysis of costs by type, we use a sample from the 1995 CWSS dataset with 565 observations that have complete data for the six input factors. Second, for our analysis of total costs, we use two full datasets for 1995 and 2000 with complete data for total O&M expenses and total water produced. The sample sizes are 1,367 and 995, respectively. Finally, for our panel data analysis, we use a dataset of 132 observations of systems that existed in both the 1995 and 2000 CWSS.

⁵ Panel data means data with multiple observations crossing several years.

4. Results

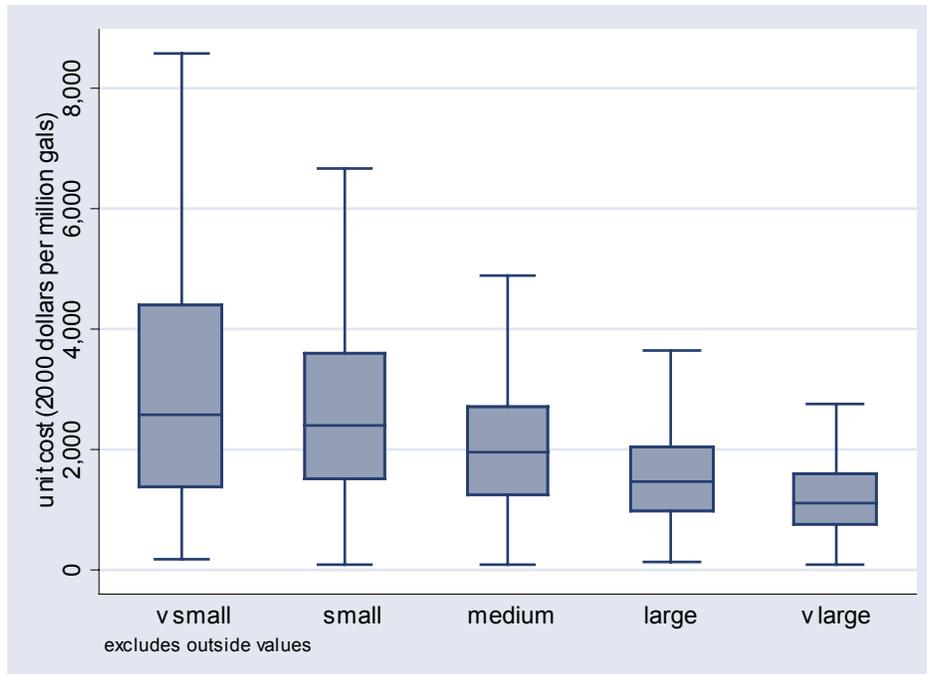
4.1 Economies of Scale in Water Supply

The box-and-whisker plot below (Figure 5) shows the range of unit production costs (operating costs plus depreciation in dollars per thousand gallons) for plants of various sizes in 1995.⁶ In the figure, the line through the boxes shows the median production costs, and the top and bottom edges of the boxes show the inter quartile range (the 75th and 25th percentiles). The lines above and below the boxes (the whiskers) show the 95th and 5th percentiles.

Figure 5 tells us several interesting things about the distribution of production costs. Most importantly, costs do generally decline as system size increases. The 1995 median cost per thousand gallons of a very small plant is 135% greater than that of a very large plant (\$2653/million gallons versus \$1128/million gallons). There is, however, very little difference between the median cost of very small and small plants. The decline in costs with increased size is also evident in the other order statistics. The 75th and 95th percentiles also show declines, as does the 25th percentile, at least for small plants and above. Finally, Figure 5 shows that despite generally falling costs, there remains plenty of overlap in costs across all size categories. For example, 20.7% of very small plants and 22.0% of medium plants have a unit cost lower than the median unit cost of very large plants.

⁶ The graph for 2000 is similar.

Figure 5
Distribution of plant production costs by size



In order to quantify the effect of size and to control for other factors available in the CWSS that affect costs, we estimate a model that includes data on water source and ownership as well as size. Water source variables are designated in the model as *GROUND*, *SURFACE*, and *PURCHASED*, respectively the percentage of raw water that comes from groundwater, surface water and water purchased from other utilities.⁷ Ownership is a dummy variable indicating whether the system is publicly or privately owned.⁸

$\ln C = \alpha + \beta_1 \text{PUBLIC} + \beta_2 \text{PRIVATE} + \gamma_1 \text{SURFACE} + \gamma_2 \text{GROUND} + \gamma_3 \text{PURCHASED} + \theta \ln W$
 where C is annual unit cost in dollars per thousand gallons and W is annual production of finished water (in million gallons). Table 3 provides estimation results for total unit cost. Because of the singularity of ownership dummy variables and sources of water variables, we manually drop variables *PRIVATE* and *GROUND* from the equation when we conducted estimation.

Table 3
Scale Economies of Total Operating Expense

Variable	
Constant	8.34* (0.061)
lnW	-0.16* (0.009)
SURFACE	0.17* (0.050)
PURCHASED	0.52* (0.053)
PUBLIC	-0.12* (0.041)
Adjusted R2	0.46
Number of observations	565
Standard errors in parentheses * Significant to 1% level	

⁷ Most utilities in the sample (1,149 out of 1,367 in 1995 and 800 out of 995 in 2000) obtain water exclusively from one type of source.

⁸ For comparability, costs exclude taxes paid by private systems.

Table 3 indicates that expected costs vary significantly based on water source. (Note that *SURFACE*, *PURCHASED*, and *PUBLIC* coefficients measure costs relative to groundwater at privately owned systems). Controlling for size, groundwater systems have the lowest cost; surface water systems are on average 17 percent more costly than groundwater systems, and use of purchased water is 52 percent more costly than groundwater. Of course, particular localities may not have unlimited access to different water sources and it is unlikely that many utilities would be able, by changing water source, to reduce costs by the amounts suggested by the coefficients in Table 3.

These results also indicate the existence of scale economies in the overall operations of these water providers. The coefficients on the size variables ($\ln W$) are actually elasticities. For example, a one percent increase in surface water production reduces unit costs by 0.16 percent. This is statistically significant, but the magnitude is fairly modest. On average, for example, a doubling of production volume reduces unit costs by 10 percent.

The estimate of scale economies above was derived by comparing the cost of systems of different sizes. These systems may differ in many ways that we do not observe, and if these differences are correlated with size, we may attribute their effects on costs to the effect of plant size. These omitted variables could cause either an over- or underestimate of scale effects. Another way of estimating scale economies is to observe the unit costs of the same water supply system at different production volumes. The advantage of this approach is that it is direct evidence of the effect of a size change, in contrast to the indirect evidence provided by cross-sectional estimation. In addition, it automatically controls for many system-specific factors that affect cost. The disadvantage is that data collection of this type is more difficult.

Table 4
Factor Elasticity (with Respect to Total Water Produced)
Estimations Using Various Models

Cost Category	Model 1	Model 2	Model 3	Model 4
Capital (1995)	-0.175	-0.176	-0.169	-0.170
Labor (1995)	-0.129	-0.140	-0.122	-0.134
Material (1995)	-0.186	-0.150	-0.188	-0.151
Energy (1995)	-0.120	-0.130	-0.111	-0.122
Outside Service (1995)	-0.219	-0.223	-0.204	-0.208
Other Costs (1995)	-0.181	-0.181	-0.166	-0.167
Total (1995)	-0.125	-0.116	-0.121	-0.112
Total (2000)	-0.118	-0.130	-0.125	-0.134
Difference 95-00	-0.475	-0.468	-0.470	-0.463
Source Controls		X		X
Ownership Controls			X	X

Fortunately, a sub-sample of 132 water supply systems were surveyed in both the 1995 and 2000 CWSS. For these systems we regress the change in reported unit costs (in constant 2000 dollars) on the change in production volume. To begin, we regress $\Delta \ln C$, the log change in unit costs in constant dollars between 1995 and 2000, against the log change in production $\Delta \ln W$, which gives results in the “Difference 95-00” row of Table 4. More generally, Table 4 shows estimates of the elasticity of unit cost with respect to water volume for a wide variety of model variants of our original equation including component costs as well as including and excluding source and ownership controls.

The largest variation in Table 4 is between the cross-section estimates, using data from a single year and the panel estimates using differenced data across the two samples from 1995 to

2000. The elasticity of unit costs with respect to water volume is much larger in this difference model. Based on the estimated elasticity of roughly 0.47, a doubling of water volume would lower unit costs by almost 30%.⁹ One possibility is that the short time horizon implies that capacity and other potentially fixed factors of production are not changing, and we are simply measuring the gains to higher capacity utilization rather than true scale economies.

In addition to looking at total costs, we can also consider the individual component costs using the more detailed financial data available in 1995. Figure 6 presents the average unit costs of production for each of these factors by the size (population served) of the system.

Figure 6
Average unit costs of production for each of the six factors of production in the 1995 dataset.

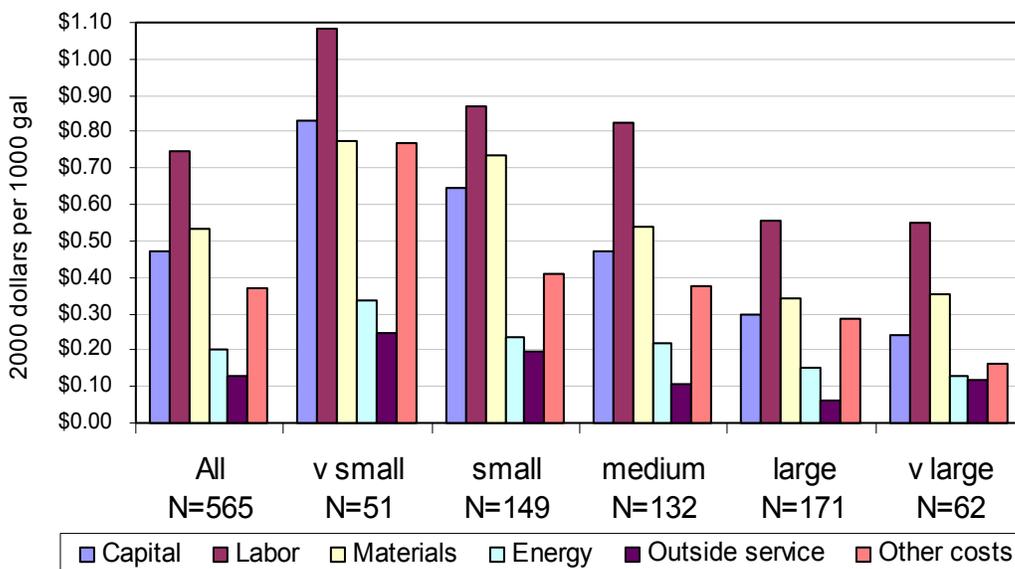


Figure 6 indicates that the average unit costs of production fall as system size increases for all of the six factors of production, but not necessarily at the same rate.

⁹ e.g., $2^{0.47} = 0.72 \sim 1 - 30\%$.

The upper portion of Table 4 quantifies these economies of scale for each component, regressing the component cost per thousand gallons on the log water supplied, with and without source and ownership controls.

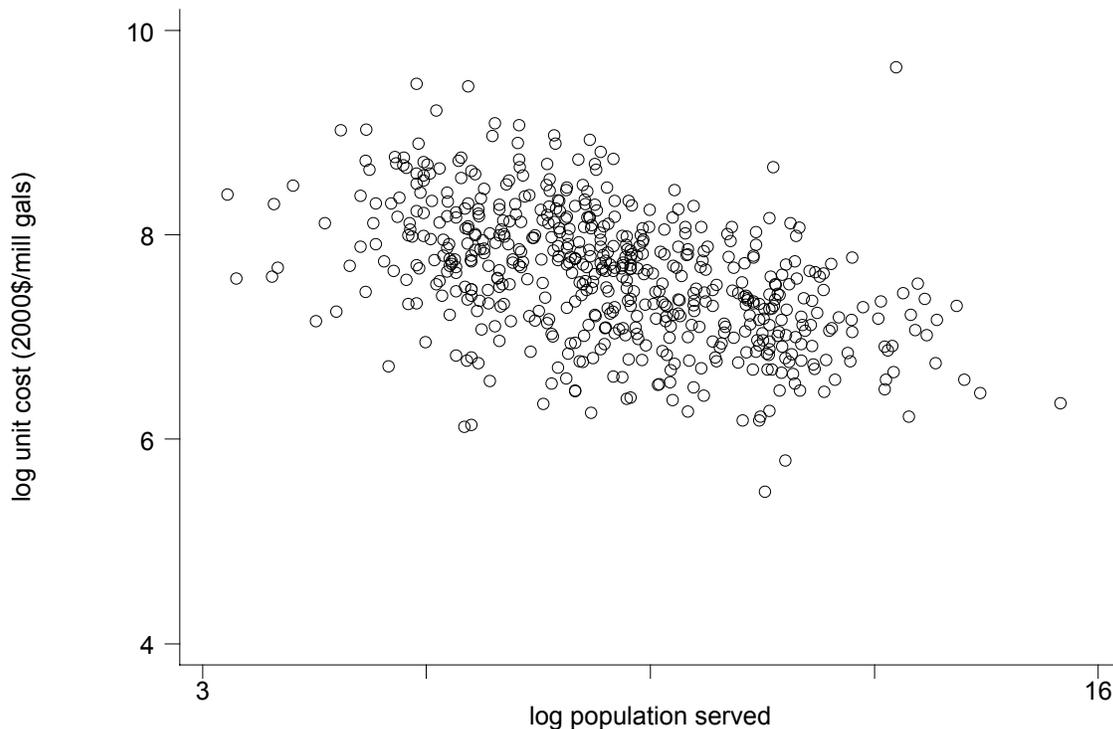
Interestingly, the greatest economies of scale exist in capital costs, outside costs, other costs and materials costs, while labor costs and energy costs exhibit the fewest economies of scale of the six factors. This suggests that larger systems may be relatively better than smaller systems at bargaining and receiving outside services and materials for a lower cost and, importantly, these gains do not necessarily depend on water systems becoming physically interconnected.

The smaller elasticity value for labor may imply that larger systems are more economically efficient from negotiating contracts and eliminating redundant positions. But, the labor economies of scale are not as substantial as those from outside costs, other costs, and materials costs. Similarly, the lower elasticity value for energy costs suggests that larger systems may have to pump water farther per unit delivered, preventing the gains from economies of scale that can be obtained from reducing other types of costs by increasing water system size.

4.2 Technical Efficiency

We have focused on production model estimates that assume all systems are technically efficient and if not, the technical inefficiency represents random observation errors. To further examine the extent of economies of scale in community-based water systems, we shift our focus from the effect of system size on economic efficiency to the effect of system size on technical efficiency. Figure 7 shows the negative relationship between the size of water systems and the unit cost of production as well as the variation represented by individual facilities—similar to Figure 5 but in logs. Previously we emphasized the downward relationship, we now focus on the variation.

Figure 7
Log of the population served versus log of the total unit cost (N=565)



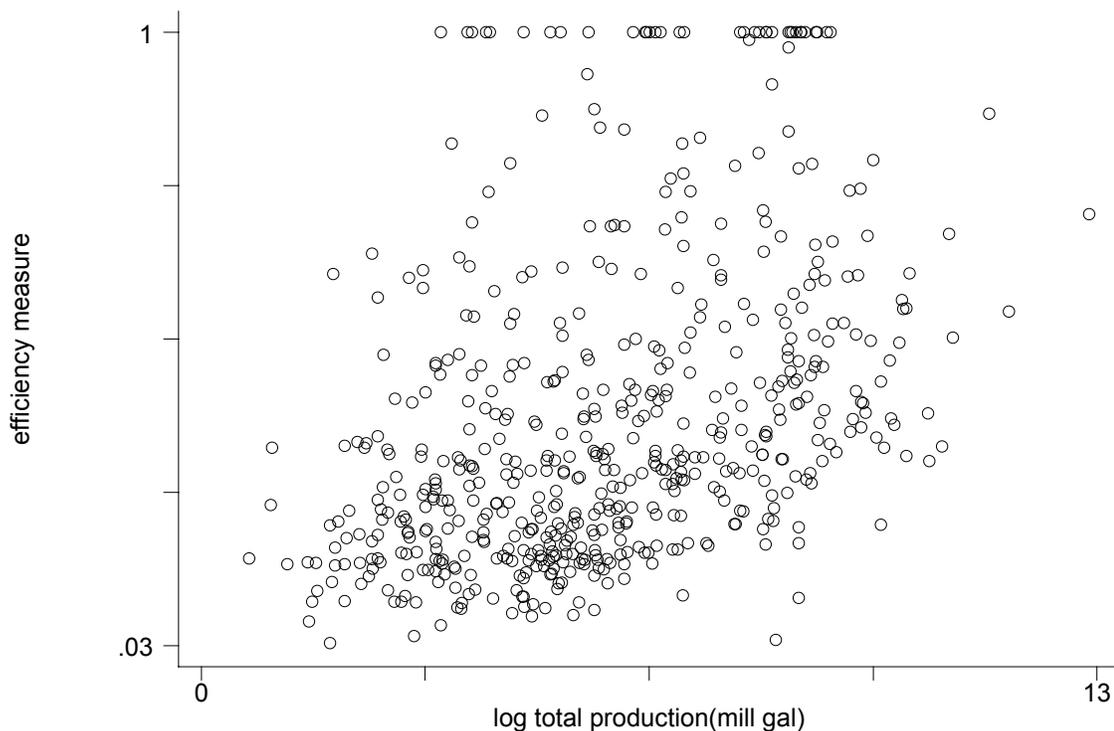
This negative relationship corresponds with the current literature, indicating that water systems do exhibit economies of scale. These economies of scale are shown by the decline in average unit costs of production with increasing water system size. Ignoring a few outliers, the unit costs are also less variable for the larger water systems. This suggests more firms operating closer to their technological frontier.

Our modeling of technical efficiency mirrors our modeling of production scale economies. After estimating the technical efficiency of each facility in one of two ways—assuming constant returns to scale—we then regress the efficiency on log water supplied and controls for source and ownership. Efficiency plays the same role as log unit costs in the production model: a 0.01 improvement in efficiency is equivalent to a 0.01 decrease in log unit costs. As with the ordinary production function estimates presented earlier, we examine the

technical efficiency of both the six input cost components measured in 1995 as well as total cost measured in 2000.¹⁰

The two ways we calculate technical efficiency involve whether we treat water production as a single input activity or whether we attempt to use the component costs as proxies for quantity inputs in a multiple input activity. The DEA model used in this research is discussed in the Appendix A. The computer package DEAP developed by Coelli (1996) was used to actually estimate technical efficiency. Figure 8 shows the scatter plot of technical efficiency by the total water produced of water systems based on the multiple input model using 595 systems from 1995 CWS dataset.

Figure 8
Plot of Technical Efficiency vs. Population Served



¹⁰ The current draft only reports results for one DEA model.

The vertical axis in Figure 8 can be thought of as similar to the vertical axis in Figure 7, except the technically efficient log unit costs have been subtracted out, leaving only the remaining inefficient “waste.” That is, assuming there is nothing preventing a firm from achieving the best-observed efficiency, failure to achieve such efficiency—values below unity on the vertical axis—represents wasted resources.¹¹ As expected, we see the flipped trend; larger systems have on average higher efficiency. We also see the (slightly) decreased variability in efficiency at higher production levels.

Table 5 presents the results of a regression of the technical efficiency from the DEA technique on log water supplied and controls for water source and ownership.

Table 5
Elasticity of Technical Efficiency

Variable	
Constant	-1.97** (0.071)
lnW	0.122** (0.011)
SURFACE	-0.055 (0.058)
PURCHASED	0.086 (0.062)
PUBLIC	0.198** (0.047)
Adjusted R2	0.23
Number of observations	565
Standard errors in parentheses ** Significant to 5% level	

¹¹ Of course, there may be good reasons why one firm is unable to match the performance of another. In that case, variation in technical efficiency reflects heterogeneity.

Reassuringly, we see an elasticity of efficiency with respect to water supplied that is similar in magnitude to the unit cost elasticities estimated earlier. This elasticity is positive—larger systems have lower unit costs but higher efficiency. It is interesting that surface and groundwater do not entail nearly the differences in efficiency, relative to purchased water, as they did with differences in unit costs. This is likely due to the fact that our measure of technical efficiency already accounts for differences in factor inputs, which would likely differ by source, soaking up some of that variation. As before, public systems have higher efficiency, e.g., lower unit costs.

4.3 Comparative Analysis of 1995 and 2000 CWSS

In this section, we discuss the elasticity of technical efficiency (TE) with respect to water produced for 1995 and 2000 CWSS datasets. Because the 2000 CWSS does not contain data on the individual input factor expenditures, to compare 1995 and 2000 technical efficiencies, we base the efficiency measurement on single input, namely, total operating expenditure (without taxes), and a single output, water produced. In order to have the same efficiency peers for the sample across two different years, we pooled the two-year samples together to measure the technical efficiency. We then separated technical efficiencies for these two-year samples to conduct elasticity of technical efficiency estimations for 1995 and 2000, separately. In addition, the elasticity of technical efficiency difference estimations is estimated separately and is based on the 132 samples existing in both 1995 and 2000 CWSS datasets. We found that the elasticities of technical efficiency are similar for different models using different controls. Much like the elasticity of unit cost, the elasticity of technical efficiency with respect to water produced is much larger in the difference model.

Table 6
DEA Technical Efficiency Elasticity
(With Respect to Total Water Produced) Estimations

Cost Category	Model 1	Model 2	Model 3	Model 4
Total (1995)	0.13	0.12	0.12	0.11
Total (2000)	0.12	0.13	0.12	0.13
Difference 95-00	0.48	0.47	0.47	0.46
Source Controls		X		X
Ownership Controls			X	X

5. Caveats

There are several caveats we would like to make related to the lack of water quality data and the DEA analysis itself. This research did not investigate the raw water quality and compliance status of individual water systems. The implicit assumption is that every system has the same raw water quality and is in compliance with drinking water standards. However, if a particular system has better raw water quality or if the system failed the drinking water standard, then given the same amount of water produced, it may use less input(s), for example, chemicals, when compared to other systems. In that case, the system will have a better technical efficiency score and gives the illusion that the system looks more efficient than other systems. In that case, the TE is biased and has overestimated the real efficiency.

In addition, the sensitivity of the DEA analysis method is important to understand. A DMU's DEA technical efficiency score is a relative number to its peers. Consequently, the score is very sensitive to the characteristics of the peers. If one of the peers on the efficiency frontier is removed, the efficiency scores of all DMUs in the analysis will be changed accordingly. The efficiency score is also sensitive to the number of inputs and outputs considered. We have found that as the number of inputs and outputs is increased, the efficiency scores increase. The rationale behind this finding is that as the number of factor inputs increases, more DMUs will become efficient, such that the relative efficiency measure of the other inefficient DMUs, their peers, increase. These observations can be seen in the following Figures 9 and 10. Figure 9 shows the scatter plot of TE using single input/single output versus six inputs/single output. Figure 10 shows the TE using six inputs/single output versus six inputs/three outputs. The six inputs, as described above, are: capital, labor, material, energy, outside service, and other costs. The single input is defined as the summation of above six input costs. The single output is total water quantity produced. The three outputs are total water quantity produced from surface, groundwater, and purchased water respectively.

Figure 9

Technical efficiency using one input and one output versus six inputs and one output

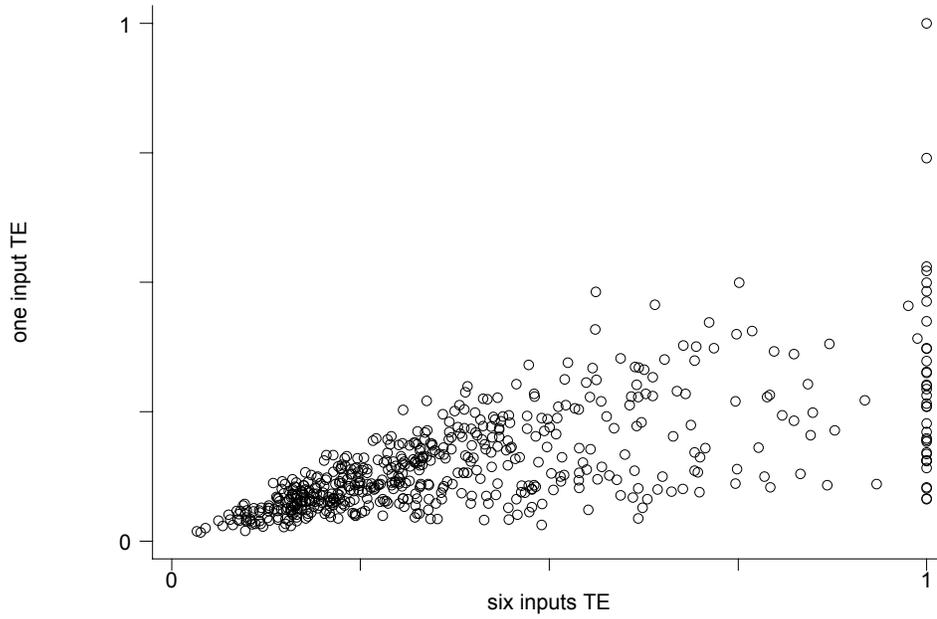
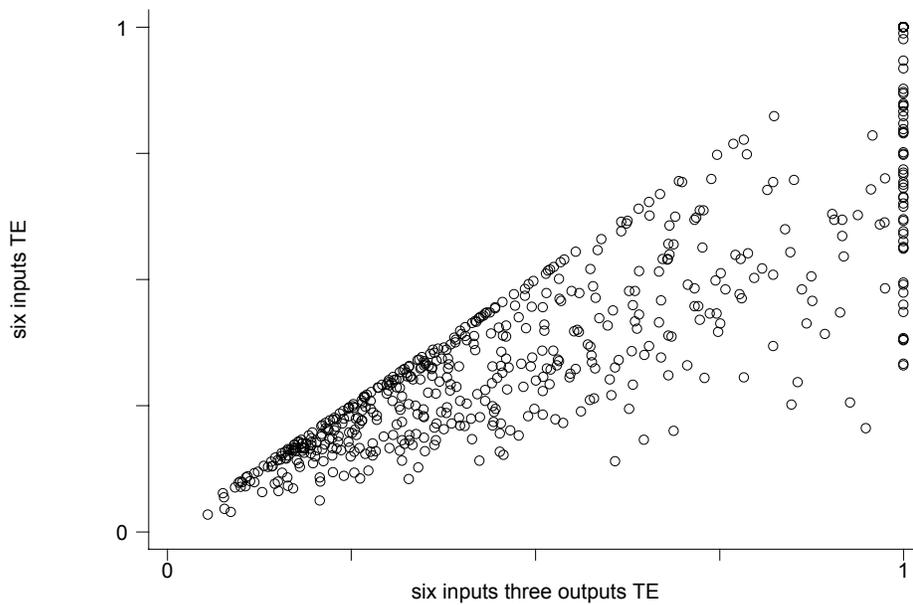


Figure 10

Technical efficiency using six inputs and three outputs versus six inputs and one output



6. Cost Savings from Consolidation

Using estimates of the elasticity of unit costs with respect to system output, we can make calculations concerning the potential cost savings from consolidating all small water systems throughout the United States. From our original model, we have

$$\Delta \ln C = \theta \Delta \ln W$$

for each firm where as before C are unit costs (dollars per thousand gallons) and W is the volume of water supplied. If $\Delta \ln C$ is the change in log unit costs, $WC \exp(\Delta \ln C)$ is the new total cost. We therefore compute the total cost savings (relative to the observed costs) in the sample as

$$\sum_i W_i \cdot C_i \cdot (1 - \exp(\theta \cdot \Delta \ln W_i))$$

That is, the old costs minus the new costs. Given θ , the only question is what to assume about $\Delta \ln W_i$ in our experiment consolidating water systems.

As a starting point, we compute the cost savings associated with giving systems below the median supplied water level the scale economy gains associated with moving their output to the median level (304 million gallons per year). That is, we do not assume their output actually rises, but that somehow they are able to realize those gains.

Carrying out these calculations, we find that among the 565 facilities in 1995 where detailed cost data is available, such an exercise would realize a \$9 million cost savings per year. On total costs of \$2.2 billion, this is less than one-half of one percent. However, among the water systems below the median size, with total costs of only \$67 million, this is a cost savings of 14 percent. For more than 40 systems, there would be cost savings of more than 50 percent.

Following Table 7 shows total as well as individual components cost saving as a percentage of total expenditure of water systems below the median size.

Table 7
Total and Individual Components Cost Saving
Among the Water Systems Below the Median Size (%)

Total Cost	13.5
Capital	2.7
Labor	3.2
Material	3.3
Energy	0.8
Outside	0.9
Others	2.1

We also simulated the annual cost saving for the following three scenarios: (a) combine small water systems with population less than 500 with metropolitan water systems with populations greater than 50,000; (b) combined small water systems with populations less than 500 with median water systems with populations served between 3,300 and 10,000; (c) double the size for the systems with populations less than 500. For scenarios (a) and (b), we use the median of total water produced in each population group, namely, 6,506 and 242 million gallons for system population greater than 50,000 and between 3,300 and 10,000 respectively, as the targets for scaling up the small systems. For scenario (c), we use two times the current total water produced as the targets for scaling up. The total and individual components of cost savings for water systems that serve fewer than 500 people are given (using our 565 sample data) for above three scenarios below in Table 8.

Table 8

Total and Individual Components Cost Saving (as Percentage of Total Costs) for the Water Systems with Population Served Below 500 in our 565 Sample Dataset

	Scenario A	Scenario B	Scenario C
Total Cost	56.5	26.3	10.5
Capital	12.6	4.3	2.7
Labor	13.7	5.9	2.5
Material	11.6	5.8	2.2
Energy	3.5	1.9	0.5
Outside	4.2	2.5	0.8
Others	10.9	6.2	2.0

The total annual cost savings for small systems with population less than 500 in our 565 samples are \$1,500,000, \$700,000, and \$280,000 for the above three scenarios, respectively. The general findings are as follows: First, the major cost savings are from capital, labor, material, and other costs. Although outside service cost has relatively large elasticity, because the cost share is small, the cost saving from outside service turns out to be relatively small. Also, cost savings from energy are small relative to other factors, indicating that on average, an increase in system size saves less money from reduced energy costs than it saves from any of the other types of expenses. One potential explanation for this is that water-pumping costs increase as system size increases. On the other hand, the three costs—material, outside, and other costs—do not require interconnection to obtain saving. If we scale the cost saving up to national level using survey sample weights, the annual cost saving are \$1,115M (all facilities to median size), \$794M (merge system population served less than 500 with population served greater than 50,000), \$417M (merge system population served less than 500 with population served between 3,300 and 10,000) and \$140M (double the size with population less than 500) for the four cases

discussed in this section.¹² That is, savings could be on the order of \$1 billion. Among factors that likely do not require interconnection (e.g. material, outside, and other costs), the savings are \$500M, \$380M, \$218M, and \$70M, respectively.

7. Conclusions

In this study we use the 1995 and 2000 CWSS datasets to examine the production costs of water supply systems. We find that smaller systems tend to face higher unit production costs across the full range of production inputs. While physical interconnection may be difficult, consolidation of administrative tasks and bargaining power may allow water systems to realize some of the potential economies of scale. However, the variation in unit costs across and within system sizes is large and size explains only a part of the cost distribution. Viewed another way, we find a high degree of technical inefficiency—large differences in costs among firms with the same input structure.

A variety of estimation techniques and cost measures provide broadly consistent estimates. We find that ground and surface water are, not surprisingly, cheaper than purchased water. Public systems consistently have lower costs than private ones. Certain cost components also exhibit economies of scale to a greater or lesser extent than others.

Quantifying the economies of scale, we find that doubling a system's production would lower unit costs between 10 and 30 percent, depending on the model and the cost component. However, consolidating small systems into a large system could double the small system's scale several times over, providing gains of 50% or more. Our estimates suggest that by giving all systems *below* the median system size the efficiency gain associated with *moving to* the median system size, annual savings of \$9 million would be realized in our sample of 565 facilities or, extrapolated to the entire to the population of roughly 50,000 systems, more than \$1 billion. We find similar, though slightly smaller estimates, when we consider merging small systems with larger systems.

¹² To scale our calculations to the national level, we multiply the savings at each facility by the sampling weights used to scale the original sample up to national values. The sampling weights are inflated to reflect our reduced (565) sample size in such a way as to maintain the original distribution of facility size (e.g., across very small, small, medium, large, and very large).

The majority of cost savings are from capital, labor, material, and other costs. While increased efficiencies associated with capital and some labor input (e.g., production versus administration and management) would require interconnection, other efficiency gains would not. To the extent these gains arise from increased bargaining power, they are all achievable without interconnection.

Compared to overall supply costs, these numbers might be considered small—about 4.5 percent. However, to the small and very small systems realizing the gains, they are large. Of course, we cannot claim that all these benefits could actually be realized. The fact that so many small water supply systems still exist suggest that there may be costs or barriers to consolidation that are not readily apparent in the CWS Survey data. One obvious example is the cost of interconnection when trying to physically bring rural populations into a large system, though some gains may not require this. In any case, as small water systems confront increasing regulatory challenges, consolidation may offer some hope of reducing that regulatory burden. To say more than this, additional research is needed on the local causes of cost heterogeneity in water supply systems and the localized costs and benefits of consolidation.

References

- AWWA. 1997. Feasibility of Small System Restructuring to Facilitate SDWA Compliance. Denver, Colorado: AWWA Research Foundation.
- Banker, R.D., A. Charnes, and W.W. Cooper. 1984. Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science* 30: 1078–92.
- Bardhan, I.R., W.W. Cooper, and S.C. Kumbhakar, S.C. 1998. A Simulation Study of Joint Uses of Data Envelopment Analysis and Statistical Regressions for Production Function Estimation and Efficiency Evaluation. *Journal of Productivity Analysis* 9: 249–78.
- Beecher Policy Research, Inc., and Cadmus Group, Inc. 2002. Scale and Scope Economies for Water Systems: Illustrations. Draft Report. Watertown, Massachusetts: Cadmus Group
- Bhattacharyya, A., T.R. Harris, R. Narayanan, and K. Raffiee. 1995. Technical Efficiency of Rural Water Utilities. *Journal of Agricultural and Resource Economics* 20(2): 373–91.
- Cadmus Group. 2002. Small Drinking Water System Consolidation Selected State Program and Consolidation Case Studies. Report. Watertown, Massachusetts: Cadmus Group
- Coelli, T.J. 1996. A Guide to DEAP Version 2.1: A Data Envelopment Analysis (Computer) Program. Center for Efficiency and Productivity Analysis (CEPA) Working Papers. Armidale, Australia: The University of New England.
- Coelli, T.J., D.S.P. Rao, and G.E. Battese. 1998. An Introduction to Efficiency and Productivity Analysis. Boston, Massachusetts: Kluwer Academic Publishers.
- Charnes, A., W.W. Cooper, and E. Rhodes. 1978. Measuring the Efficiency of Decisionmaking Units. *European Journal of Operational Research* 2: 429–44.
- Farrell, M.J. 1957. The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society Series a*, 120 (III): 253–81.
- Hewitt, J.A. 2000. An Investigation into the Reasons Why Water Utilities Choose Particular Residential Rate Structures. In A. Dinar, ed., *The Political Economy of Water Pricing Reforms*. Oxford, UK: Oxford University Press.
- Kopp, R.J., and J. Mullahy. 1990. Moment-Based Estimation and Testing of Stochastic Frontier Models. *Journal of Econometrics* 46: 165–83.

Kopp, R.J., and V.K. Smith. 1980. Frontier Production Function Estimates for Steam Electric Generation: A Comparative Analysis. *Southern Economic Journal* 47(4): 1049–59.

Kopp, R.J., V.K. Smith, and W.J. Vaughan, W.J. 1982. Stochastic Cost Frontiers and Perceived Technical Inefficiency. *Advances in Applied Micro-economics*, 2: 71–85.

Renzetti, S. 2001. Environmental Conditions and the Cost of Water Supply. AWRA/IWLRI-University of Dundee International Specialty Conference, August 6–8.

U.S. EPA. 1997a. <http://www.epa.gov/safewater/annual/acrful.pdf>.

———. 1997b. <http://www.epa.gov/safewater/cwsreprt.pdf>.

———. 2003 accessed.

<http://www.epa.gov/safewater/dwa/electronic/presentations/sdwa/pt4/sdwa56.html>.

Appendix A: Data Envelopment Analysis (DEA) Model

DEA is a linear programming (LP) based method for assessing the productive efficiencies of decisionmaking units such as bank branches, schools and hospitals which perform a given function. Each unit represents an observed correspondence of input-output levels. DEA uses a LP approach to ascertain the potential for input reduction at a unit given its output levels, or alternatively the potential for output augmentation given its input levels. DEA models were first developed by Charnes, Cooper, and Rhodes (1978).

The input-oriented DEA model for a single output is formalized below. Consider the situation with n decisionmaking units (DMUs), each producing single output by using m different inputs. The i -th DMU uses x_{ki} units of the k -th input in the production of y_i units of output. A separate LP problem is solved for each DMU. The constant returns to scale (CRS) input-oriented DEA model for the i -th DMU is given by

$$\begin{aligned} & \text{Min } \theta \\ & \text{s.t.:} \\ & \theta x_0 - \lambda X - s = 0 \\ & y_0 - \lambda Y + e = 0 \\ & \lambda \geq 0, s \geq 0, e \geq 0. \end{aligned}$$

Where θ is the proportional reduction in input possible for the i -th DMU; s is the input slack, e is the k -th output slack; and λ_j is the weight of j -th DMU.

The single output input-oriented DEA model seeks to minimize the proportional reduction in input while remaining within the production possibility set. The proportional reduction in input which is possible is accomplished when input slack, s , becomes zero. When the value of θ is 1, $\lambda_i=1$, and $\lambda_j=0$ for $j \neq i$, the i -th DMU lies on the frontier and is efficient. For the inefficient units, $\theta < 1$, $\lambda_i=0$, and $\lambda_j \neq 0$ for $j \neq i$. The projected or frontier production level for the i -th DMU, denoted by \bar{x}_i , is given by

$$\bar{x}_i = \theta x_i$$

In this research, we solve DEA model using DEAP by Coelli (1996).

DEA uses multiple inputs and outputs from every decisionmaking unit to estimate the relative efficiency from its peers. It also identifies inefficiencies in each input and output which

are attributable to organization form used and provides some guidance for each DMU about how to improve its performance.

The strengths of DEA include:

- DEA can handle multiple inputs and multiple outputs.
- DEA does not require an assumption of a functional form relating inputs to outputs.
- DMUs are directly compared against a peer or combination of peers.
- Inputs and outputs can have very different units.

DEA also has some limitations. An analyst should keep following limitations in mind when choosing whether or not to use DEA.

- Since DEA is an extreme point technique, random variation unrelated to the explanatory variable (e.g., measurement error) can cause significant problems.
- DEA is good at estimating “relative” efficiency of a DMU. In other words, it can tell you how well you are doing compared to your peers but not compared to a “theoretical maximum.”
- Since DEA is a nonparametric technique, statistical hypothesis tests are difficult.

Appendix B: Some Recommendations for CWSS

The authors for this project recognize that the CWSS is a very important data collection effort undertaken by the Environmental Protection Agency's (EPA) Office of Ground Water and Drinking Water. The CWSS data is collected every five years and has been used extensively in policy research. While performing the economic analysis using 1995 and 2000 CWSS datasets for this specific project, the authors discovered gaps in the data that impacted this project. As a gesture, the authors feel it would be useful to provide suggestions to assist future data collection efforts for a possible expansion of future CWS Surveys to make them more useful for economic analysis.

Functional aggregation of costs. Costs are given in the CWSS for a wide variety of specific items (for example, energy, labor), but there is little to indicate how these inputs are used. Since one of the major questions in this project's analysis is associated with economies of scale (e.g., whether to join systems administratively without joining them physically), it would be helpful to have the costs grouped according to their function, such as raw water transport, water treatment, distribution, and administration.

Prices. The 1995 CWSS includes input factor expenditures, which is the product of prices and quantities of the input factors. Strictly speaking, the technical efficiency analysis discussed in this research should be carried out with quantities of input factors, rather than expenditures. For 2000 CWSS, the expenditure information is even more highly aggregated. It would be useful to break down the expenditure by price and quantity for each input factor.

Panel dataset. To permit a more meaningful econometric analysis, the authors suggest taking a representative sample of water systems and then focusing on using that same sample in every CWS Survey to the degree that it is possible. This could provide a very useful panel dataset (i.e., it would have both a cross-sectional and time-series component).

Consistent variable definition across years. Survey questions were asked in different ways between 1995 and 2000. For example, in the '95 dataset the past year's capital improvement expenditure is included in the defined value of total water system expenses. However, in the 2000 dataset, capital improvement expenditures are defined as spending over the past five years and are included as a separate group outside of total expenses. Consistent variable definition would allow for more interesting time-series analysis.

Raw water quality data. Both the 1995 and 2000 datasets did not include data on the raw water quality of all source waters. The authors believe the quality of raw water is an important missing variable in our analysis because quality variation largely determines the variation in water treatment, which affects the variation in costs (Renzetti, 2001). The authors recommend that future CWS Surveys include a raw water quality section eliciting the quality of all sources of water.

Compliance data. Collection of compliance data under the Safe Drinking Water Act would be useful. Noncompliant plants may have different costs from compliant plants.