Does Capacity Utilization Affect the "Stickiness" of Cost?

RAMJI BALAKRISHNAN*
MICHAEL J. PETERSEN**
NAOMI S. SODERSTROM***

1. Introduction

Traditional cost models are constructed using the assumption (Noreen [1991]) that costs change proportionately with activity levels. The implicit assumption is that the proportionality of the change in cost is independent of the magnitude and direction of change in activity levels. Considerable attention has been devoted to examining the explicit proportionality assumption (e.g., Noreen and Soderstrom [1994, 1997]; Balakrishnan and Soderstrom [2000]). Recent research (Anderson, Banker, and Janakiraman [2003]) has begun to examine the implicit assumption of whether the direction of change in activity moderates the cost response. A differential response is expected because, as Cooper and Kaplan (1998) observe, managers seem more inclined to increase costs when activity levels increase than they are to decrease costs when activity levels decrease. Anderson et al. (2003) coin the term "sticky" cost to capture this asymmetric cost response. Their analysis of selling, general and administrative costs provides broad support for the conjectured behavior.[1]

This paper extends the analysis in Anderson et al. (2003) to capture the effect of two factors that may also moderate the manager's response to changing activity levels. First, the *magnitude* of the change may influence the proportionality of response. Significant transactions costs associated with changing cost levels may rationally lead to the response to a "large" change in activity being proportionately larger than the response to a "small" change in activity.[2] The response to small

*The University of Iowa
**Arizona State University
***University of Colorado at Boulder

Data used in this study come from a proprietary source and cannot be shared.

We thank Shannon Anderson, Leslie Eldenburg, John Jacob, Scott Keating, David Larcker, Joan Luft, Morton Pincus, Siva Nathan, George Neuman, Sonja Rego, Frank Selto, Geoff Sprinkle, Konduru Sivaramakrishnan, Sally Widener, participants at presentations at the University of Iowa, University of Arizona the 2000 MAR conference, and the anonymous reviewer for comments. This paper was previously titled "On the Behavior of Labor Cost in Therapy Clinics."

1. Other research examining the stickiness of costs include Subramaniam and Weidenmier (2002) and Wiersma (2002).

2. We define a large change as any change greater than 3 percent. Approximately 75 percent of all changes were less than ± 7 percent, reducing concerns that observations lie outside of the relevant range.

changes may better manifest itself in the intensity of asset usage, without increasing their supply. Second, current *capacity utilization* may affect the manager's response to a change in activity levels. If the firm is experiencing excess capacity, managers may be more likely to use slack to absorb the demand from an increase in activity levels. However, an additional decrease in activity levels may be seen as confirmation of a permanent reduction in demand and thus trigger a greater response. Thus, if the firm has excess capacity, the response to a decrease would exceed the response to a similar increase in activity levels. On the other hand, if the firm has high capacity utilization (strained capacity), the response to a decrease would be lower than the response to a similar increase in activity levels. If the firm is experiencing strained capacity, managers may use a decrease in demand to relieve pressure on available resources and not reduce resource levels proportionately. Conversely, in this setting, any further increase may cross resource thresholds and trigger a disproportionate increase in resources supplied.

We test the above conjectures using data from a firm that operates a number of physical therapy clinics in the western United States. We utilize data for 1,898 clinic-months from 49 therapy clinics. We use both the number of therapist hours staffed and the salary paid to therapists as dependent variables in testing the hypothesized relations. We use both measures because scheduling flexibility allows clinic managers to more easily adjust staffing hours relative to adjusting staffing cost, by "banking" hours for example. We focus on therapists because this is the key and binding resource in operating clinics, and salaries and benefits are over 80 percent of total cost.

Our evidence is mixed. For the sample as a whole, we do not find significantly different responses to large versus small changes in activity levels. The percentage change in hours or cost for "small" changes ($\pm$ 3%) in the number of visits (the activity measure) is not distinguishable from the change in hours or cost when the number of visits changes by more than 3 percent. Turning to our second conjecture, we find evidence of sticky cost when we consider the sample as a whole. As in Anderson et al. (2003), the response to a decrease in activity levels (3% or greater) is smaller than the response to a similar increase in activity levels. However, more central to our analysis, we find a significant interaction with current capacity utilization. If the clinic is experiencing excess capacity, the response to a large decrease in activity levels is greater than the response to a large increase in activity levels. On the other hand, if the clinic's resources are currently strained, the response to a large decrease in activity levels is smaller than the response to an increase. There is no statistically detectable differential in response when the clinic is operating at "normal" utilization levels. The above results hold for staffed hours and for cost, although the results for staffed hours are qualitatively stronger. Inferences are robust to alternate definitions of a large change and the cutoff used to classify a clinic as currently experiencing excess capacity or strained capacity.

These results underscore the need to examine the fundamental assumptions of cost models used in describing cost behavior. Our findings suggest caution in applying Anderson et al.'s conclusion that costs are sticky. While costs are indeed sticky on average as Anderson et al. document, current capacity utilization plays an important role in determining the extent of stickiness.

The remainder of this paper is organized as follows. Section 2 describes the setting and model that we estimate. Section 3 provides descriptive statistics and discusses operationalization of key constructs. Section 4 reports results from the model estimation. Section 5 concludes.

2. Setting and Model Development

Our data are drawn from a single firm that provides occupational, physical, and speech therapy services in the western United States. Our study focuses on physical therapy because of variations in the labor markets for therapists across the three categories. The clinics are located within either a skilled nursing facility (SNF) or a hospital (HOSP). For hospital-based clinics, the hospital provides space and some equipment but is otherwise not in charge of the clinic. Each clinic is evaluated as a profit center, loosely controlled by a central office. Clinic managers (who often double as therapists) are responsible for staffing of therapists and scheduling of aides and assistants. The firm uses flexible budgets for planning and control, with the number of visits as the activity measure. Labor cost is the largest expense. In 1997, for the firm as a whole, the cost of labor and accompanying benefits accounted for 85.1 percent of total expenses, implying that labor is the focus of cost management efforts. Managers are motivated to control costs via a profit sharing system that rewards them with a percentage of net income for their clinic. In 1997, the payout from profit sharing was about 12 percent of the managers' base pay.[3]

Clinics have little control over the volume of activity. Visits occur due to patient referrals from a pool of physicians or the hospital. All clinic visits are on an outpatient basis. While there is variation across patients in the number of visits required, average effort required per visit is not believed to be highly variable.[4] The current backlog and the extent to which the schedule was full did, however, affect variations in the time given to each patient. That is, the pace of work could be managed *within limits*, while still providing appropriate care. In Section 3, we report evidence of growth in visits over this period and the persistence of demand shocks.

The staffing decision involves hiring a mix of staff therapists and therapy aides. A therapy aide can supplement the work of a therapist by performing some remedial exercises and by monitoring therapy steps. However, an aide cannot replace a

3. We lack data that would allow us to capture any cross-sectional variation in the incentive plans. Conversations with firm personnel indicate some variation. The presence or the strength of an incentive plan appeared unrelated to the type of clinic or its location. Managerial tenure was offered as a potential explanatory variable.

4. The firm does not weight visits by the complexity of the required therapy. Firm personnel believe that complexity in treatment typically manifests itself as an increase in the number of visits rather than as an increase in the intensity of each visit.

therapist; state laws dictate the maximum allowable proportion of aides to therapists and the work that can be done by an aide. A licensed therapist must do the initial diagnosis and perform progress checks. Therapists usually are salaried employees. Aides are paid on an hourly basis.[5] Thus, therapist salary has the characteristics of a "fixed" cost that is hard to manage. Clinic managers, however, exert considerable influence on the total payroll by creative use of schedules. As discussed below, there are several ways that managers can manage consumption of the therapists' time. Consequently, in this study, we focus the analysis on the number of therapist hours staffed. For robustness, we estimated the model using total hours and reached identical conclusions.[6]

Clinic managers have many avenues to manage the number of staffed therapist hours (the dependent variable). The most important is scheduling. Much like a dentist's office, clinics have little walk-in demand; visits are usually scheduled a week or more in advance.[7] Managers closely watch the extent to which the scheduling book is full and adjust staffing accordingly. The primary manner by which managers adjust staffing is by adjusting the schedule of therapists who have a part-time appointment (e.g., two or three days a week, or four hours a day) and by managing when therapists take vacation or other time off (comp time). The magnitude of such adjustments, of course, depends on the mix of full- and part-time therapists. We could not find any objective measure of the "degrees of freedom" enjoyed by individual clinics; it appears that there is no systematic determinant of the mix. Second, conversations with firm personnel indicate that overtime, while used when needed, is not extensively employed. Third, managers can also manage time available by shifting therapists across clinics. The feasibility of such shifting depends on the clinic's location as well as demand from the other clinics. While the practice is not unusual (we were not able to get specific measures), firm personnel feel that it is inappropriate to characterize the setting as one in which the clinics drew from a common pool of therapists. Finally, many managers keep a "registry" of licensed therapists who, for various reasons, do not wish to work on a regular basis and/or want supplementary income. When available, a manager could use this registry to meet temporary demand surges. Overall, firm management feel that a key aspect of the clinic managers' job is to manage staffing in such a way that there is not too much excess capacity and a patient can be given a new appointment without an undue wait for an open slot.

---

5. We focus our tests on staff hours because roughly 7 percent of the clinics never employed any aides. In addition, use of total hours raises questions about the mix of new and old patients since therapists must do the initial assessment. Our conclusions are unaltered if we use total hours or total cost as the dependent variable. It is also possible to argue that the availability of aide hours can be viewed as a parameter that influences the proportionality of response, suggesting that the parameter be incorporated as an independent variable. Our conclusions are not altered if we add the "average mix of aide hours to staff hours" as an independent variable. We thank the associate editor for this insight.

6. These time measures exclude time that was paid for but could not be used to schedule patients. Specifically, vacation, sick leave, and comp time (essentially, a way to "bank" time) are not included in our measure of the dependent variable.

7. While there was an effort to maintain continuity of care, there was no medical need for a patient to see the same therapist during each visit.

2.1 Model Development

Let $VISITS_{it}$ represent the number of visits and let HRS_{it} represent the number of therapist hours staffed in clinic i during period t. Then, we posit that

$$HRS_{it} = FC_i + \beta * VISITS_{it},$$

where FC_i is the fixed time commitment required to administer clinic i. (We also develop and estimate an analogous model with $COST_{it}$ as the dependent variable, where $COST_{it}$ is the total salary expense for therapists. For parsimony, we do not replicate the model development.) Estimating this equation (with controls for clinic type and location; results not tabled) reveals that variation in the number of visits explains over 80 percent of the variance in staffing hours, implying that visits are the main, if not the only, determinant of staffed therapist hours in the long run. Estimating the model with $COST_{it}$ (also not tabled) as the dependent variable also shows an excellent fit. Such a fit is expected because the level of staffed hours and cost are highly correlated.

Taking first differences and scaling by prior levels, we have

$$\frac{HRS_{it} - HRS_{it-1}}{HRS_{it-1}} = \beta \frac{VISITS_{it} - VISITS_{it-1}}{VISITS_{it-1}}, \tag{1}$$

where β is the percentage change in hours staffed (hereafter, %ΔHRS) for a 1 percent change in the number of visits (hereafter, %ΔVISITS) (i.e., the elasticity of the response). We estimate eq. (1) as an ordinary least squares model:

$$\%\Delta HRS_{it} = \beta_1 \%\Delta VISITS_{it} + \varepsilon_{it}. \tag{2}$$

Note that if the number of hours staffed varies proportionately with the number of visits, $\beta_1 = 1$. Empirically, we expect $0 < \beta_1 < 1$.

2.2 Direction and Magnitude of Change in Activity

Considerable anecdotal evidence (see also Cooper and Kaplan [1998, p. 247]) suggests that the model in eq. (3) may be misspecified because the manager's reaction is likely to be affected by the direction of the change in the activity level. The change in reaction to an increase in visits (%ΔVISITS > 0) is likely greater than that for a decrease in visits (%ΔVISITS < 0) because managers commit to capacity resources such as labor before demand is known. If demand exceeds expectation, available capacity is strained. Slack is created if demand has been overestimated. Slack does not put the same pressure on managers as not having enough staff. While sustained slack affects the probability of meeting cost targets, the pressures are more in the medium and long term. In our setting, not having enough staff (i.e., not being able to schedule patients on a timely basis) is more of an immediate problem than being overstaffed due to pressures from patients, physicians, and employees. In addition, the spending-consumption model (Cooper and

Kaplan [1992]) suggests that decreases in committed cost are a function of managerial decisions. Clinic managers may be loath to fire fellow staff or to reduce their hours. Reducing the pace of work may be a more attractive short-term solution to a decline in demand. This indicates that the manager's loss function may be asymmetric to under- and overstaffing, which would manifest itself as "stickiness" in the response function that maps changes in the number of patient visits to staffed hours (or cost).

Operationally, we capture the effect of the direction and magnitude in activity change by partitioning the change in visits into three groups: one group for decreases in activity of at least 3 percent (NEG), one group for changes in activity less than 3 percent (SMALL), and one group for increases in activity of at least 3 percent (POS). Dummy variables for each group enter the regression both as intercept and slope effects (where each dummy variable is multiplied by %ΔVISITS). Introduction of the dummy variables changes the equation to

$$\begin{aligned} \%\Delta HRS_{it} = {} & \beta_1 NEG_{it} + \beta_2 SMALL_{it} + \beta_3 POS_{it} + \beta_4 NEG_{it} * \%\Delta VISITS_{it} \\ & + \beta_5 SMALL_{it} * \%\Delta VISITS_{it} + \beta_6 POS_{it} * \%\Delta VISITS_{it} \\ & + \text{controls} + \varepsilon_{it}, \end{aligned} \tag{3}$$

where

%ΔHRS	=	Percent change in hours staffed.
NEG	=	1 if activity decreased at least 3 percent, 0 otherwise.
SMALL	=	1 if activity changed less than 3 percent, 0 otherwise.
POS	=	1 if activity increased at least 3 percent, 0 otherwise.
%ΔVISITS	=	percent change in patient visits.
controls	=	control variables (specified below).

At the aggregate level, the predicted effect due to direction (sticky cost) implies that $\beta_4 < \beta_6$. This is the prediction in Anderson et al. (2003). Our argument of fixed transaction costs implies a difference due to the magnitude of the change, or $\beta_4 > \beta_5$ and $\beta_6 > \beta_5$.

2.3 Current Capacity Utilization

Our main argument is that capacity utilization could be crucial in determining managerial response to activity changes. Consider a clinic that is at capacity. If there is a decrease in the activity level, the manager is likely to take advantage of the decrease and relieve some of the strain on the staff (or ease scheduling problems) by not reducing capacity proportionately. In contrast, if there is an increase in activity levels, the manager is likely to increase staffing to relieve the overburdened staff.

Opposite incentives seem to exist when we consider a clinic experiencing excess capacity. If activity levels decrease, managers are likely to view the reduction as confirmation of a permanent reduction in activity levels. If, however, activity levels increase, the manager is likely to employ some of the existing slack rather than to add more capacity. Thus, for a given negative change, we expect a greater reduction in staffed hours if the clinic currently has excess capacity than if the clinic were at capacity. For the same positive change in activity, we expect a clinic that has excess capacity to add fewer hours relative to a clinic that is currently at capacity.

We investigate the above interaction effect by defining three levels of capacity utilization: Normal Utilization (NORMAL), excess capacity (EXCESS), and strained resources (STRAINED). We then partition the sample into different levels of capacity utilization and jointly estimate the model, eq. (3), for each partition as a system of equations, using seemingly unrelated regression (SUR). We label the β coefficients to reflect the partition: β_{Ni} for normal capacity utilization, β_{Ei} for excess capacity, and β_{Si} for strained resources. Equation (3) is thus modified to reflect different categories of resource utilization:

$$\begin{aligned}\%\Delta \text{HRS}_{it} = {} & \beta_{x1}\text{NEG}_{it} + \beta_{x2}\text{SMALL}_{it} + \beta_{x3}\text{POS}_{it} \\ & + \beta_{x4}\text{NEG}_{it} * \%\Delta\text{VISITS}_{it} + \beta_{x5}\text{SMALL}_{it} * \%\Delta\text{VISITS}_{it} \\ & + \beta_{x6}\text{POS}_{it} * \%\Delta\text{VISITS}_{it} + \text{controls} + \varepsilon_{it},\end{aligned} \tag{3a}$$

where $X = N$ for normal utilization, E for excess capacity, or S for strained resources, and all variables are as defined in eq. (3).

Absent an interaction with capacity utilization, Cooper and Kaplan's conjecture (as tested by Anderson et al. [2003]) is that a differential response to positive versus negative changes in activity would result in $\beta_{x4} < \beta_{x6}$ for each of the system's equations ($X = N, E, S$). Interaction with capacity utilization should lead to conditional predictions. Specifically, we conjecture that $\beta_{S4} < \beta_{S6}$ and that $\beta_{E4} > \beta_{E6}$. The interaction effect clouds the prediction for the relation when capacity utilization is normal, so we do not make a directional prediction for $X = N$.

The above predictions pertain to a differential response, controlling for capacity utilization (within the equation). Based on the interaction effect, we also conjecture that $\beta_{S4} < \beta_{E4}$; we expect a greater reduction of the managerial response for decreases in activity levels when capacity is at a premium than when it is not. We also expect that if the activity level increases, managers will make larger staffing adjustments when capacity is strained than when there is excess capacity ($\beta_{S6} > \beta_{E6}$). We do not make directional predictions across equations, when capacity utilization is normal.

For each capacity level, we augment the system of equations to control for other factors that systematically differ across sample observations. We introduce controls for clinic location and size. In addition, we estimate the model with fixed effects for month to control for seasonality. In sum, we estimate the following full model:

$$\begin{aligned}\%\Delta\text{HRS}_{it} = {} & \beta_{x1}\text{NEG}_{it} + \beta_{x2}\text{SMALL}_{it} + \beta_{x3}\text{POS}_{it} \\ & + \beta_{x4}\text{NEG}_{it} * \%\Delta\text{VISITS}_{it} + \beta_{x5}\text{SMALL}_{it} * \%\Delta\text{VISITS}_{it} \\ & + \beta_{x6}\text{POS}_{it} * \%\Delta\text{VISITS}_{it} + \beta_{x7}\text{HOSP}_{i} + \beta_{x8}\text{SIZE}_{i} \\ & \sum_{n=2}^{12} \beta_{x7+n}\text{MONTH}_{n} + \varepsilon_{Xit}, \ (X \in N, E, S),\end{aligned} \tag{4}$$

where

HOSP = 1 if clinic is in a hospital, 0 if it is in a SNF.
SIZE = quartile for the clinic's average number of visits per month (1 through 4).
MONTH_n = 1 if the observation takes place in month n, 0 otherwise.

Other variables as defined earlier.

We make identical arguments to develop a model with $\%\Delta\text{COST}_{it}$ as the dependent variable. We estimate the above system of equations after replacing $\%\Delta\text{HRS}_{it}$ with $\%\Delta\text{COST}_{it}$ as the dependent variable.

2.4 Measurement Issues

The period over which we should measure the variables needed to estimate eq. (4) is unclear. While a month might seem the appropriate period because it is the smallest time frame with available data, a month may be too short a period to capture the relation between staffing and demand patterns, since therapists are salaried employees and managers have finite scheduling flexibility. A longer period is also beneficial because the choice mitigates the effect of coding and other errors. Too long a period, however, diminishes the difference between the flexibility in adjusting spending (hiring /firing therapists) and in adjusting consumption (scheduling practices to manage available time). We chose to define a period as a three-month average.[8] That is, we define[9]

$$\%\Delta\text{HRS}_{it} = \frac{\text{HRS}_{it+3} - \text{HRS}_{it}}{\sum_{t=1}^{3} \text{HRS}_{it}}.$$

As is evident from dividing the numerator and denominator by three, this formulation captures the average change over a three-month period divided by the average activity level over the same time. We chose this metric because it does not

8. As a robustness check, we also run the model using nonoverlapping quarters. Results were substantially weaker, perhaps because of the dramatic reduction in sample size.

9. This is derived from the percent change in 3-month moving average number of visits:

$$\frac{(\text{HRS}_{t+1} + \text{HRS}_{t+2} + \text{HRS}_{t+3})/3 - (\text{HRS}_{t} + \text{HRS}_{t+1} + \text{HRS}_{t+2})/3}{(\text{HRS}_{t} + \text{HRS}_{t+1} + \text{HRS}_{t+2})/3}.$$

dramatically reduce the number of observations, reduces noise in the data, and provides a sufficient period for managers to adjust to any perceived change in demand patterns.[10] We employ the same convention to measure changes in activity levels and cost.

We measure $\%\Delta HRS_{it}$ as the change in the number of staffed (i.e., available hours)[11] and measure $\%\Delta COST_{it}$ as the percent change in total recorded salary expense for therapists plus any payments to contract employees.[12]

We use the average staff time available per visit to classify clinics into three broad groups representing differing levels of capacity utilization. Specifically, we develop a distribution of the average staff time available per visit. Using separate distributions for hospitals and skilled nursing facilities (SNF), we use extreme quartiles to classify clinics as having excess or strained resources. We use prior month's staff hours to visit ratio to define capacity utilization because managers do not have access to the current month's volume. Overall, we classify 24 percent (457 of 1,898) of clinic-months as having excess capacity and 27 percent (513 of 1,898) of clinics as having strained resources. (Our conclusions do not change with alternate definitions for capacity utilization.)

Earlier, we conjectured that both the direction and magnitude of the change in number of visits would affect the manager's response. To test this conjecture, we divided the change in visits into three groups. In particular, if $-0.03 < \%\Delta VISITS < 0.03$ (i.e., the number of visits changed by less than 3% in either direction), we classified the observation as representing a "normal" (SMALL) change due to random factors. The change was classified as a negative (NEG) change if the value was less than or equal to -0.03, and as a positive (POS) change if the value was larger than or equal to 0.03. This results in classification of 37.5 percent of the observations as large negative changes, 30.7 percent as small changes, and 31.7 percent as large positive changes. Other cutoffs yield qualitatively similar results.

Finally, we use the average number of visits per month to classify clinics into four groups, coded from one through four. We also note that the variable HOSP itself is a partial control for size because clinics in hospitals are significantly larger than those in SNFs. (Dropping the SIZE variable in the model does not change conclusions. The estimated coefficient for the variable is not statistically significant.)

---

10. The Lagrange multiplier test (Greene [2000, p. 541]) does not indicate significant autocorrelation. The Durbin-Watson test is inconclusive. We also estimated the model using non-overlapping quarters. We reach similar conclusions, although the response coefficients are larger and are often greater than one (significantly so in some instances).

11. We did not distinguish between staffed and contract hours. Firm personnel believed that the breakdown between contract hours and salaried hours is very unreliable. Also, many observations recorded a 0 for contract hours even if they had contract pay. We emphasize that this measure excludes "comp" time, the primary way to move capacity across periods.

12. We exclude payments to aides and assistants as well as all operating expenses. We exclude benefits because the benefit data does not break out the benefits payable to therapists from the benefits payable to aides and other employees.

3. Sample Characteristics

Our primary data source is a clinic-level management report on the number of hours staffed, by type of employee for 49 physical therapy clinics. Available data include clinic-level monthly expense, payroll, and patient visit information from January 1992 to December 1997. For the period 1994–1997, this report also provides data on the number of visits (and statistics such as average salary rates.) For earlier years, we hand-collected visit data from reports sent by the clinics to the central office. Available data do not provide a detailed cost breakdown of costs other than labor cost.

From the original database, we selected clinics with at least six months of patient visit data reported on a consistent and continuing basis. Descriptive statistics for the clinics represented in the screened database indicated the presence of outlier values, perhaps because of coding errors. We reduced the effect of outliers on the analysis by winsorizing each individual data element (the number of visits, the number of staff and aide hours) to the 5th and 95th percentile of the respective distribution. Even after this, we found that data indicated extreme volatility of operations for some clinics.[13] In the interest of capturing behavior in a stable environment, we then deleted all clinics for which the ratio of the largest number of visits to the smallest number of visits was larger than five.[14] This means that we capped growth at 500 percent over a six-year (or shorter, if the time-series was shorter) period. These restrictions reduced the number of observations to 1,898 monthly observations from 49 clinics.

3.1 Sample Characteristics

Of the 49 clinics included in the sample, 26 are located inside hospitals (HOSP) and the rest are in skilled nursing facilities (SNF).[15] Data reported in Table 1 indicate that the average clinic was staffed for 590 therapist hours. As indicated by the median of 448 hours, the distribution is skewed. In addition, as indicated by the end points of the distributions and by clinic level analysis (data not reported), there is considerable variation in the size of the clinics. We find that clinics located within hospitals are larger than clinics located in SNFs. There is similar variation in the number of aide-hours utilized by a clinic. Almost 7 percent of the clinics did not use any aide-hours at all. Accordingly, we focus attention on hours staffed

13. For example, within the same 15-month period, the maximum number of visits in a clinic was over 10 times the minimum number of visits. We believe that the behavior is due to miscoding, but cannot rule out volatility as the true underlying cause. Inspection reveals that the unusual observations occur almost at random. We do not believe that introducing a control variable for growth would solve the problem.

14. We recognize the arbitrary nature of this cutoff. In the trade-off between more observations and more confidence in data integrity, we chose to err toward the latter.

15. From the viewpoint of a hospital, our research firm was a subcontractor that was physically located within the hospital.

TABLE 1

Descriptive Statistics

Variable	Facility Location	*N*	Mean	Median	Standard Deviation
Staff therapist hours	SNF[1]	844	341.72	295.50	193.58
	Hospitals	1,054	789.60	679.00	539.18
	All clinics	1,898	590.44	448.34	477.06
Aide and assistant hours	SNF	844	315.50	255.83	264.85
	Hospitals	1,054	669.80	448.50	604.02
	All clinics	1,898	512.87	334.00	514.97
Volume of patient visits	SNF	844	404.17	306.33	258.88
	Hospitals	1,054	1010.38	746.00	811.64
	All clinics	1,898	740.81	477.50	697.31
Staffing cost/visit ($)	SNF	844	23.68	22.49	9.64
	Hospitals	1,054	20.21	19.11	7.68
	All clinics	1,898	21.75	20.29	8.78

[1]SNF = skilled nursing facility.

by licensed therapists. (For comparison purposes, we also estimated the model using total hours staffed as the dependent variable and reached similar conclusions.)

Data regarding the number of patient visits in a month exhibit similar variations by clinic location. The mean (median) number of visits is 741 (478) per month. Consistent with the staffing pattern, the distribution is skewed; hospitals tend to have larger clinics. The median staffing cost per visit is $20.29. Cost is higher in SNF-based clinics.

The number of patient visits exhibits some seasonality (results not tabulated). We find that in hospitals, winter and late summer months tend to have fewer visits. The pattern in winter is consistent with patients wishing to postpone elective surgery (such as a hip replacement, which is followed by extensive physical therapy) until after the holiday season. SNF-based clinics have a more consistent pattern of visits, although the number of treatments in late summer months is also somewhat lower. We control for this seasonality with the variable $MONTH_n$ in the model.

Panel A of Table 2 reports data after partitioning clinics by capacity utilization. The number of visits is monotonic in our measure of capacity utilization. Clinics with excess capacity have lower demand than clinics with normal utilization, which, in turn, have lower demand than clinics experiencing strained resources. This pattern provides face validity to our measure of capacity utilization, although we hasten to note that this table does not control for clinic size.

Panel B of Table 2 provides data on the persistence of excess or strained resources. We find that there is an 81.6 percent (81.6%) chance that a clinic categorized as having excess (strained) capacity one month will be in the same classification the next month. The probability declines to 67.8 percent (70.0%) with a two-month horizon. Thus, it appears that managers do not react very quickly to remove unused capacity or to relieve strain.

TABLE 2

Descriptive Statistics: Capacity Utilization

Panel A: Median number of visits per month, by capacity utilization

Facility Location	*N*	SNF[1]	Hospital	Total
Excess capacity utilization	457	256	623	460
Normal capacity utilization	928	412	904	687
Strained capacity utilization	513	518	1,552	1,087

Panel B: Persistence in capacity utilization classification (probability of change in status)

	Status Next Month				Status Two Months Later			
Current Month	Excess	Normal	Strained	Total	Excess	Normal	Strained	Total
Excess	81.6	15.8	2.6	100	67.8	25.8	6.4	100
Normal	7.6	82.4	10.0	100	13.3	71.6	15.1	100
Strained	2.2	16.2	81.6	100	5.2	15.8	70.0	100

[1]SNF = skilled nursing facility.

Capacity utilization was defined separately for each location. We first derived the distribution of the available staff time per visit. For month t, if the available staff time per visit in a clinic fell in the top 25 percent (bottom 25%) of the distribution, the clinic was classified as having excess (strained) resources in month $t + 1$. Otherwise, month $t + 1$ was classified as having normal utilization.

The persistence of the classification with respect to capacity utilization, however, should not be interpreted as resulting in no change in clinic classification. Of the 49 clinics in the sample, 35 were placed into each of the three capacity classifications at least once. Thirteen clinics were never classified as having excess capacity, whereas only one clinic was never classified as having strained resources. We interpret this pattern as consistent with managers seeking to optimize resource utilization.

3.2 Magnitude of Change

Looking at the change variables, (results not tabulated), we note a consistency with the time trend reported earlier, that is, median growth is zero. Maximum values for change indicate some unreasonable values (e.g., growth of 240%), indicating that our attempts at removing outliers (using three-month moving averages, eliminating clinics with anomalous data) were not entirely successful. Staffing changes were comparable in magnitude to the change in demand (visits).

Table 3 provides descriptive statistics concerning our partition for change in visits. Panel A reports the persistence of classification into negative, small, and positive changes in demand. The classification persistence is much lower than for our capacity utilization partition; clinics that are classified as negative (positive) have a likelihood of only 59.5 percent (49.8%) of being classified in the same

TABLE 3

Descriptive Statistics: Persistence of Demand

Panel A: Persistence of Demand Change

Current Month	Status Next Month (%)				Status Two Months Later (%)			
	Negative	Small	Positive	Total	Negative	Small	Positive	Total
Negative	59.5	19.1	21.4	100	42.7	21.9	35.3	100
Small	27.1	47.0	25.9	100	33.7	43.9	22.4	100
Positive	21.7	28.6	49.8	100	35.1	28.3	36.5	100

Panel B: Distribution of demand shocks and capacity utilization

	Negative (%)	Small (%)	Positive (%)	Total (%)
Excess	33.9	26.0	40.1	100
Normal	34.5	32.0	33.5	100
Strained	45.6	32.2	22.1	100

A change in panel A was classified as a large positive (negative) shock if the percent increase (decrease) of visits in month *t* was more than (less than) 3 percent. All other changes are classified as a small change.

A change in panel B was classified as a positive (negative) shock if the percent increase (decrease) of visits in month *t* was more than (less than) 3 percent. All other changes are classified as a small change.

category in the following month. Two months later, the corresponding percentages are 42.7 percent and 36.5 percent. Three months later (results not tabulated), clinics classified as having negative (positive) changes have only 29.8 percent (21.2%) probability of being classified in the same group.

Panel B of Table 3 cross-tabulates capacity utilization and the change in visits. We find that large positive (negative) changes in visits are more likely if the clinic is currently experiencing excess (strained) capacity. This finding supports the assertion that capacity utilization is not a clinic-specific measure, and that clinics experience differing capacity utilization in different months.

4. Results

The first data column in panel A of Table 4 reports data for the entire sample (without conditioning for capacity utilization levels). We draw three conclusions:

- Managers do not significantly change staffed hours in response to "small" (3% or less) changes in activity levels (SMALL * %ΔVISITS). The insignificance of the coefficient results from a large standard error.
- There is a significant (but less than proportionate) response for "large" changes, independent of direction. However, as reported in the first two rows of panel B of Table 4, the difference in response to small versus large

TABLE 4

Response in Staffed Therapist Hours to Changes in Visits

Panel A: Regression results[1]

$$\%\Delta HRS_{it} = \beta_{X_1}NEG_{it} + \beta_{X_2}SMALL_{it} + \beta_{X_3}POS_{it} + \beta_{X_4}NEG_{it} * \%\Delta VISITS_{it} + \beta_{X_5}SMALL_{it} * \%\Delta VISITS_{it} + \beta_{X_6}POS_{it} * \%\Delta VISITS_{it} + \beta_{X_7}HOSP_i + \beta_{X_8}SIZE_i + \sum_{n=2}^{12} \beta_{X_{7+n}}MONTH_n + \varepsilon_{X_{it}}$$

	All Observations	Excess[1] Capacity	Normal Utilization	Strained Resources
NEG	0.017	−0.004	0.034***	0.029**
SMALL	−0.020*	−0.015*	0.005	0.034***
POS	0.009	−0.033***	−0.044***	0.020*
NEG * %ΔVISITS	0.358***	0.538***	0.623***	0.047
SMALL * %ΔVISITS	0.291	−0.048	0.064	0.008
POS * %ΔVISITS	0.509***	0.390***	0.655***	0.387***
Adj. R^2	0.21	0.18	0.30	0.15

Panel B: Tests for significant differences across variables

Test	Subsample(s)	Prediction	Difference	F Value
Magnitude of response				
(POS * %ΔVISITS − SMALL * %ΔVISITS)	All	>0	0.218	0.19
(NEG * %ΔVISITS − SMALL * %ΔVISITS)	All	>0	0.078	0.02
Sticky costs				
(POS * %ΔVISITS − NEG * %ΔVISITS)	All	>0	0.151	5.97**
	Excess	<0	−0.148	4.47**
	Normal	>0	0.032	0.22
	Strained	>0	0.340	33.79***
Interaction with capacity utilization				
Decreases in activity: (NEG * %ΔVISITS)	Strain − Excess	<0	−0.491	43.64***
Increases in activity (POS * %ΔVISITS)	Strain − Excess	>0	−0.003	0.01

%ΔHRS = Percent change in moving average of staffed hours.
NEG = 1 if change in visits decreases by at least 3 percent, 0 otherwise.
SMALL = 1 if change in visits changes by less than 3 percent, 0 otherwise.
POS = 1 if change in visits increases by at least 3 percent, 0 otherwise.
%ΔVISITS = Percent change in moving average of visits.
HOSP = 1 if clinic is located in a hospital, 0 otherwise.
SIZE = Average number of visits quartile for clinic (1–4).
$MONTH_i$ = 1 if observation is in month *i*, 0 otherwise.

[1]Capacity definitions are based on quartiles of average staff time available per visit, estimated separately for hospital-based and SNF-based clinics. Overall, 24.0 percent of clinics are classified as having excess capacity and 27.05 percent are classified as having strained resources.

Note capacity-partitioned model is estimated using SUR. Coefficients for control variables are not reported.

*, **, *** Indicate significance at the 0.10, 0.05, and 0.01 levels, respectively.

changes is not statistically significant. Thus, evidence does not support our assertion that the response to large changes differs from the response to small changes in activity levels.

- Finally, as hypothesized by Cooper and Kaplan (1998) and confirming the results in Anderson et al. (2003), the response for a decrease in activity is significantly smaller than that for an increase in activity levels, when we consider the sample as a whole.

Our focus is on the response to the direction of change ("sticky cost"), and its interaction with capacity utilization (see columns 2–4 in panel A of Table 4). These data confirm and strengthen the first two results reported for the full sample. The response to a small change in activity is not distinguishable from zero, suggesting that staffed hours are held constant for random demand fluctuations. In addition, the response to a large change significantly differs from that for a small change.

Our primary contribution is refining the third bullet point, which is the "sticky" cost hypothesized by Cooper and Kaplan, and confirmed by Anderson et al. (2003) and others. In particular, our evidence points to systematic differences due to current capacity utilization.

- *Excess Capacity.* The response to large decreases in activity levels is significantly *larger* than the response for an increase in activity ($F = 4.47$, $p < 0.05$). This is in direct contrast to the "sticky" cost hypothesis, and shows that current utilization affects the managerial response.
- *Normal Utilization.* Our results for clinics with normal capacity utilization are inconclusive. The response for a decline in activity levels is *similar* to the response for an increase in activity levels ($F = 0.22$, $p > 0.10$).
- *Strained Resources.* Our prediction for clinics experiencing strained capacity is identical to that made by Cooper and Kaplan. As predicted, we find a significantly *smaller* response to a decline in activity levels relative to the response for an increase ($F = 33.79$, $p < 0.01$). Indeed, the response to a decline in activity is not significantly different than zero.

These results underscore the importance of considering current capacity utilization when predicting managerial response.

The last section of panel B of Table 4 reports results from tests that compare coefficients across equations. We find that the response to a decrease in activity is substantially larger when capacity is plentiful relative to when it is scarce. However, current utilization does not appear to impact the response to an increase in activity levels. We can only conjecture that managers are prone to increasing resources following increased demand. However, unless there is clear evidence of demand decline (excess capacity and a decreasing demand), they are less likely to reduce capacity. Our evidence therefore supports "sticky costs" overall but highlights the importance of considering capacity utilization when constructing cost models that incorporate sticky costs.

Table 5 reports the results from estimating eq. (4) using therapist salary cost

TABLE 5

Salary Cost Response to Changes in Visits

Panel A: Regression estimates[1]

$$\%\Delta COST_{it} = \beta_{X_1}NEG_{it} + \beta_{X_2}SMALL_{it} + \beta_{X_3}POS_{it} + \beta_{X_4}NEG_{it} * \%\Delta VISITS_{it} + \beta_{X_5}SMALL_{it} * \%\Delta VISITS_{it} + \beta_{X_6}POS_{it} * \%\Delta VISITS_{it} + \beta_{X_7}HOSP_{i} + \beta_{X_8}SIZE_{i} + \sum_{n=2}^{12}\beta_{X_{7+n}}MONTH_{n} + \varepsilon_{X_{it}}$$

	All Observations	Excess Capacity	Normal Utilization	Strained Resources
NEG	0.011	0.028**	0.037***	0.024
SMALL	0.004	0.012	0.007	0.038***
POS	−0.015	−0.004	−0.027***	0.048***
NEG * %ΔVISITS	0.336***	0.653***	0.622***	−0.092
SMALL * %ΔVISITS	0.523	1.014	0.369	0.214
POS * %ΔVISITS	0.456***	0.362***	0.643***	0.186***
Adj. R^2	0.12	0.11	0.24	0.063

Panel B: Tests for significant differences across variables

Test	Subsample(s)	Prediction	Difference	F Value
Magnitude of response				
(POS * %ΔVISITS − SMALL * %ΔVISITS)	All	>0	−0.07	0.01
(NEG * %ΔVISITS − SMALL * %ΔVISITS)	All	>0	−0.19	0.09
Sticky costs				
(POS * %ΔVISITS − NEG * %ΔVISITS)	All	>0	0.120	2.23
	Excess	<0	−0.291	12.84***
	Normal	>0	0.021	0.07
	Strained	>0	0.278	17.32***
Interaction with capacity utilization				
Decreases in activity: (NEG * %ΔVISITS)	Strain − Excess	<0	−0.745	63.33***
Increases in activity (POS * %ΔVISITS)	Strain − Excess	>0	−0.176	3.93

%ΔHRS = Percent change in moving average of staffed hours.
NEG = 1 if change in visits decreases by at least 3 percent, 0 otherwise.
SMALL = 1 if change in visits changes by less than 3 percent, 0 otherwise.
POS = 1 if change in visits increases by at least 3 percent, 0 otherwise.
%ΔVISITS = Percent change in moving average of visits.
HOSP = 1 if clinic is located in a hospital, 0 otherwise.
SIZE = Average number of visits quartile for clinic (1–4).
$MONTH_i$ = 1 if observation is in month i, 0 otherwise.

[1]Capacity definitions are based on quartiles of average staff time available per visit, estimated separately for hospital-based and SNF-based clinics. Overall, 24.0 percent of clinics are classified as having excess capacity and 27.05 percent are classified as having strained resources.

Note capacity-partitioned model is estimated using SUR. Coefficients for control variables are not reported.

*, **, *** Indicate significance at the 0.10, 0.05, and 0.01 levels, respectively.

as the dependent variable. Inspection reveals qualitatively similar results to the results reported in Table 4.

5. Conclusion

In this paper, we explore the "sticky" cost hypothesis postulated by Cooper and Kaplan (1998) and confirmed by Anderson et al. (2003). The sticky cost hypothesis is that a manager's cost response to change in activity levels is affected by the direction of the change. We conjecture that the response is also affected by the magnitude of the change. Thus, significant transactions costs appear to exist, consistent with intuition. In addition, we posit an interaction effect—the response to a decline in activity levels is smaller (larger) than that for an increase only when capacity is currently strained (in excess). Capacity utilization may therefore be an important omitted variable in cross-sectional studies of cost behavior.

REFERENCES

Anderson, M., R. Banker, and S. Janakiraman. 2003. "Are SG&A Costs Sticky?" *Journal of Accounting Research* 41 (March): 47–63.

Balakrishnan, R., and N. Soderstrom. 2000. "The Cost of Congestion in Healthcare." *Journal of Management Accounting Research* 12:97–114.

Cooper, R., and R. Kaplan. 1992. "Activity based Cost Systems: Measuring the Cost of Resource Usage." *Accounting Horizons* 6: 1–13.

———. 1998. *The Design of Cost Management Systems: Text, Cases and Readings.* 2nd ed. Upper Saddle River, N.J.: Prentice Hall.

Greene, W. H. 2000. *Econometric Analysis.* Upper Saddle River, N.J.: Prentice Hall.

Noreen, E. 1991. "Conditions under Which Activity-based Costs Provide Relevant Costs." *Journal of Management Accounting Research* 3:159–168.

Noreen, E., and N. Soderstrom. 1994. "Are Overhead Costs Strictly Proportional to Activity?" *Journal of Accounting and Economics* 17 (January): 255–278.

———. 1997. "The Accuracy of Proportional Cost Models: Evidence from Hospital Service Departments." *Review of Accounting Studies* 2:89–114.

Subramaniam, C., and M. Weidenmier. 2002. "Additional Evidence on the Sticky Behavior of Costs." Working paper, Texas Christian University.

Wiersma, E. 2002. "Asymmetric Cost Behavior: The Managerial Evaluation System, Flexibility of Resources, Economies of Scale, and Predictability of Activity Levels as Drivers of Stickiness." Working paper, Vrije Universiteit, Amsterdam.