Collinearity in Linear Structural Models of Market Power

Jeffrey M. Perloff * Edward Z. Shen †

*University of California, Berkeley, and Giannini Foundation
†Wells Fargo Bank

Collinearity in Linear Structural Models of Market Power

Abstract

The well-known structural model used to estimate market structure suffers from a severe collinearity problem if the marginal cost and demand equations are linear.

Collinearity in Linear Structural Models of Market Power

Jeffrey M. Perloff*
Edward Z. Shen**

December 2001

Abstract

The well-known structural model used to estimate market structure suffers from a severe collinearity problem if the marginal cost and demand equations are linear.

Key Words: market power; estimation

JEL classification: L13, C1

We thank Charles Hyde, Larry Karp, and Michael Ward for helpful discussions. The simulations were conducted using Shazam.

* Professor, Department of Agricultural and Resource Economics, and Member of the Giannini Foundation, University of California, Berkeley

** Risk Management Consultant, Wells Fargo Bank

Contact:
Jeffrey M. Perloff 510/642-9574 (o) 510/643-8911 (fax)
Department of Agricultural and Resource Economics
207 Giannini Hall
University of California
Berkeley, California 94720
perloff@are.Berkeley.Edu

Collinearity in Structural Models of Market Power

A structural model due to Just and Chern (1980) has been widely used to estimate market power based on market-level data.[1] Bresnahan (1982) used a simple linear structural model to illustrate the method and to demonstrate that identification problems can arise. The linear model has been estimated by many competent econometricians (two recent applications are Jans and Rosenbaum, 1996, and Deodhar and Sheldon, 1997). Unfortunately, this linear model is fundamentally flawed due to a previously unrecognized multicollinearity problem.

We discovered this problem when we tried to estimate a simulated linear model. The simulations demonstrated that a loglinear model could be reliably estimated; however, a linear model produced completely unreliable estimates due to severe multicollinearity problems.

In this paper, we demonstrate that an econometrician trying to estimate the linear model faces three very unattractive possibilities. First, if the true model is not linear, the estimates are biased (simulations in Hyde and Perloff, 1995, illustrate that misspecification biases in the estimates of market power may be severe). Second, if the true model is linear and the equations hold exactly, the variables are perfectly collinear so that the model cannot be estimated. Third, if the true linear model equations hold with errors, one can estimate the equations, but the estimated coefficients are likely to be highly unstable and unreliable due to nearly perfect collinearity.

---

[1] Over a hundred studies have used this approach. For a partial list, see Bresnahan (1989) or Carlton and Perloff (2000).

Structural Model of Market Power

In most studies based on industry-level data, the econometrician starts by assuming that the market consists of identical firms (or makes other similar assumptions), so that the "average" market power of these firms can be estimated. Typically, in this approach, the "true" model reflects the behavior of a single firm that may not be profit-maximizing.

Problems arise when both the marginal cost and demand curves are linear. Following Bresnahan (1982), suppose that the marginal cost curve is

$$MC = \eta + \alpha w + \beta r + \gamma Q + \varepsilon_c, \tag{1}$$

where w is the wage, r is the rental rate on capital, Q is market output, and ε_c is an error term.[2] Also following Bresnahan, the demand curve is

$$p = \phi_0 - \left[\phi_1 + \phi_2 Z\right] Q + \phi_3 Y + \varepsilon_d, \tag{2}$$

where p is price, Z is an exogenous variable (such as the price of a substitute, a proxy for taste changes, or income) that rotates the demand curve, Y is an additive exogenous variable, and ε_d is an error term. The slope of the demand curve is $p' \equiv dp/dQ = -[\phi_1 + \phi_2 Z]$.

[2] This functional form is used by Bresnahan (1982) and others. Notice that this specification implies that the cost function is not homogeneous if γ is nonzero.

Following Just and Chern (1980), Bresnahan (1982), and Lau (1982), we use a parameter λ to nest various market structures.[3] Specifically, we define an "effective" marginal revenue function as

$$MR(\lambda) = p + \lambda p' Q = p - \lambda[\phi_1 + \phi_2 Z]Q.$$

If $\lambda = 0$, marginal revenue equals price and the market is competitive; if $\lambda = 1$, marginal revenue equals the marginal revenue of a monopoly; if λ lies between 0 and 1, the degree of market power lies between that of monopoly and competition. For example, with *n* identical Cournot firms, λ equals $1/n$.

The optimality or equilibrium condition is that the industry sets its effective marginal revenue equal to its marginal cost, $MR(\lambda) = MC$, so $p = MC + \lambda[\phi_1 + \phi_2 Z]Q$, or

$$p = \eta + \alpha w + \beta r + [\lambda(\phi_1 + \phi_2 Z) + \gamma]Q + \varepsilon_c. \tag{3}$$

The econometrician simultaneously estimates Equations 2 and 3 to obtain an estimate of λ, the measure of market power. Bresnahan (1982) and Lau (1982) show that λ is identified in the linear model only if $\phi_2 \neq 0$ (the *ZQ* interaction term matters) or $\gamma = 0$ (marginal cost does not vary with output: constant returns to scale).

---

[3] Some researchers view λ as a conjectural variation, while others describe it as the outcome of an unknown game, where λ is a summary measure of the gap between *p* and *MC*. Bresnahan (1989) and Corts (1999) discuss these alternative interpretations.

Perfect Collinearity

Even if the linear model is correctly specified and identified, it has a fundamental problem of multicollinearity. As we show in the appendix, the six regressors in Equation 3, the constant, w, r, Y, ZQ, and Q, are perfectly collinear if the equations hold exactly ($\varepsilon_d = \varepsilon_c = 0$). To make this intuition clear, we consider the special case where Y, w, and r are irrelevant ($\phi_3 = \alpha = \beta = 0$) and marginal cost does not vary with output ($\gamma = 0$). We can solve for the equilibrium value of Q by substituting for p from Equation 2 into Equation 3:

$$Q = \frac{\phi_0 - \eta}{(1 + \lambda)(\phi_1 + \phi_2 Z)}.$$

We now show that the right-hand-side variables in the optimality equation, Q and ZQ, are perfectly collinear by demonstrating that the weighted sum of these two variables is a constant. Let the weight on Q be $\lambda\phi_1$ and the weight on ZQ be $\lambda\phi_2$, then

$$\lambda\phi_1 Q + \lambda\phi_2 ZQ = \lambda(\phi_1 + \phi_2 Z)Q = \lambda(\phi_1 + \phi_2 Z)\frac{\phi_0 - \eta}{(1 + \lambda)(\phi_1 + \phi_2 Z)}$$
$$= \frac{\lambda}{1 + \lambda}(\phi_0 - \eta)$$

Thus, when the demand and optimality equations hold exactly, the optimality equation suffers from perfect multicollinearity. This perfect multicollinearity creates a problem (Greene, 1990, p. 278) that is more fundamental than the one Bresnahan (1982) and Lau (1982) discuss where only λ cannot be identified. Here, due to perfect linear dependency, none of the coefficients can be estimated.

Nearly Perfect Collinearity

If the demand and optimality equations do not hold exactly ($\varepsilon_d \neq 0$ and $\varepsilon_c \neq 0$), one can mechanically estimate this system of equations, but the right-hand-side variables are nearly perfectly collinear, which creates the usual problems:

1. Coefficients may have large standard errors (low precision) even though they are jointly highly significant.

2. Coefficients may have the "wrong" sign or an implausible magnitude.

3. Estimates may be very sensitive to addition or deletion of a few observations or the deletion of an apparently insignificant variable.

We illustrate these multicollienarity problem using simulations. Tables 1 and 2 summarize a 1,000 replications of experiments with 50 observations each. We set $\alpha = \beta = \gamma = \delta_1 = \delta_2 = \eta = 1$, $\phi_3 = 0$, $\delta_0 = 10$, $\lambda = 0.5$. Both ε_d and ε_c are distributed $N(0, \sigma)$, where $\sigma = \sigma_d = \sigma_s$.[4]

If we were to set $\sigma \equiv \sigma_d = \sigma_c \leq 0.00001$, the model cannot mechanically be estimated because of virtually perfect collinearity. With a slightly larger amount of noise, we can mechanically estimate the model. In the tables, we report simulations for σ equal to 0.001, 0.5, 1, and 2.

We estimated the model using two-stage least square (2SLS), three-stage least squares (3SLS), and nonlinear three-stage least squares (NL3SLS). Except where otherwise noted, the tables report the 2SLS estimates. Both 3SLS and NL3SLS produce similar results.

---

[4] The exogenous variables are constructed as random variable where $w \sim N(3, 1)$, $r \sim N(0, 1)$, $Z \sim N(10, 1)$. Two additional instruments are created by adding an additional random variable drawn from $N(0, 1)$ to w and to r.

Table 1 shows several summary statistics. Because we used 2SLS, the R^2 for the demand and optimality equations may lie between $-\infty$ and 1. When relatively little error is added to the equations ($\sigma = 0.001$), the R^2 for these two equations is virtually one in every experiment. As the error grows, the mean R^2 measures falls, and is negative when σ is at least one for the optimality equation and two for the demand equation.

Although the demand equation, Equation 2, can be accurately estimated, the optimality equation, Equation 3, suffers from extreme multicollinearity (even with a large error). When $\sigma = 0.001$, the average condition number (the square root of the ratio of the largest to the smallest characteristic root of the regressors) is at least 1,433 in our examples and as high as 6.4×10^7 (same order of magnitude as with the infamous Longley data).[5] Belsley et al. (1980) suggest that condition numbers above 20 indicate potential problems. Similarly, in an auxiliary regression where we regress one of the right-hand-side variables, *ZQ*, on the others (w, r, and Q), the average R^2 is at least 0.91 (and virtually 1.0 when $\sigma = 0.001$).

Table 2 shows the coefficient estimates. With 2SLS, the multicollinearity in the optimality equation does not affect the demand equation, so we are able to estimate it very well (at least when $\sigma \leq 1$). For example, the true value of the coefficient on Q in the demand equation, ϕ_1, is 1. As Table 2 shows, the average estimated value for ϕ_1 is 1.0 (with a standard deviation of 0.004) when $\sigma = 0.001$, 0.99 (1.98) when $\sigma = 0.5$, and 0.97 (3.96) when $\sigma = 1$. The other two demand coefficients are estimated equally well.

[5] We are reporting the condition number using the actual right-hand-side variables, w, r, *ZQ*, and Z. If we replace the latter two with the instrumental estimates, the condition number rises by at least several orders of magnitude.

In contrast, we cannot accurately estimate the optimality equation coefficients due to the extreme collinearity. The true value α, the coefficient on *w* in the optimality equation, is 1. The average estimated value ranges from 0.46 to 0.49 with standard deviations that range between 0.88 and 1.04. The true value of the scale parameter, γ, is 1, but the average of the estimates range from 5.85 to 5.73 with large standard deviations (between 7.89 and 8.66). In the optimality Equation 3, the estimated standard deviations remain relatively unchanged as the size of the error terms fall (whereas the estimated standard deviations shrink in the demand equation as the error terms fall).

Typically, we are most interested in the market power coefficient, λ, which equals 0.5 in our experiments. With the 2SLS estimates, we obtain two estimates of λ.[6] As a practical matter, both provide nearly identical estimates in our experiments.[7] As the table shows, the average estimate of λ using 2SLS is usually negative and has a very large standard deviation. When $\sigma > 0.001$, the average of the 3SLS estimates is either negative or much above 1.0 (outside the plausible range). Even when $\sigma = 0.001$ and the average, 0.46, is close to the true value, the standard deviation is gigantic (8.98).

---

[6] First, we can divide the estimate of the coefficient on the *Q* term, $\lambda\phi_1$, in the optimality Equation 3 by the estimate of ϕ_1 in the demand Equation 2. Second, we can divide the estimate of the coefficient on the *ZQ* term in Equation 3, $\lambda\phi_2$, by the estimate of ϕ_2 from Equation 2.

[7] Imposing the restriction that the two estimates are equal in our NL3SLS model does not improve our estimates meaningfully.

Conclusions

Studies of market power based on market-level data are commonly used. Many of these studies have employed a linear specification that avoids the identification problem described by Bresnahan (1982) and Lau (1982). However, we demonstrate that estimates based on the linear model inherently suffer from multicollinearity problems, which makes such estimates by nature unreliable.

Appendix

We demonstrate that the *w*, *r*, *ZQ*, and *Q* terms in Equation 3 are perfectly collinear $\varepsilon_d = \varepsilon_c = 0$. We show this result by demonstrating that there exist nonzero coefficients χ_1, χ_2, χ_3, and χ_4 such that

$$QZ + \chi_1 Q + \chi_2 w + \chi_3 r + \chi_4 Y + \chi_5 = 0. \tag{A1}$$

To show that Equation A1 holds, we first solve for the equilibrium output, *Q*, as a function of the exogenous variables and parameters:

$$Q = \frac{\phi_0 + \phi_3 Y - \eta - \alpha w - \beta r}{(\lambda + 1)(\phi_1 + \phi_2 Z) + \gamma}. \tag{A2}$$

Substituting *Q* from Equation A2 into Equation A1 and rearranging terms, we obtain

$$\zeta_1 Z + \zeta_2 YZ + \zeta_3 wZ + \zeta_4 rZ + \zeta_5 Y + \zeta_6 w + \zeta_7 r + \zeta_8 = 0, \tag{A3}$$

where $\zeta_1 = \phi_0 - \eta + (\lambda + 1)\phi_2\chi_5$, $\zeta_2 = \phi_3 + (\lambda + 1)\phi_2\chi 4$, $\zeta_3 = -\alpha + (\lambda + 1)\phi_2\chi_2$, $\zeta_4 = -\beta + (\lambda + 1)\phi_2\chi_3$, $\zeta_5 = \phi_3 + [(\lambda + 1)\phi_1 + \gamma]\chi_4$, $\zeta_6 = -\alpha\chi_1 + [(\lambda + 1)\phi_1 + \gamma]\chi_2$, $\zeta_7 = -\beta\chi_1 + [(\lambda + 1)\phi_1 + \gamma]\chi_3$, and $\zeta_8 = [\phi - \eta]\chi_1 + [(\lambda + 1)\phi_1 + \gamma]\chi_5$. If we set $\zeta_i = 0$ for $i = 1,..., 7$, we have a seven-equation system in five unknowns, with the unique solution:

$$\chi_1 = [(\lambda + 1)\phi_1 + \gamma]/[(\lambda + 1)\phi_2]$$

$$\chi_2 = \alpha/[(\lambda + 1)\phi_2]$$

$$\chi_3 = \beta/[(\lambda + 1)\phi_2]$$

$$\chi_4 = -\phi_3/[(\lambda + 1)\phi_2].$$

$$\chi_5 = [\phi_0 - \eta]/[(\lambda + 1)\phi_2].$$

Thus, there exist χ_1,..., χ_5 such that Equation 4 holds, so the linear model is perfectly collinear and cannot be estimated.

References

Belsley, D., E. Kuh, and R. Welsch, Regression diagnostics: Identifying influential data and sources of collinearity, New York: Wiley, 1980.

Bresnahan, T. F., 1982, The oligopoly solution concept is identified, Economics Letters, 10, 87-92.

Bresnahan, T. F., 1989, Studies of industries with market power, in: R. Schmalensee and R. Willig, eds, Handbook of industrial organization, (North Holland, New York).

Deodhar, S. Y., and I. M. Sheldon, 1997, Market power in the world market for soymeal exports, Journal of Agricultural and Resource Economics, 22, 78-86.

Greene, William H., 1990, *Econometric Analysis*, New York:MacMillan Publishing Company.

Hyde C. E., and J. M. Perloff, 1995, Can market power be estimated? Review of Industrial Organization, 10, 465-85.

Jans, I., and D. I. Rosenbaum, 1996, Multimarket contact and pricing: Evidence from the U.S. cement industry, International Journal of Industrial Organization, 15, 391-412.

Just, R. E., and W. S. Chern, 1980, Tomatoes, technology, and oligopsony, Bell Journal of Economics and Management Science, 11, 584-602.

Lau, L. J., On identifying the degree of competitiveness from industry price and output data, Economics Letters, 10, 93-9.

Table 1
Linear Model Simulation Summary Statistics

	$\sigma = 0.001$	$\sigma = 0.5$	$\sigma = 1$	$\sigma = 2$
Model				
R^2 Demand Equation	1.00	0.76	0.32	-0.15
	(0.3E-6)	(0.07)	(0.22)	(0.52)
R^2 Optimality Equation	1.00	0.51	-0.41	-1.85
	(0.1E-4)	(2.67)	(8.13)	(19.85)
Multicollinearity Checks				
R^2 Auxiliary Equation*	1.00	0.91	0.91	0.96
	(0.3E-6)	(0.03)	(0.03)	(0.01)
Condition number	6.4E7	1,433.2	1,614.9	2,643.2
	(1.8E7)	(483.2)	(575.13)	(1,027.1)

Note: Standard deviations in parentheses.

* R^2 of the auxiliary regression of ZQ on w, r, and Q.

Table 2
Linear Model Simulation Estimates

	True Coefficient	*Average Estimated Coefficient (Standard Deviation)*			
		σ = 0.001	σ = 0.5	σ = 1	σ = 2
2SLS					
Demand					
ϕ_0	10	10.00 (0.001)	9.96 (0.33)	9.86 (0.65)	9.46 (1.20)
ϕ_1	1	1.00 (0.004)	0.99 (1.98)	0.97 (3.96)	0.88 (7.80)
ϕ_2	1	1.00 (0.004)	0.99 (0.21)	0.97 (0.42)	0.87 (0.82)
Optimality					
α	1	0.46 (0.88)	0.46 (0.91)	0.47 (0.93)	0.49 (1.04)
γ	1	5.85 (7.89)	5.85 (8.15)	5.78 (8.21)	5.73 (8.66)
λ	0.5	-0.31 (1.31)	-0.29 (1.34)	0.09 (11.48)	-1.53 (30.41)
3SLS					
λ	0.5	0.46 (8.98)	-0.61 (24.34)	27.28 (844.28)	1.22 (25.57)

The demand curve is $p = \phi_0 - \phi_1 Q - \phi_2 ZQ + \varepsilon_d$.

The optimality condition is $p = \eta + \alpha w + \beta r + [\lambda(\phi_1 + \phi_2 Z) + \gamma]Q + \varepsilon_c$.