Dynamic Pricing in Retail Gasoline Markets

Severin Borenstein; Andrea Shepard

The RAND Journal of Economics, Volume 27, Issue 3 (Autumn, 1996), 429-451.

Stable URL:
http://links.jstor.org/sici?sici=0741-6261%28199623%2927%3A3%3C429%3ADPIRGM%3E2.0.CO%3B2-%23

RAND Journal of Economics
Vol. 27, No. 3, Autumn 1996
pp. 429–451

Dynamic pricing in retail gasoline markets

Severin Borenstein*

and

Andrea Shepard**

Supergame models of tacit collusion show that supportable price-cost margins increase with expected future collusive profits, ceteris paribus. *As a result, collusive margins will be larger when demand is expected to increase or marginal costs are expected to decline. Using panel data on sales volume and gasoline prices in 43 cities over 72 months, we find behavior consistent with tacit collusion in retail gasoline markets. Controlling for* current *demand and cost, current margins increase with expected next-month demand and decrease with expected next-month cost. The results are not consistent with intertemporal linkages due to inventory behavior or customer loyalty.*

1. Introduction

■ In recent years, economists have developed numerous models to distinguish empirically between collusive and noncollusive pricing behaviors.[1] Generally, these studies have attempted to uncover collusive behavior by estimating relationships among contemporaneous observations of output, cost, and price. In contrast to most of these studies, we present a test for collusion that is based on the relationship between current price and expected future demand and cost conditions. Most prior studies have tested for collusion in industries best characterized as tight oligopolies: the U.S. automobile industry and 19th-century U.S. railroads, for example. We examine behavior in markets with many, differentiated firms.[2]

Our approach exploits the insights from supergame models in which tacitly collusive outcomes are supported by repeated play. In these models, self-enforcing collusion depends on the current gain from defecting being smaller than the anticipated

* University of California, Berkeley, University of California, Davis, and NBER; sjborenstein@ucdavis.edu.

** Stanford University and NBER; fshepard@lira.stanford.edu.

We thank Dennis Carlton, Sam Peltzman, Jim Poterba, Peter Reiss, Mike Riordan, Garth Saloner, two anonymous referees, and seminar participants at the NBER Industrial Organization Summer Institute, U.C. Berkeley, Boston University, University of Chicago, and Stanford University for helpful comments and discussion. Jamie Woods provided excellent research assistance. Any remaining errors are our own.

[1] These include Porter (1983, 1985), Bresnahan (1987), Porter and Zona (1993), and Ellison (1994). Bresnahan (1989) provides a useful review.

[2] Throughout this article we use "collusion" to mean implicit collusion supported by repeated interaction. Successful collusion of this type does not imply that firms in these markets are earning excess profits or engaging in activities that violate U.S. antitrust laws.

future loss from the punishment triggered by defection. The gain from defecting occurs at the time of the defection and increases in current collusive profits. The punishment loss is realized in the future when other firms respond to the defection. The expected punishment loss increases in the expected difference between the future collusive profits that would have been earned absent defection and the future profit earned when firms are punishing defection. Collusion is more difficult to sustain (i.e., the highest sustainable collusive margin will be lower) when either the gain from defection is greater or the anticipated loss from punishment is lower. Using this general framework, Rotemberg and Saloner (1986) construct a model in which firms anticipate changes in demand. They show that when current demand is higher (lower) than expected future demand, collusion is more (less) difficult to sustain because the gain from cheating increases in current demand while the loss from punishment increases in future demand. Using an analogous argument, collusion is more (less) difficult to sustain when current costs are lower (higher) than expected future costs. Haltiwanger and Harrington (1991) apply this logic to a deterministic demand cycle and show that, holding constant the current level of demand and cost, collusion is more difficult to sustain when demand is declining or cost is increasing.

These models predict that current margins will respond to expected future demand or cost when firms are pricing collusively. In particular, the models predict that current margins will respond positively to expected future demand and negatively to expected future cost. These predictions are inconsistent with standard noncooperative models in which current margins are not a function of expected future conditions. Noncollusive models that could explain a link between today's margins and expected future market conditions usually rely on either consumer loyalty or inventory behavior. We argue below that these alternatives have empirical implications that are not supported by the data.

We test for collusive pricing in retail gasoline markets by examining retail margins (price at the pump minus the wholesale price). In some respects, these markets are a natural setting for the test. The theoretical models rely on predictable changes in demand and marginal cost, and there are predictable changes both in the retail demand for gasoline and in wholesale gasoline prices, which is our proxy for marginal cost.[3] Because there is a marked seasonal cycle in the demand for gasoline, some of the movement in demand can be anticipated. Because these are lags in the response of wholesale gasoline prices to crude oil shocks, some of the variation in wholesale prices is also predictable.

The structure of these markets, however, is not typical of those in which collusion is commonly viewed as probable. Collusive behavior in these markets implies at least tacit cooperation over prices at gasoline stations. In the urban areas we study, there are many more gasoline stations than the "few" firms assumed in the theoretical models.[4] Since collusion is more difficult to sustain as the number of firms increases, the existence of many firms seems incompatible with cooperative pricing. However, gasoline stations sell a product that is differentiated by brand, service, and, most importantly, location. The number of firms operating in a metropolitan area, therefore, may not be a good indicator of the number of effective competitors in any actual retail market. Further, prices at stations need not be set independently: some stations have a common manager who sets prices at all stations he/she owns or manages. In addition, retail prices are posted and can be changed quickly and at low cost, making it easy to detect

[3] Wholesale gasoline price constitutes about 85% of the retail price. The other significant contributors to the marginal retail cost are (some) labor costs and the costs of delivering gasoline from the terminal to the retail outlet. Neither of these other components are nearly as volatile as the wholesale price.

[4] See Borenstein and Gilbert (1993) for information on retail market structures.

defection and respond rapidly. As a result, the gain from defecting will be small, making tacit collusion easier to sustain.

Further, prior research indicates that pricing in retail gasoline markets is not well characterized by standard competitive models. Slade (1986) presents evidence from a single retail market in Vancouver that station-level demand is not perfectly elastic and rejects the hypothesis of competitive pricing. In related work, Slade (1987) concludes that pricing in the Vancouver market is characterized by implicit collusion in which periods of cooperation alternate with price wars triggered by demand shocks. Borenstein (1991) and Shepard (1991) show that U.S. gasoline stations have sufficient local market power to implement price discrimination across gasoline grades or service levels. Although the structure of the market makes it unlikely that the firms come close to achieving the monopoly price in a supergame equilibrium, these results suggest that the possibility of some collusive pricing cannot be rejected *a priori*.

Our work is most closely related to Ellison's (1994) test of the Rotemberg and Saloner model using data on railroad prices and outputs during the era of the Joint Executive Committee cartel. Ellison models the collusive price as a function of the ratio of current to expected future demand. He finds no evidence of an effect on current margins, but notes that the explicit railroad cartel might not be an environment in which the implicit collusion of the Rotemberg-Saloner model is likely to apply. Hajivassiliou (1989) also tests this model in the Joint Executive Committee cartel, but relies on contemporaneous price and quantity data only. He finds little support for the prediction that collusion is less likely when current demand is high.

The results of our analysis are consistent with the predictions of the collusive pricing models. Controlling for current demand and wholesale price, we find that current margins increase with expected next-period demand and decline with expected next-period wholesale price. The magnitudes of these effects are not large in absolute value: increasing expected next-period demand by 10% increases the current margin, which averages about eleven cents, by about .42 cents. Increasing the expected wholesale price of gasoline by ten cents reduces current margins by about .63 cents. The economic importance of these effects is not in the magnitude of their impact on consumer prices, but in what they suggest for the behavior of firms in these markets. They indicate that departure from single-period Nash behavior may be more common than is suggested by studies focusing on tight oligopolies. They also demonstrate that the technique used here to uncover firm behavior might be useful in markets where collusion would have more pronounced welfare effects.

In the following section we describe the models on which the empirical work is based. In Section 3 we discuss the data we use and in Section 4 the estimation procedure. The results are presented in Section 5. We discuss alternative hypotheses in Section 6 and offer concluding comments in Section 7.

2. Models of price dynamics

■ Bresnahan (1989) and others point out that diagnosing collusive pricing from only contemporaneous price, cost, and demand data is quite difficult. The fact that noncooperative behavior is consistent with a variety of pricing patterns poses a serious problem in these endeavors. For example, noncooperative prices may increase or decrease in response to positive demand shocks, depending on the cost structure and whether the shocks change the composition of demand as well as its level.[5] As a result, efforts to distinguish collusive from noncooperative behavior generally have relied upon restrictive assumptions about the functional forms of demand and cost. Recent models

[5] For this reason, we do not focus on the effect of current demand to diagnose implicit collusion.

of collusion, however, have implications for dynamic pricing that more easily distinguish the behavior consistent with these specific models from noncooperative outcomes.

Rotemberg and Saloner (1986) develop a model in which firms sustain implicit collusion by adjusting current margins in response to anticipated changes in demand. In their model, very-high-demand periods are outlying realizations of an independently and identically distributed demand shock. An individual firm has a relatively large incentive to deviate from the collusive price in these periods because it is able to capture a share of an unusually large market by doing so. Current deviations would be punished by lower prices in future periods. If deviation is to be prevented, the potential loss earned by the firm in the punishment phase must be at least as great as the potential gain from deviating in the current period. The punishment loss anticipated by a deviating firm is the present value of the difference between the profits it expects to earn in the future while colluding and those it expects to earn under the lower, punishment prices. If demand is independently and identically distributed, current demand realizations have no effect on expectations of future demand, and the expected loss is constant. In high-demand periods, then, collusion can be sustained only by reducing the gains to deviating, i.e., by reducing current collusive profits. The highest sustainable collusive margin will therefore be lower in periods after which demand is expected to decline.

Haltiwanger and Harrington (1991) reformulate this model in the context of a deterministic demand cycle. In this environment, it is possible to distinguish the level of demand from the expected change in demand.[6] To understand their model, it is easier to reverse the thought experiment and hold constant the gain from deviating. Consider a simple model in which firms have constant marginal cost, produce an undifferentiated product, and engage in Bertrand competition during punishment periods. Assume that monopoly (or perfectly collusive) profits would be procyclical (i.e., increase with demand). Notice that the Bertrand assumption and constant marginal costs imply that punishment-period profits are always zero.

Consider two periods, t_i and t_j, with equal demand. Because current demand is equal, the gain from deviating is equal. Suppose that demand is increasing at t_i and declining at t_j. Near-term, future collusive profits are, therefore, expected to be higher at time t_i than at t_j. Because near-term profit is weighted more heavily in evaluating the present value of future profits, the expected collusive profit forgone by deviating is higher at t_i. With punishment profit constant by assumption, this means that the loss from deviating will be higher at t_i. The highest sustainable collusive margin will therefore be higher at t_i than at t_j. In this formulation, margins respond positively to changes in expected near-term demand, holding constant current demand.

Both the Rotemberg-Saloner and Haltiwanger-Harrington models explicitly hold marginal cost constant over time. If changes in input prices create predictable shifts in marginal cost, however, it is straightforward to show that these models also imply that margins will be affected by expectations about future marginal cost. Suppose demand does not change over time and marginal cost is invariant to output. Then if input prices are expected to rise next period, the expected increase in marginal cost will cause expected collusive profits to decline. This will reduce the potential loss from future punishment. Therefore, holding constant current marginal cost, expected future changes in marginal cost will have a negative effect on current margins.

This discussion suggests three tests for collusive pricing. First, collusive margins will respond to anticipated changes in cost and demand. Second, controlling for current

[6] In the Rotemberg-Saloner model, the distinction between level and derivative cannot be made. A high demand necessarily implies a negative expected derivative, and a low demand implies a positive expected derivative.

demand, margins will respond positively to expected increases in near-term demand. Finally, controlling for current input prices, margins will respond negatively to expected increases in the input prices.

Before proceeding to the empirical test of these predictions, it is important to recognize that these stylized game-theoretic models do not precisely comport with conditions in retail gasoline markets. For our application, the most troublesome assumption of these models is that punishment profits are invariant to demand and cost movements. If punishment means a reversion to a noncooperative equilibrium so that "punishment" and "noncollusive" profits are equal, this assumption might be violated in gasoline markets. On the demand side, there is reason to believe that noncollusive margins would be changed with demand.[7] In this case, the preceding argument about the relationship between expected demand changes and collusive margins relies on the effect of demand changes on collusive profits being larger than the effect of demand changes on noncollusive profits. Similarly, the prediction that margins will respond negatively to expected future input price changes assumes that a change in cost will change collusive profits more than noncollusive profits. While that is the case in many models, it is not general.

What this implies for our empirical work is that the signs of the effects of future demand and input costs on current margins are not unambiguous, as they are in the stylized models. *Margins should still respond to expected future cost and demand conditions when firms are colluding.* So a finding of any effect of expected future conditions on current margins supports the claim that firms are not in a noncooperative equilibrium. However, we are not certain that the effects will run in the direction predicted by the models.

3. The data

■ The estimation procedure uses data on retail prices, wholesale prices, crude oil prices, and gasoline demand. We describe these data in this section, along with some relevant details on gasoline distribution. Descriptive statistics appear in Table 1.

Retail margin is defined as the retail price (the price paid by consumers at the pump) minus the wholesale price of gasoline. The price at the pump is clearly defined, but selecting the appropriate "wholesale" gasoline price is complicated by the structure of gasoline production and distribution and by observability. Retail gasoline markets are local: consumers typically buy gasoline within a limited geographic area. The relevant price for a given retail market, then, is the wholesale price of gasoline in that geographic area. Wholesale gasoline markets are defined by terminal locations. Most larger cities (and all those in our sample) have a city terminal supplied with gasoline by pipeline or water transport. From the terminal, gasoline is trucked to gasoline stations by refiners or by independent distributors who purchase gasoline at the terminal. Any given station has a single wholesale supplier.

Approximately 55% of the retail gasoline sold in the United States is sold at stations supplied by independent wholesales (called "jobbers" in the gasoline industry; see Temple, Barker, and Sloan, Inc. (1988)). At the terminal, jobbers are charged a posted price set by refiners on a terminal-by-terminal basis. This is called the "terminal price." Jobbers own and operate some of the stations they supply, and for gasoline sold at these stations there is no additional wholesale market transaction. At stations that jobbers supply but do not operate, there is a transaction between the jobber and

[7] One argument is that high-demand periods have disproportionately more demand from vacation travelers who are buying gasoline in unfamiliar locations and therefore have higher search costs. Higher search costs imply that firm-level elasticities will be higher. For a more complete discussion, see Borenstein and Shepard (1994).

Table 1 Descriptive Statistics (2,873 observations)

Variable	Mean	Standard Deviation	Minimum	Maximum
Retail price (cents per gallon) *RETAIL*	72.81	12.56	43.18	126.57
Terminal price (cents per gallon) *TERMINAL*	61.82	11.94	36.00	111.88
Crude oil price (cents per gallon) *CRUDE*	46.02	10.90	25.83	87.98
Retail-terminal margin (cents per gallon) *MARGIN*	10.99	5.41	−8.91	34.90
Gasoline volume (gallons × 10^5/day) *VOLUME*	6.56	6.54	.38	38.11
Normalized gasoline volume *NVOLUME*	1.004	.093	.570	1.617
Expected normalized gasoline volume *EXP NVOLUME*	1.004	.086	.633	1.402
Expected terminal price (cents per gallon) *EXP TERMINAL*	62.05	11.09	36.45	112.77

station operator, but the price for this transaction is not publicly available. Similarly, for the gasoline sold at stations supplied directly by a refiner, there is no additional wholesale transaction if the station is owned and operated by the refiner. At the independently operated stations they supply, refiners sell gasoline to the station at the "dealer tankwagon price" (DTW). The DTW price is posted at the terminal. Less than 30% of the gasoline sold in the United States is sold at these refiner-supplied, but independently operated, stations (see U.S. Department of Energy (1988)).

For our purposes, the terminal price is the best proxy for marginal cost. For jobbers, it is the only observable wholesale price. It is also the opportunity cost of gasoline sold through refiner-supplied stations. Finally, unlike the DTW price, the terminal price is less subject to discounting off the posted price, according to people familiar with the industry. Refiners and station operators report that there is widespread discounting off posted DTW prices.[8] Unobservable, systematic discounts would be particularly troublesome here because their depth and prevalence probably varies with the refiner's perception of profitability in a given market at a particular time.

Because we use terminal price as the input cost, the margin we use for the empirical work is potentially affected by both the supplier (jobber or refiner) and the station operator. At stations operated by the supplier, the margin is set by the supplier alone. At independent stations, the retail price is set by the station operator but may also be affected by the supplier's use of quantity forcing and nonlinear pricing in the delivered price.[9] Without future information, it is impossible to determine whether the patterns we observe in retail margins reflect primarily the response of refiners and jobbers or of station operators.

[8] One person employed by a major refiner to set its delivered prices reported that a substantial portion of his division's effort was expended trying to infer rivals' actual delivered prices from their posted DTW prices.

[9] For a more complete discussion of how retail prices are set, see Shepard (1993) and Borenstein and Gilbert (1993).

For retail prices (*RETAIL*) we use the average price for unleaded, 87 octane, self-service gasoline in each of 43 cities. These averages are reported by Lundberg Survey based on a monthly survey of gasoline stations in each city.[10] Lundberg reports the survey date and the average price in the city for that date. The surveys always occur on a Friday, and the stations sampled change infrequently. The prices have been adjusted to exclude all sales and excise taxes. As is the case for all prices used in the study, retail prices are in nominal dollars.[11]

The terminal price (*TERMINAL*) is the average of all branded, unleaded, 87 octane terminal prices posted at the city terminal as reported by Lundberg. Because Lundberg reports average terminal prices for each Friday, we can match the average terminal and retail prices by date and city. The retail margin (*MARGIN*) is simply *RETAIL* − *TERMINAL* for a given city at a point in time.

Figure 1 shows the monthly movement in retail, terminal, and crude oil prices for 1986–1991.[12] As is clear in the graph, a large share of the volatility in terminal prices (and therefore in retail prices) is the result of shocks to the price of crude oil. The effect of the Gulf War in 1990–1991, for example, is quite clear in the data, as is the waning effect of the OPEC cartel at the beginning of the sample period. Although this is less clear in the figure, terminal prices respond with a lag to shocks to crude oil. Borenstein and Shepard (1996) estimate that approximately two-thirds of the eventual passthrough of crude price changes occurs in the first two weeks following the shock. The lags imply that future terminal prices will be influenced by both current and past crude price movements. As a result, decisionmakers have information from which to form time-varying expectations about future terminal prices. In the econometric model, we exploit this relationship to estimate expected terminal prices.

The data on gasoline consumption (U.S. Federal Highway Administration) are the total retail volume of gasoline sold in each state in each month. These data are tied to payment of the federal excise tax on gasoline. For each city, we use the volume data for the state in which the city is located. In those states for which Lundberg provides data on more than one city, we have randomly selected a single city to avoid replicating the volume data.[13] We divide the monthly volume figures by the number of days in the month to get an average daily consumption series for each state.

Movements in the quantity of gasoline consumed per day, as shown in Figure 2, are much more regular than the movements in price.[14] Gasoline demand follows a clear seasonal pattern in the United States: national demand is higher in the summer months, with the annual peak in August and the trough in January. On average, the August peak consumption is 23% above the January trough. There is also long-run movement in consumption—annual volumes rise over the first half of the 1986–1991 period and then decline again—but this is swamped by the seasonal movement.

Because the volume data are cumulative monthly consumption, we have only average daily consumption for each month. The price data, in contrast, record the average price on a single Friday in the month. This raises the issue of how best to match the

[10] This seems to be the best data series on retail prices. Other sources, such as *Oil and Gas Journal*, estimate retail prices from wholesale prices.

[11] During our sample period, inflation was fairly stable at about 4%. The results are virtually unchanged if we deflate all prices by the consumer price index.

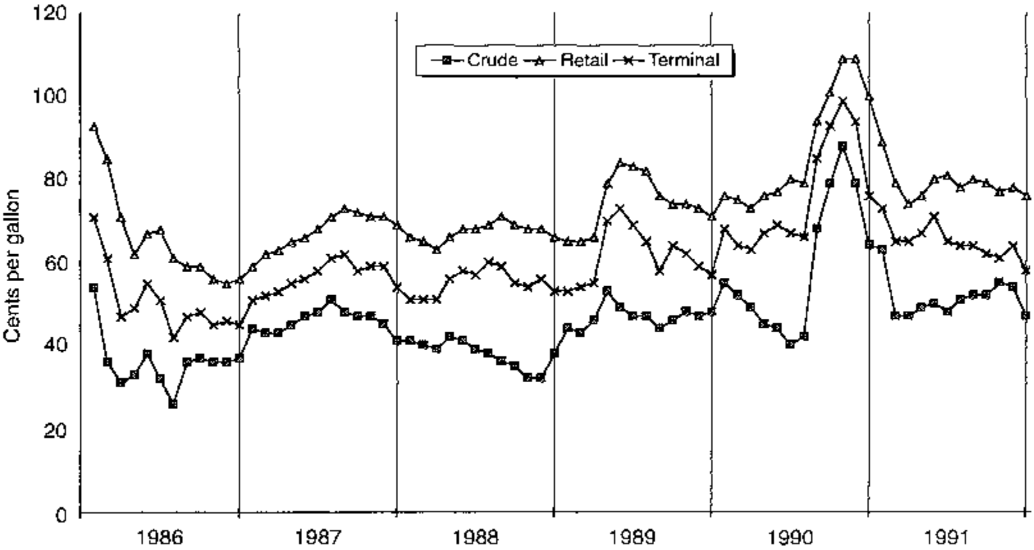
[12] The retail and terminal price series in Figure 1 are average monthly prices for the 43 cities included in our sample. The crude oil price series is the gulf coast spot price for West Texas Intermediate crude oil. The sample period is determined by the available retail gasoline price and volume data. We have gasoline price data for 1986–1992 and volume data for 1982–1991.

[13] We also estimated the empirical model using data on all (59) cities for which data are available. The results were substantively the same as those reported in Tables 2 and 3.

[14] The data for Figure 2 are average daily consumption for the 43 cities in our sample. They are not the interpolated series described below, but interpolation would have no substantive effect on the figure.

FIGURE 1

PRICES 1986–1991 (AVERAGE CITY PRICES)



price on a given Friday with the average daily volume series. The daily volume for the first Friday in April, for example, is probably best approximated by some combination of the average daily volume in March and in April. We therefore use a linear interpolation approach to construct a weighted average of the daily volumes in adjacent months.[15] The weighting scheme assigns a weight of one to the jth month's average daily volume if the observation happened to record prices on the middle day of that month. The weight on the current month declines linearly with movement away from the middle day. If, for example, prices were recorded for the first day of the month, the weighting scheme would apply a weight of about one-half to the average daily volumes of months j and $j - 1$.[16] The interpolated daily average is reported in Table 1 as *VOLUME.*

Finally, we divide the interpolated daily average in each state by its state mean over the sample period. The resulting normalized volume series (*NVOLUME*) expresses consumption as a proportion of the state's average daily consumption. This normalization removes the cross-sectional variation in daily volume introduced by variations in state populations: over 99% of the variance in *VOLUME* is accounted for by differences between state means. States with higher means also have more variation around the mean in absolute magnitude. If the effect of volume on margin were estimated using absolute deviations from state means in a standard fixed-effects formulation, the estimates would be heavily influenced by differences in state populations. *NVOLUME* removes this variation, leaving us with a series in which the remaining cross-sectional

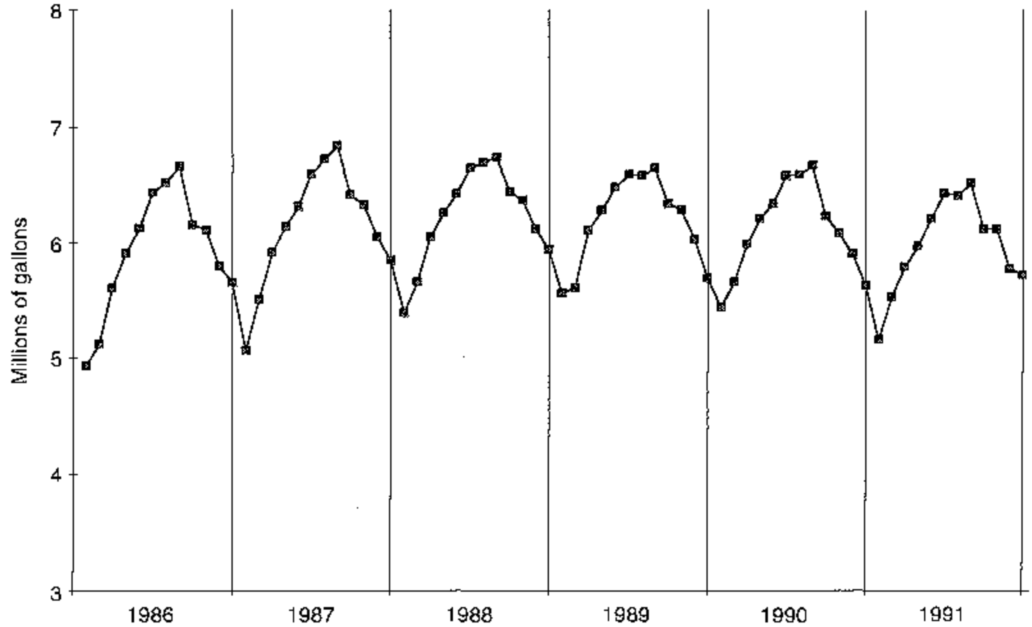
[15] In unreported estimates, we used a cubic rather than a linear spline. We have also used the reported-month daily volume with no interpolation. The results are robust to the interpolation technique.

[16] More precisely, if v_j is the average daily volume for month j, which has n_j days, and d_{ij} is the day on which retail price is observed, then the interpolated volume on the day price is observed, $\hat{v}_{ij}$, is given by

$$\hat{v}_{ij} = \begin{cases} v_j\left(\dfrac{1}{2} + \dfrac{d_{ij}}{n_j}\right) + v_{j-1}\left(\dfrac{1}{2} - \dfrac{d_{ij}}{n_j}\right) & \text{if } d_{ij} \le \dfrac{n_j}{2}; \\ v_j\left(\dfrac{3}{2} - \dfrac{d_{ij}}{n_j}\right) + v_{j+1}\left(\dfrac{d_{ij}}{n_j} + \dfrac{1}{2}\right) & \text{if } d_{ij} \ge \dfrac{n_j}{2}. \end{cases}$$

FIGURE 2

DAILY VOLUME 1986–1991 (STATE AVERAGES)



variation reflects temporal differences in state-level demand patterns. Using *NVOLUME* also provides us with a better variable for station-level demand. A 20% change in California daily volume represents (approximately) a 20% change at the observed California stations. This change is comparable to the change at Vermont stations that is represented by a 20% change in Vermont demand.

4. Estimation issues

■ A number of estimation issues must be addressed to develop consistent estimates of the effect of expected future demand and cost on current margins. As we argue below, the correct procedure involves estimating expected future demand and cost, allowing for lagged responses of both retail and wholesale prices to crude oil shocks, correcting for heteroskedastic and correlated errors, and instrumenting for endogenous variables. Before addressing all of these issues, however, we take a simple "first cut" at the data to explore the basic relationships. These preliminary estimates (reported in Table 2) should be viewed only as descriptive statistics and as a motivation for the empirical models developed later.

□ **Data exploration.** We begin by running an ordinary least squares regression of margin on current volume, next-period volume, current terminal price, next-period terminal price, and a set of time and city fixed effects. Next-period volume ($NVOLUME_{t+1}$) and terminal price ($TERMINAL_{t+1}$) are included as proxies for the expected future values that the theory predicts will influence current margins. There are two concerns about these proxies. First, it is expected values, not actual values, that should affect current margin decisions. We address this problem in subsequent estimations.

Second, we have chosen to represent the future by only the next period, a simplification retained throughout our analysis. In the theory, the number of empirically relevant future periods depends on the length of the punishment period and the discount

Table 2 Preliminary Regressions Dependent Variable: $MARGIN_t$

	OLS	OLS-AR1	OLS	OLS-AR1
$NVOLUME_t$	3.777	1.426	3.685	1.277
	(1.545)	(1.031)	(1.440)	(1.016)
$NVOLUME_{t+1}$	1.645	1.844	2.826	2.301
	(1.597)	(1.291)	(1.436)	(1.196)
$TERMINAL_t$	−.262	−.431	−.093	−.134
	(.021)	(.020)	(.022)	(.025)
$TERMINAL_{t+1}$	−.028	−.066	−.023	−.047
	(.021)	(.018)	(.020)	(.017)
$\Delta TERMINAL_t$			−.323	−.284
			(.021)	(.018)
ρ		.520		.558
		(.016)		(.015)
Observations	2,873	2,873	2,873	2,873
R^2	.652	.636	.686	.684

Notes: Fixed time and city effects not reported. White's heteroskedastic-consistent standard errors in parentheses.

rate. The theoretical models of tacit collusion leave open the length of the punishment period. Both the Rotemberg-Saloner and Haltiwanger-Harrington models use infinite punishments to simplify exposition, but their results do not depend on infinite punishments. If punishments are infinite, all future periods are relevant, and discounting makes the near-term periods relatively more important than later periods. If punishments are of finite duration, the relative importance of near-term periods will be greater. In the absence of any information about how long the punishment period might be in these markets, we include only next-period values. If the punishment is longer than one month, we will be measuring the relevant variable with error, and our estimate of the magnitude of the effect is likely to be biased downward.

Because the volume series display correlation over time, we include current volume ($NVOLUME_t$) among the regressors in this simple regression. Omitting current volume would load the effect of both current and future volume on the coefficient on next period's volume. Current terminal price is included to allow margin to respond to the level of the primary input price.[17] We also include a set of 72 monthly and 43 city fixed effects. As is clear from the plot in Figure 1, there are periods in which the price series display unusually large movements. Removing the mean effect for each of the months in the sample ensures that the results are not driven by these events. Approximately 30% of the variance in margin is accounted for by these time fixed effects. An additional 31% of the variance is explained by city fixed effects. All the regressions reported here remove time and city fixed effects.[18]

[17] Current terminal price is already in the dependent variable. We include it on the right-hand side to allow an unrestricted response of margins to current terminal prices. Omitting $TERMINAL_t$ from the regressors would be equivalent to imposing full (i.e., one-for-one) and immediate passthrough of terminal price changes into retail prices.

[18] The regressions presented here and below rely on the retail, terminal, and crude oil price variables being cointegrated. Carrying out the cointegration test suggested by Greene (1993) yields the following t-statistics, each of which strongly rejects the hypothesis that the variables are not cointegrated (using the Dickey-Fuller tables): retail-terminal, −25.31; retail-crude, −19.95; terminal-crude, −22.93.

The resulting coefficient estimates are presented in the first column of Table 2 and are roughly consistent with the predictions of the theoretical models. Conditional on contemporaneous volume and terminal price, current margins are a declining function of next period's terminal price and an increasing function of next period's volume. The estimated coefficient for the effect of next period's volume is, however, quite noisy.

The residuals from this regression display first-order autocorrelation. Given the evidence from prior studies that retail prices are affected by lagged terminal prices, this is not surprising. Correcting the estimates for an AR1 process (see column 2) does not affect the conclusion about the effect of future values on current margins. It does, however, have a large effect on the coefficient on current terminal. The magnitude of the change in this coefficient is about eight times its standard error.

The effect of the AR1 correction on the $TERMINAL_t$ coefficient suggests that it would be appropriate to include lagged values of terminal prices in the regression. We take the first step in this direction by adding $\Delta TERMINAL_t = TERMINAL_t - TERMINAL_{t-1}$ to the regressors. Including this simple lag structure does not eliminate the AR1 process, but (as reported in the third and fourth columns of Table 2) the coefficient estimates are now more stable. The positive effect of next period's volume and the negative effect of next period's terminal price on current margin persist in this specification.

These descriptive regressions suggest that the estimates from a more careful procedure might uncover effects consistent with the theory of tacit collusion. We begin this effort by describing the procedure we use to construct variables for expected volume and terminal price.

□ **The empirical model.** The results from preliminary regressions suggest that the specification of the lag structure of retail price response might have important effects on the results. We therefore estimate two different models of the lag process. The simpler specification is given by[19]

$$\begin{aligned} MARGIN_{it} = \alpha_1 NVOLUME_{it} + \alpha_2 EXP\ NVOLUME_{it+1} + \alpha_3 TERMINAL_{it} \\ + \alpha_4 EXP\ TERMINAL_{it+1} + \alpha_5 \Delta TERMINAL_{it} + \epsilon_{it}, \end{aligned} \tag{1}$$

where i indexes cities, $EXP\ NVOLUME_{it+1}$ is expected next-period volume, and $EXP\ TERMINAL_{it+1}$ is expected next-period terminal price.

While equation (1) is parsimonious, it does not allow for the complex lag structure in the response of gasoline prices that has been found in prior research. (See, for example, Karrenbrock (1991), Bacon (1991), U.S. General Accounting Office (1993), Borenstein, Cameron, and Gilbert (1995), and Borenstein and Shepard (1996).) These studies indicate that the response of retail prices may involve lags longer than one month and may be asymmetric with respect to terminal price increases and decreases. In addition to the lagged response to terminal price changes, retail price changes display persistence. One common effect of persistence is some form of autocorrelation in the error structure of price equations. More generally, retail price changes are a function of lagged changes in retail prices.

A standard approach to estimation in the presence of lagged responses is to model price as a vector autoregression (VAR) process in which the current change in price is a function of past shocks and an error-correction term that takes into account the underlying structure relationship. In our context, this means modelling the change in retail price, and therefore margins, as a function of contemporaneous and lagged

[19] In this and the following estimation equations, we suppress the constant term, which is absorbed by city and time fixed effects.

changes in terminal prices, lagged changes in retail prices, and an error-correction term involving the one-period lags of retail and terminal prices:

$$
\begin{aligned}
MARGIN_{it} = {} & \alpha_1 NVOLUME_{it} + \alpha_2 EXP\ NVOLUME_{it+1} + \alpha_3 EXP\ TERMINAL_{it+1} \\
& + \beta_1 \Delta TERMINAL^{+}_{it} + \beta_2 \Delta TERMINAL^{+}_{it-1} + \beta_3 \Delta TERMINAL^{+}_{it-2} \\
& + \beta_4 \Delta TERMINAL^{-}_{it} + \beta_5 \Delta TERMINAL^{-}_{it-1} + \beta_6 \Delta TERMINAL^{-}_{it-2} \\
& + \beta_7 \Delta RETAIL^{+}_{it-1} + \beta_8 \Delta RETAIL^{+}_{it-2} + \beta_9 \Delta RETAIL^{-}_{it-1} \\
& + \beta_{10} \Delta RETAIL^{-}_{it-2} + \beta_{11} RETAIL_{it-1} + \beta_{12} TERMINAL_{it-1} + \epsilon_{it},
\end{aligned} \tag{2}
$$

where ΔX_{it} denotes $X_{it} - X_{it-1}$ and the + superscript denotes an increase and the − a decrease. The contemporaneous and lagged changes capture the transmission of shocks, while the error-correction term captures the tendency to revert gradually to the long-term equilibrium relationship between retail and terminal prices. The number of lag periods has been chosen to allow for longer-than-expected lags in response. In the spirit of imposing only a minimal structure on the data, both retail and terminal price changes have been decomposed to allow an asymmetric response to price increases and decreases.[20] The details of deriving this estimating equation from the standard VAR treatment of price changes are given in the Appendix.

Estimating either (1) or (2) requires constructing variables representing expected volume and terminal prices. To construct *EXP NVOLUME*, we assume that station operators and suppliers form their expectations of next-period demand based on current and past sales volumes observed in their markets. Volume changes are caused largely by seasonal shifts in consumption patterns and, to a much lesser extent, by variations in retail prices. Demand seasonality varies across geographic locations in both pattern and amplitude. Although the typical state sees its peak demand in August and lowest demand in January, August demand is below the state mean in Florida and January demand is above the state mean in Colorado and Hawaii. The state in our sample with the greatest average annual variation (Maine) has a peak-to-trough difference equal to 57% of mean volume for that state; the state with the smallest variation (Arizona) has a peak-to-trough difference equal to 7% of its mean. The average state has a peak-to-trough difference equal to 23% of its mean. To allow the data to reflect these variations, we predict volume on a state-by-state basis. Accordingly, the predicting equation for state i is

$$
\begin{aligned}
NVOLUME_{it} = {} & a_0 + a_1 NVOLUME_{it-1} + a_2 RETAIL_{it-1} \\
& + \sum_{j=2}^{12} \delta_j MONTH_j + a_3 TIME_t + a_4 TIME_t^2 + \epsilon_{it},
\end{aligned} \tag{3}
$$

where $MONTH_j$ is a monthly dummy variable. For nearly all cities, the R^2 of these predicting equations is between .80 and .95. *EXP NVOLUME* is the fitted values from estimating (3).

Our procedure uses future data on volume to estimate the coefficients in the predicting equation (and, similarly, we use future prices in the predicting equations for prices). This is, in principle, incorrect because it uses information the decisionmakers did not have when they formed their coefficient estimates. If, however, the coefficients

[20] Including more lags than may be required or allowing for asymmetries unnecessarily will not bias the estimate of other coefficients.

do not change over time, we would get similar results using only data available to them. To check for robustness to the predicting procedure, we also predicted volume using a rolling regression with volume data for 1982–1991. Neither the predicted values nor the coefficients in the final regressions were substantively affected.[21]

Station operators and suppliers form their expectations of next-period terminal prices based on observing the price of crude oil and the time structure of terminal prices in their market. Analogous to retail prices, terminal prices are a function of lagged and current crude oil prices and lagged terminal prices. Because we want to use all information available to decisionmakers, we use the full lag structure analogous to the VAR model in equation (2):

$$\begin{aligned} TERMINAL_{it} = {} & b_1 \Delta TERMINAL^{+}_{it-1} + b_2 \Delta TERMINAL^{+}_{it-2} + b_3 \Delta TERMINAL^{-}_{it-1} \\ & + b_4 \Delta TERMINAL^{-}_{it-2} + b_5 \Delta CRUDE^{+}_{it-1} + b_6 \Delta CRUDE^{+}_{it-2} \\ & + b_7 \Delta CRUDE^{-}_{it-1} + b_8 \Delta CRUDE^{-}_{it-2} + b_9 TERMINAL_{it-1} \\ & + b_{10} CRUDE_{it-1} + \sum_{j=1}^{12} \delta_j MONTH_j + \epsilon_{it}. \end{aligned} \tag{4}$$

We estimate this equation city by city because the response of terminal prices to common crude oil shocks displays substantial variation across cities. Some of the difference in responses comes from the distribution system (Spiller and Huang, 1986) and some from differences in the conduct of competitors at the terminal (Borenstein and Shepard, 1996).

The crude oil price we use is the gulf coast spot price for West Texas Intermediate crude as reported by *Dow Jones International Petroleum Report* and published in the *Wall Street Journal.*[22] In contrast to equation (2), contemporaneous change in *CRUDE* ($\Delta CRUDE_{it}$) is not included in (4), because the purpose of this estimation is to generate the best forecast of *TERMINAL* from information available during the previous period. The R^2 of these regressions varies across cities within the .30 to .60 range. Price forecasts are less accurate than volume forecasts because volume follows a strong seasonal pattern, while terminal prices are very sensitive to current crude oil prices, which approximately follow a random walk. The fitted values from these regressions are the expected terminal prices for the period.

Both equations (1) and (2) include endogenous variables. Contemporaneous volume and terminal price are both reasonably modelled as endogenous. Volume is directly a function of the retail price and, therefore, a function of retail margin. Because expected next-period volume is a function of current volume and retail price, *EXP NVOLUME* is also endogenous. Terminal price will be a function of retail price if the terminal markets for gasoline are not perfectly competitive so that refiners set the wholesale price to influence the retail price. Terminal prices also will be endogenous if the market is competitive but refiners have an upward-sloping supply curve. Since *EXP TERMINAL* is a function of current terminal price, it is also endogenous. As a result, OLS

[21] We use the approach represented in equation (3), because we do not have data on prices for the presample period. Without presample price data we cannot use the rolling regression approach for prices and have chosen to be consistent across predicting equations.

[22] Borenstein, Cameron, and Gilbert (1995) report that using futures prices rather than spot prices or using spot prices for other commonly traded crude oils does not affect the estimated relationship between terminal and crude oil prices.

estimates of the margin equations are biased.[23] We provide consistent estimates for these equations using a two-stage least-squares (2SLS) procedure.

We identify the effects of volume and terminal price in 2SLS estimation by recognizing that these variables respond differently across cities to seasons, time, and changes in crude oil prices. As noted previously, seasonal demand patterns differ across cities depending on the weather and the tourist orientation of the area. Over time, these variables have also changed at different rates across areas. The response of terminal prices to changes in crude oil prices differs depending on transportation costs, proximity to a major port, alternative energy sources, and nongasoline uses of oil in the region. Gasoline demand and terminal prices might also be affected by changes in local economic conditions. Thus, besides the included exogenous variables, the instrument set for 2SLS estimation of (1) and (2) includes 11 separate monthly dummy variables for each city ($MONTH_{ij}$, $i = 1, 43$, $j = 1, 11$), time and time-squared variables for each city ($TIME_i$ and $TIME_i^2$, $i = 1, 43$), seven variables for adjustment to current and past crude oil price changes for each city ($\Delta CRUDE_{it}^+$, $\Delta CRUDE_{it-1}^+$, $\Delta CRUDE_{it-2}^+$, $\Delta CRUDE_{it}^-$, $\Delta CRUDE_{it-1}^-$, $\Delta CRUDE_{it-2}^-$, and $CRUDE_{it,i} = i = 1, 42$), and the current and lagged unemployment rate for the state.[24]

The endogeneity problem for equation (1) raises an additional complication because there is serial correlation in the OLS residuals. As a result, the lagged values of terminal in $\Delta TERMINAL_t$ are endogenous and the standard AR1 correction procedure is invalid. To correct for the additional endogenous variables, we treat the lagged values as endogenous, add an additional lag to the *CRUDE* variables in the instrument set and implement an AR1 correction following Greene (1993). Durbin's test for autocorrelation in the presence of lagged dependent variables (Greene, 1993) indicates no serial correlation in equation (2).[25]

Although we use an appropriate procedure to produce consistent estimates in the presence of endogeneity, we expect that the bias in the OLS estimates would be quite small. The estimated equations are supply relations, and the primary source of bias (directly for volume variables and indirectly for terminal price variables) is from movements along the demand curve rather than shifts of the demand curve. Industry demand for gasoline, however, is quite inelastic in the short run, with estimates ranging from $-.08$ to $-.2$ (Dahl and Sterner, 1991). It is clear, then, that the exogenous shifts in demand account for a much greater variation in consumption than is attributable to price changes. In fact, the 2SLS estimates do not differ substantially from the OLS estimates, as we show below.

Finally, the errors for both equations are heteroskedastic. We expected that the variance of the residuals would be a function of both the average retail margin in a city and the volatility of crude oil prices.[26] This suspicion was confirmed by regressing the square of the residuals on the average margin for each city and the sum of the absolute value of the change in crude oil price over the last three months: both variables

[23] In each specification, endogeneity tests reject the exogeneity of some of the current or expected volume and terminal prices and reject the joint exogeneity of all current and expected volume and terminal prices.

[24] Along with the city fixed effect for each city, these instruments use 23 degrees of freedom for each city. On average, there are 66 observations per city.

[25] The test statistic, distributed asymptotically normal, is below 1.0 for both the OLS and the 2SLS estimates of (2).

[26] It is natural to think that larger changes in crude oil prices would be associated with higher variance in the errors. The belief that higher margin observations would have lower variance comes from Borenstein and Shepard (1995), where we show that the adjustment rate of a city's terminal prices to crude oil shocks is a declining function of the city's average retail margin.

had t-statistics above 4. The heteroskedasticity correction was implemented by weighting the data by the inverse of the predicted values from this regression, as suggested by Greene (1993).[27]

5. Results

■ The OLS and two-stage least-squares estimates from equations (1) and (2) are reported in Table 3. In all specifications, the coefficient on expected volume is positive, as the Rotemberg-Saloner hypothesis predicts, and significantly different from zero at the 5% significance level or better. Using the coefficients from 2SLS estimation of equation (2), the elasticity of margin with respect to expected next-period volume is about .38 at mean margin and volume. The estimated effect of an expected change in terminal price is also consistent with the Rotemberg-Saloner hypothesis and is statistically significant at the 1% level. The elasticity of margin with respect to expected next-period terminal price is about −.36 at mean margin and terminal price, using the 2SLS results from (2).

In Figure 3 we have graphed the effects of volume changes on margins implied by the estimated two-stage coefficients on current and expected next-period volume from equation (2). The figure shows the national annual pattern for normalized volume (*NVOLUME*) and the estimated margin, normalized by mean margin. Because the changes in margin induced by changes in terminal and crude prices have been removed from the data graphed in Figure 3, it isolates the response of margins to changes in demand.[28] Movements in volume are estimated to cause margins to vary from about 95% of average margin in December to about 104% in June and July. The margin pattern in Figure 3 is consistent with the prediction that margins will be higher when near-term volume is increasing, holding current demand constant. For example, national average demand is approximately equal in April and October, but the increase in demand in May compared to the decline in demand in November causes estimated margins to be higher in April than in October. The estimates also imply that changes in the retail margin lead changes in volume over the demand cycle. Margin is positively correlated with volume, but it declines before the August demand peak in anticipation of the September volume decline, and rises before the January trough in anticipation of the February volume increase. This pattern is consistent with simulations conducted by Haltiwanger and Harrington that show a tendency for collusive price to drop (rise) before the demand peak (trough) even if price is generally procyclical.

The estimated effect of volume implies about a 9% variation in margins, or about one cent per gallon, over the seasonal cycle. This effect is not large in absolute magnitude, but its importance to tacitly colluding firms depends on the size of the operating margin. The margin we observe (retail price minus terminal price) averages 10.99 cents in our sample, but this overstates the true margin (retail price minus marginal cost), because true marginal cost is greater than the wholesale cost of gasoline. Delivery cost, for instance, is probably two to three cents per gallon,[29] reducing the economic margin to no more than about eight cents. Some of the labor costs are variable on a month-to-month basis for most stations, which would further reduce the economic margin.

[27] In unreported regressions, we correct the standard errors using White's procedure. The coefficient pattern is largely unaffected by this change in procedure. However, since White's correction gives consistent estimates of the standard errors, but does not improve the efficiency of the coefficient estimates, we report the results from this direct correction for heteroskedasticity.

[28] The pattern evident in Figure 3 is not apparent in a simple graph of margin on volume because a very large share of the variance in margins is the result of movements in crude oil prices. These effects must be removed to uncover the demand effects.

[29] DTW prices exceed terminal prices by about this amount.

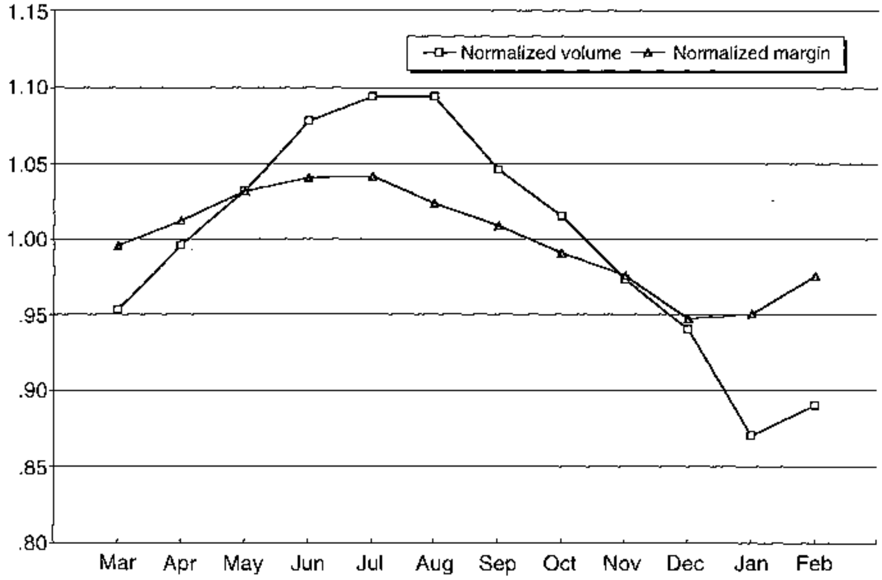
Table 3 Estimation of Equations (1) and (2)
Dependent Variable: $MARGIN_t$

	Equation (1)		Equation (2)	
	OLS	2SLS	OLS	2SLS
$NVOLUME_t$	.952 (1.223)	3.009 (1.600)	−.335 (1.211)	.690 (1.493)
$EXP\ NVOLUME_{t+1}$	3.220 (1.560)	3.927 (1.636)	4.314 (1.464)	4.158 (1.533)
$TERMINAL_t$	−.149 (.023)	−.058 (.035)		
$EXP\ TERMINAL_{t+1}$	−.062 (.016)	−.029 (.019)	−.061 (.017)	−.063 (.022)
$\Delta TERMINAL_t$	−.271 (.015)	−.291 (.024)		
$\Delta TERMINAL_t^+$			−.268 (.026)	−.179 (.043)
$\Delta TERMINAL_{t-1}^+$			.154 (.029)	.162 (.028)
$\Delta TERMINAL_{t-2}^+$			.073 (.026)	.078 (.027)
$\Delta TERMINAL_t^-$			−.532 (.025)	−.553 (.050)
$\Delta TERMINAL_{t-1}^-$			.174 (.026)	.162 (.028)
$\Delta TERMINAL_{t-2}^-$			.074 (.025)	.074 (.025)
$\Delta RETAIL_{t-1}^+$			−.038 (.027)	−.040 (.027)
$\Delta RETAIL_{t-2}^+$			−.056 (.026)	−.052 (.026)
$\Delta RETAIL_{t-1}^-$			−.047 (.027)	−.044 (.027)
$\Delta RETAIL_{t-2}^-$			.023 (.023)	.017 (.024)
$RETAIL_{t-1}$			.610 (.018)	.606 (.018)
$TERMINAL_{t-1}$			−.605 (.027)	−.582 (.032)
ρ	.547 (.015)	.564 (.015)		
Observations	2,812	2,812	2,812	2,812
R^2	.634		.799	

Notes: Fixed time and city effects not reported. Heteroskedasticity–corrected asymptotic standard errors in parentheses. Equation (1) corrected for first-order serial correlation.

FIGURE 3

ESTIMATED EFFECT OF VOLUME ON MARGIN



Overall, the proportional effect of volume changes on economic margins is probably at least 15%.

The estimated effect of $EXP\ TERMINAL_{t+1}$ is most easily interpreted by considering the magnitudes of expected month-to-month changes in terminal price. The month-to-month expected change in terminal price (i.e., the expected terminal price in period $t + 1$ minus the actual terminal price in period t) in a city has a mean of about zero and a standard deviation of 4.84 cents. It terminal price is expected to increase by an amount equal to one standard deviation of the average expected change, current margin will be reduced by .30 cents compared to an expected change of zero.

The remaining estimates of the effects of lagged changes and levels of terminal and retail prices are consistent with previous research. The cumulative response function implied by these estimates exhibits asymmetric adjustment to terminal price changes: increases are passed through more quickly than decreases. This is consistent with the response found by Karrenbrock (1991), and the size of the asymmetry is very close to that found by Borenstein, Cameron, and Gilbert (1995) using national rather than city-level price data. The estimates also imply a passthrough rate that is not statistically distinguishable from 100%.[30]

While the estimates in Table 3 allow for fixed city and time effects, we have maintained the assumption that the coefficients in equations (1) and (2) are equal across cities. This is a parsimonious approach, and we have no *a priori* reason to think that these coefficients would differ systematically. Nonetheless, this assumption imposes 210 restrictions on (1) and 630 restrictions on (2). It would not be surprising to find that standard tests reject pooling. Indeed, an F-test rejects the pooling in each equation. To investigate whether pooling has affected our results, we now relax these restrictions. Since there is no obvious way of grouping cities to test for meaningful subgroups, we estimated each equation allowing parameter values to differ for each of the 43 cities.

[30] See Borenstein, Cameron, and Gilbert (1995) for a description of the method used to construct the cumulative response function and to calculate the passthrough rate.

This approach uses up many degrees of freedom (leaving only about 50 degrees of freedom per city for estimation of (2)), and the estimates will be noisier than the restricted estimates. Nonetheless, removing all the restrictions has the advantage that the estimated coefficients cannot be affected by illegitimate pooling.[31] The results for the expectation variables from unrestricted estimation of (2) are reported in Table 4 and confirm the pooled estimates reported in Table 3.

The first line of the table reports the means of the 43 OLS and 2SLS estimates of the coefficients on $EXP\ NVOLUME_{t+1}$ and $EXP\ TERMINAL_{t+1}$ parameters and the standard error of those means. The mean estimated effect of future expected volume is of the same sign as the restricted estimate reported in Table 3, but it is substantially larger. Similarly, the mean estimated effect of future expected terminal price is of the same sign as the restricted estimate in Table 3. Despite the loss in degrees of freedom, the means are statistically different from zero, except for the effect of future terminal price in the 2SLS estimation.

Table 4 also reports two additional tests of whether the pattern in the city-by-city parameter estimates is consistent with the restricted estimates in Table 3. The first is the Z-statistic. This statistic is the sum of the 43 t-statistics and is distributed asymptotically normal with mean zero and standard deviation of $\sqrt{43}$ under the null hypothesis that the sum of the t-statistics equals zero. The Z-statistic, which is commonly used in stock market event studies, is less heavily influenced by individual estimates with large standard errors than is the estimate of the mean. Below each of the Z-statistics, we show the level at which the statistic is significantly different from zero using a two-tailed test.[32] The Z-statistics are all of the same sign as the coefficients in Table 3, and all are significantly different from zero at the 10% level or greater.

The final test is a simple binomial distribution: a count of the number of the 43 parameter estimates that have the expected sign. This approach has low power because it ignores considerable available information, but it is least likely to be influenced by individual cities. The counts shown in Table 4 indicate the high noise level in the individual city estimates, though the majority in each case are still of the hypothesized sign. In the 2SLS estimation, 60% of the coefficients for $EXP\ VOLUME_{t+1}$ are of the expected sign and 60% of the coefficients for $EXP\ TERMINAL_{t+1}$ are of the expected sign. Below each count we report the significance level (two-tailed) at which the frequency of "correct" signs is significantly different from the null of .5. Only the results from 2SLS estimation of the $EXP\ VOLUME_{t+1}$ coefficient are distinguishable from $p = .5$ using standard significance levels.[33]

These results are consistent with the predictions of the tacit collusion models and inconsistent with standard noncollusive models where current margin is not a function of future expected demand or cost conditions. There are, however, two types of noncollusive models that can produce a linkage: models involving intertemporal switching costs and models of inventory behavior. We discuss these alternatives next.

---

[31] For the unrestricted estimation, we continue to remove the common 72 time fixed effects. Thus for each city, every variable is measured as the deviation from its mean over all cities at the same point in time. The heteroskedasticity correction in each case was based on a common estimate of the effect of crude volatility on the variance in all cities. For the unrestricted estimation of (1), we estimated a common parameter of serial correlation ρ.

[32] These and all of the significance tests in this table are carried out under the assumption that the city-by-city parameter estimates are uncorrelated across cities. This is valid because the 72 fixed time effects remove common shocks.

[33] The (unreported) results for the unrestricted estimation of equation (1) are similar to those reported in Table 4 for (2). The results for unrestricted estimation of (1) have the expected sign in all cases, are somewhat less significant than for (2) in the mean coefficients, about the same significance for the z-statistics, and more significant in the binomial tests.

Table 4 Summary of Unconstrained Estimation of Equation (2)
Dependent Variable: $MARGIN_t$

	$EXP\ NVOLUME_{t+1}$		$EXP\ TERMINAL_{t+1}$	
	OLS	2SLS	OLS	2SLS
Mean parameter estimate	9.844	10.154	−.053	−.048
(standard error of mean)	(2.623)	(4.159)	(.018)	(.035)
Z-statistic	17.90	11.89	−18.62	−13.74
(P-value vs. H_0: $Z = 0$)	(.01)	(.07)	(.01)	(.04)
Binomial distribution (of 43)	25 > 0	26 > 0	28 < 0	26 < 0
(P-value vs. $p = .5$)	(.36)	(.22)	(.07)	(.22)

6. Possible alternative explanations for the effect of future demand and cost

■ Repeat purchases and consumer loyalty can create an incentive for firms to take actions this period in anticipation of future conditions. Klemperer (1987) presents a model in which consumer loyalty leads firms to choose prices to affect their future demand. Suppose that some kind of switching cost causes consumers to prefer buying next period at the station at which they buy this period. Then stations (and their suppliers) want to attract buyers in this period in order to sell to them next period. If industry demand is expected to increase next period, the incentive to attract customers this period is higher than if demand is expected to decline. Or, holding demand constant, suppose the terminal price were expected to rise. This would make a sale tomorrow less profitable and induce higher margins today. Thus, while this model does make this-period choices a function of expected next-period values, it implies that margins will be lower when demand is increasing and higher when cost is increasing. This is the reverse of our empirical findings.[34]

A link between current margins and anticipated market conditions could also be the result of firms responding to expected changes by altering their inventory levels. Inventories play a very important role in upstream petroleum markets, but these effects will be reflected in the terminal price and will not affect our analysis of the retail margin. For inventory effects to be important in determining retail margins, inventories must be held by station suppliers or station operators. Station operators do not hold significant inventory; they accept new deliveries every few days.[35] Any inventory effects, then, must come from independent wholesalers. These jobbers have inventory capacity and are known to hold inventory against future demand and cost changes.

In the standard inventory models, a firm that expects demand to increase (decline) next period will increase (reduce) its inventory holdings this period (see, for example, Pindyck (1994) and Thurman (1988)). If, for example, jobbers increase their purchasing today in anticipation of higher demand tomorrow, this period's terminal prices might be driven up. If they do not fully pass this increase into the prices they charge stations or into the retail prices they directly control, then current margin will be reduced. This

[34] The switching-cost explanation assumes that consumers are indeed loyal and that the increase in demand is generated by the same consumers in the same geographic market. If switching costs are low or a significant share of the seasonal increase in demand is generated by leisure travelers buying gasoline away from their usual locations, then there will be no intertemporal response due to consumer loyalty.

[35] The National Petroleum Council estimates that total inventory capacity at the 170,000 retail gasoline outlets in the United States in 1989 was 3.49 billion gallons. This implies an average capacity of approximately 20,000 gallons per station. Stations sell about 200,000 gallons per month on average, so they must accept delivery more than ten times per month on average.

implies that higher demand tomorrow would lead to lower margins today. This prediction has the opposite sign to the one we find for future expected volume.[36]

Inventory might also link current margins and expected future input prices. If, for example, jobbers believe that terminal prices will increase next period, they have an incentive to increase inventories this period. Holding inventories against expected input price fluctuation will have no effect on current margins, however, because the (opportunity) cost of marginal sales this period is the current terminal price. As long as wholesalers can purchase at the current terminal price, holding inventories may increase profits, but it should have no effect on the optimal retail margin for the current period.

It is possible, however, that transaction costs lead wholesalers to purchase only periodically,[37] so that the relevant opportunity cost is the terminal price they expect for the next transaction date. In that case, higher expected terminal prices could cause suppliers to increase the current price they charge to station operators who will, in turn, pass at least some of the increase into the retail price. This would lead to higher observed margins when wholesale prices are expected to increase. The true economic margin would be today's retail price minus the expected future terminal price, but the observed margin would be today's retail price minus today's terminal price. Inventories, then, might lead to a positive estimated effect of expected terminal price changes on current margins, the opposite of the effect that we find.

7. Conclusion

■ Using a panel of data on retail gasoline margins in 43 cities over 6 years, we have found evidence consistent with tacitly collusive pricing. Contrary to the predictions of standard noncooperative models, we find that margins respond to anticipated changes in demand and input prices. Furthermore, margins respond with the signs predicted by the Haltiwanger-Harrington extension of the Rotemberg-Saloner model to markets with deterministic demand cycles. *Ceteris paribus,* we find lower margins when demand is expected to decline next period than when it is expected to increase. The evidence for tacit collusion is particularly strong because we also find that margins respond to input cost changes in the predicted direction—margins are higher when wholesale prices are expected to decline next period than when they are expected to increase next period.

The results are somewhat surprising because the structure of the retail gasoline industry is not the tight oligopoly setting supposed by the formal models or investigated in prior empirical studies of collusion. They are somewhat less surprising when we recognize that evidence supporting tacitly collusive pricing is not evidence that the firms are able to set prices at, or even close to, the monopoly level. Indeed, given the inelastic industry demand for gasoline, it is quite clear that the actual margins fall far short of the monopoly level in this industry. The magnitude of the effect of tacit collusion on margins and the low elasticity of gasoline demand also imply that it will have little effect on welfare. Still, a simple structural analysis of retail gasoline markets might lead one to expect that any price above the one-shot Nash equilibrium level would be unsustainable. The dynamic pattern we observe in retail margins is inconsistent with this conclusion.

The dynamic pattern is consistent with formal models that assume firms are able to make sophisticated calculations to achieve the highest sustainable collusive price. In this

[36] Since we control for the lag structure in retail prices, we would uncover this inventory-induced linkage only if we have misspecified the lag structure. The point of the argument in the text is to make clear that the results we do find cannot be explained by inventory effects, not to argue that inventory effects don't exist.

[37] This might be the case if, for example, there were a fixed cost to making a purchase.

abstract world, all available information is used, price wars never happen, and the collusive constraint is always binding. This cannot be taken as a serious description of the world of gasoline retailing. It is plausible, however, that station operators, for example, recognize that reducing their own prices induces price cutting at rival stations. They also might reasonably expect that once price cutting begins it will take some time to return to higher margins even without a clearly articulated strategy for the length and severity of the retaliation. Recognizing this, each seller might very well be less willing to cut price if the resulting retaliation will occur when market conditions make it particularly costly. Tacit collusion of this sort can produce the dynamic pricing pattern we find.

Appendix

■ Equation (2) is a straightforward variation on standard VAR models. A VAR model of changes in retail price is

$$\begin{aligned} RETAIL_{it} - RETAIL_{it-1} = {} & \beta_1 \Delta TERMINAL_{it}^{+} + \beta_2 \Delta TERMINAL_{it-1}^{+} + \beta_3 \Delta TERMINAL_{it-2}^{+} \\ & + \beta_4 \Delta TERMINAL_{it}^{-} + \beta_5 \Delta TERMINAL_{it-1}^{-} + \beta_6 \Delta TERMINAL_{it-2}^{-} \\ & + \beta_7 \Delta RETAIL_{it-1}^{+} + \beta_8 \Delta RETAIL_{it-2}^{+} + \beta_9 \Delta RETAIL_{it-1}^{-} \\ & + \beta_{10} \Delta RETAIL_{it-2}^{-} + \beta_{11} RETAIL_{it-1} + \beta_{12} TERMINAL_{it-1} + \epsilon_{it}. \end{aligned} \tag{A1}$$

The terms involving lagged retail and terminal prices in levels are the error-correction term. To understand its role in the regression, take the equilibrium relationship between retail and terminal prices to be

$$RETAIL = \gamma_0 + \gamma_1 TERMINAL.$$

The extent to which prices deviate from equilibrium in period $t - 1$ would be

$$RETAIL_{t-1} - \gamma_0 - \gamma_1 TERMINAL_{t-1}.$$

If the next-period margin adjusts toward the equilibrium by some factor ρ $(0 > \rho > -1)$, the error-correction effect for period t would be $\rho(RETAIL_{t-1} - \gamma_0 - \gamma_1 TERMINAL_{t-1})$, which can be rewritten as

$$\rho RETAIL_{t-1} - \rho\gamma_0 - \rho\gamma_1 TERMINAL_{t-1}.$$

The first term requires including $RETAIL_{t-1}$, the second term is subsumed in the fixed effects, and the third term requires including $TERMINAL_{t-1}$.

Because we are interested in retail margins, we rearrange terms in this equation, adding $RETAIL_{it-1}$ and subtracting $TERMINAL_{it}$ from both sides to get current retail margin on the left-hand side:

$$\begin{aligned} MARGIN_{it} = {} & (\beta_1 - 1)\Delta TERMINAL_{it}^{+} + \beta_2 \Delta TERMINAL_{it-1}^{+} + \beta_3 \Delta TERMINAL_{it-2}^{+} \\ & + (\beta_4 - 1)\Delta TERMINAL_{it}^{-} + \beta_5 \Delta TERMINAL_{it-1}^{-} + \beta_6 \Delta TERMINAL_{it-2}^{-} \\ & + \beta_7 \Delta RETAIL_{it-1}^{+} + \beta_8 \Delta RETAIL_{it-2}^{+} + \beta_9 \Delta RETAIL_{it-1}^{-} + \beta_{10} \Delta RETAIL_{it-2}^{-} \\ & + (\beta_{11} + 1) RETAIL_{it-1} + (\beta_{12} - 1) TERMINAL_{it-1} + \epsilon_{it}. \end{aligned} \tag{A2}$$

Subtracting $TERMINAL_{it}$ from the right-hand side of (1) requires subtracting one from the coefficients on $\Delta TERMINAL_{it}^{+}$ and $\Delta TERMINAL_{it}^{-}$, each of which contains a positive $TERMINAL_{it}$. Because each also contains a negative $TERMINAL_{it-1}$ that has been removed, subtracting one from the coefficient on $TERMINAL_{it-1}$ reintroduces these terms. The equivalence is not exact because the coefficients are allowed to be asymmetric in the adjustment (Δ) terms, but not in the error-correction term ($TERMINAL_{it-1}$).

Equation (A2) characterizes the autoregressive structure of prices. Augmenting it to take into account the effect of current demand and of expected demand and terminal price produces the basic estimating equation reported in the body of the article as equation (2), where the parameter adjustments to get from (A1) to (A2) are implicit.

References

BACON, R.W. "Rockets and Feathers: The Asymmetric Speed of Adjustment of U.K. Retail Gasoline Prices to Cost Changes." *Energy Economics*, Vol. 13 (1991), pp. 211–218.

BORENSTEIN, S. "Selling Costs and Switching Costs: Explaining Retail Gasoline Margins." *RAND Journal of Economics*, Vol. 22 (1991), pp. 354–369.

——— AND GILBERT, R.J. "Uncle Sam at the Gas Pump: Causes and Consequences of Regulating Gasoline Distribution." *Regulation*, Vol. 16 (1993), pp. 63–75.

——— AND SHEPARD, A. "Dynamic Pricing in Retail Gasoline Markets." Mimeo, 1994.

——— AND ———. "Sticky Prices, Inventories, and Market Power in Wholesale Gasoline Markets." NBER Working Paper no. 5468, 1996.

———, CAMERON, A.C., AND GILBERT, R.J. "Do Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes?" Program on Workable Energy Regulation Working Paper no. PWP-001R, University of California, 1995.

BRESNAHAN, T.F. "Competition and Collusion in the American Automobile Oligopoly: The 1955 Price War." *Journal of Industrial Economics*, Vol. 35 (1987), pp. 457–482.

———. "Empirical Studies of Industries with Market Power." In R. Schmalensee and R.D. Willig, eds., *Handbook of Industrial Organization*. Amsterdam: North-Holland, 1989.

DAHL, C. AND STERNER, T. "Analyzing Gasoline Demand Elasticities: A Survey." *Energy Economics*, Vol. 13 (1991), pp. 203–210.

ELLISON, G. "Theories of Cartel Stability and the Joint Executive Committee." *RAND Journal of Economics*, Vol. 25 (1994), pp. 37–57.

GREENE, W.H. *Econometric Analysis*. 2d ed. New York: Macmillan, 1993.

HAJIVASSILIOU, V.A. "Testing Game-Theoretic Models of Price-fixing Behavior." Yale Cowles Foundation Discussion Paper no. 935, 1989.

HALTIWANGER, J. AND HARRINGTON, J.E., JR. "The Impact of Cyclical Demand Movements on Collusive Behavior." *RAND Journal of Economics*, Vol. 22 (1991), pp. 89–106.

HENLY, J., POTTER, S., AND TOWN, R. "Price Rigidity, the Firm, and the Market: Evidence from the Wholesale Gasoline Industry During the Iraqi Invasion of Kuwait." Mimeo, 1995.

KARRENBROCK, J.D. "The Behavior of Retail Gasoline Prices: Symmetric or Not?" *Federal Reserve Bank of St. Louis Review*, Vol. 73 (July/August, 1991), pp. 19–29.

KLEMPERER, P. "The Competitiveness of Markets with Switching Costs." *RAND Journal of Economics*, Vol. 18 (1987), pp. 137–150.

PINDYCK, R.S. "Inventory and the Short-Run Dynamics of Commodity Prices." *RAND Journal of Economics*, Vol. 25 (1994), pp. 141–159.

PORTER, R.H. "A Study of Cartel Stability: The Joint Executive Committee, 1880–1886." *Bell Journal of Economics*, Vol. 14 (1983), pp. 301–314.

———. "On the Incidence and Duration of Price Wars." *Journal of Industrial Economics*, Vol. 33 (1985), pp. 415–426.

——— AND ZONA, J.D. "Detection of Bid Rigging in Procurement Auctions." *Journal of Political Economy*, Vol. 101 (1993), pp. 518–538.

ROTEMBERG, J.J. AND SALONER, G. "A Supergame-Theoretic Model of Price Wars During Booms." *American Economic Review*, Vol. 76 (1986), pp. 390–407.

SHEPARD, A. "Price Discrimination and Retail Configuration." *Journal of Political Economy*, Vol. 99 (1991), pp. 30–53.

———. "Contractual Form, Retail Price, and Asset Characteristics in Gasoline Retailing." *RAND Journal of Economics*, Vol. 24 (1993), pp. 58–77.

SLADE, M.E. "Conjectures, Firm Characteristics, and Market Structure." *International Journal of Industrial Organization*, Vol. 4 (1986), pp. 347–369.

———. "Interfirm Rivalry in a Repeated Game: An Empirical Test of Tacit Collusion." *Journal of Industrial Economics*, Vol. 35 (1987), pp. 499–516.

SPILLER, P.T. AND HUANG, C.J. "On the Extent of the Market: Wholesale Gasoline in the Northeastern United States." *Journal of Industrial Economics*, Vol. 35 (1986), pp. 131–141.

TEMPLE, BARKER, AND SLOAN, INC. "Gasoline Marketing Practices in the 1980s: Structure, Practices, and Public Policy." Washington, D.C.: American Petroleum Institute, 1988.

THURMAN, W.N. "Speculative Carryover: An Empirical Examination of the U.S. Refined Copper Market." *RAND Journal of Economics*, Vol. 19 (1988), pp. 420–437.

U.S. DEPARTMENT OF ENERGY, ENERGY INFORMATION ADMINISTRATION. *Petroleum Marketing Monthly*, September 1988.

———, NATIONAL PETROLEUM COUNCIL. "Petroleum Storage and Transportation, Volume IV, Petroleum Inventories and Storage." Washington, D.C.: U.S. Government Printing Office, 1989.

U.S. Federal Highway Administration. *Highway Statistics.* Washington, D.C.: U.S. General Printing Office, 1986–1992.

U.S. General Accounting Office. *Energy Security and Policy: Analysis of the Pricing of Crude Oil and Petroleum Products, GAO/RCED-93-17.* Washington, D.C.: U.S. Government Printing Office, 1993.