

# Econ INSIDE One

ECONOMIC THEORIES IN PRACTICE / JANUARY 2008

## WHAT'S YOUR SIGN?

*By Mark Dwyer, Ph.D., Econ One, Los Angeles, CA*

Everyday intuition can be misleading when using economic (or practical) significance to evaluate the reliability of regression analysis.

Economic experts often employ regression analysis to quantify the impact of factors affecting prices, market shares or other economic variables. For example, regression analysis can provide a means of isolating the price effect of an alleged conspiracy from the effects of other demand and supply factors on prices. Such a regression model would provide estimates for both the effect of the alleged conspiracy and of these other supply and demand factors. The reliability of the model and consequently, the reliability of its conspiracy effect, are often evaluated by examining the estimates of these additional factors. Two tools for assessing the reliability of such estimates produced by regression models (or more generally, by any statistical model) are examinations of their statistical<sup>1</sup> and economic significance. My focus here is evaluating the apparent economic significance of empirical results.

Examinations of economic significance have two aspects.<sup>2</sup> First, does an estimated effect have the right sign? Second, is the magnitude of the effect plausible? These issues are addressed first regarding the variable of interest--say, the conspiracy effect--then attention turns to the other factors in the model. Answers to both the correct sign and plausible magnitude questions depend on prior information

regarding the relationship of interest. For example, in estimating whether an alleged conspiracy artificially raised prices in Los Angeles, the model would control for factors (other than the conspiracy) that would plausibly be thought to affect those same prices. Prices in areas not impacted by the conspiracy (e.g., San Francisco, Denver, Houston, New York) might capture some of the effects of non-conspiracy factors--be they changes in labor conditions, temporary supply chain variations, fortuitous product placement events, terms of foreign trade or the vagaries of fashion--that influence prices in Los Angeles. Such geographically varying prices can be particularly useful if direct data on the underlying supply and demand factors driving prices are difficult to obtain.<sup>3</sup>

When considering a model in which prices in Los Angeles are accounted for, in part, by prices in other cities, what prior information should you use to judge the economic significance of these factors? Consider the issue of the sign of each of these control cities. Intuition drawn from real-world experience would suggest that prices in these control cities tend to move together, and with prices in Los Angeles. One might then conclude that the estimated relationship between each of these city's price on the price in Los Angeles would be positive. This conclusion however is likely to be incorrect.

Intuition may not be a reliable guide to establishing the signs of regression factors because

## WHAT'S YOUR SIGN? -CONTINUED

regression models estimate the marginal or partial relationship between each city's price and the Los Angeles price. These estimates are capturing the result of a hypothetical experiment in which other factors in the model, including prices at other locations, are constant and only the prices in Los Angeles and one other given city happen to vary (e.g., the relationship between Houston and Los Angeles prices is what we would observe if all other cities' prices were not changing).

Such conditions are unlikely to be observed in practice. Instead, prices across all of the cities tend to move together but not in lock-step. Therefore, our casual observations can mislead us regarding plausible signs for economic factors.

Simple regressions generally accord better with our day-to-day intuition, but cannot capture the necessary relationships. These regressions model the effect of a single factor at a time. In the example above, simple regressions would capture the common movement of Houston prices with Los Angeles prices, and then San Francisco prices with Los Angeles prices, and so on. That exercise would generally agree with our prior experience, and reveal the positive relationships between prices in each city and Los Angeles that we would expect, but it would tell us little about the actual marginal relationships. These marginal relationships are what we need to successfully control for factors beyond the alleged conspiracy.

As an example, the figure below plots average weekly retail gasoline prices in 10 cities. It is apparent that the gasoline prices across these cities tend to move together. Correlations of each of the other 9 city prices with the price in Los Angeles range from 94.8 percent to 98.2 percent.<sup>4</sup>

In contrast, within a fuller specification that includes prices in all 9 remaining cities as factors

accounting for price variation in Los Angeles, 5 out of 9 of these city prices have negative estimated relationships with Los Angeles.<sup>5</sup>

As the example above illustrates, one reason why the signs of marginal relationships (as measured by regression analysis) may be at odds with our intuition is that these effects may tend to move together in everyday experience. Such behavior has been pejoratively labeled as "collinear." Models that include such factors--like the example above--are sometimes castigated in economic literature as "suffering" from "collinearity" or more redundantly "multicollinearity." In fact, unexpected signs on estimated effects are sometimes held up as a "warning sign" of collinearity.<sup>6</sup> This criticism is usually without merit. Collinearity can, at most, only be considered a relevant concern when the variable of interest (e.g., the conspiracy effect) is collinear with other effects.

Evidence that prices in other cities are collinear is largely immaterial, because identifying the precise relationship between each city's price and the price in Los Angeles is not the objective<sup>7</sup>. Rather, the objective is measuring the effect of the alleged conspiracy. Other cities' prices function as controls and collinearity among controls is generally not an issue.<sup>8</sup>

One response to counterintuitive signs, and to collinearity more generally, is to discard some of the collinear effects, as being redundant. This is often accomplished through model selection techniques that attempt to prioritize these variables so that the modeler can get rid of those with the least explanatory power.<sup>9</sup> In the present example, one would try to discard one or more of the other cities' prices from the model.

Such exercises are unwarranted when the variable of interest (e.g., the conspiracy effect) is not one of the factors affected by the collinearity. Even a high correlation between the conspiracy variable and other model factors provides no motivation for discarding variables if the conspiracy effect is

## WHAT'S YOUR SIGN? -CONTINUED

statistically significant. The expected consequence of including prices for locations that do not bring with them significant additional information is to lower the overall statistical significance of estimated effects.<sup>10</sup> The argument that a conspiracy effect is statistically significant because a model is “overfit” lacks statistical foundations.

While the hypothetical above used prices in cities across the country as controls, pricing models more often employ lagged prices. For example, Los Angeles gasoline prices might be modeled as depending on gasoline prices in Los Angeles in the preceding weeks. Similar comments regarding the interpretation of the signs and magnitudes of these lagged prices apply in this context as well. Prices and lagged prices tend to move together, so any intuition regarding the “correct” sign of lagged effects should be considered tentative at best.

To illustrate, while correlations between Los Angeles gas prices in a given week and those in prior weeks range from 91.9 percent (at 4 weeks prior) to 99.1 percent (at 1 week prior), a model of Los Angeles gasoline prices on the preceding 4 weeks reveals a

strong negative partial effect at two weeks.<sup>11</sup>

I should comment on one notable difference between models that rely on lagged prices as controls and models that rely on factors outside the reach of the alleged conspiracy. In the former case, the controls cannot be expected to be independent of the conspiracy. Even so, collinearity between the conspiracy effect and lagged price effects will be strongest when the conspiracy’s impact on price itself is strongest, and hence, when it is most clearly identified. Thus, here too, the potential collinearity of lagged price controls should not cast dispersions on the statistical significance of the conspiracy effect.

In summary, counterintuitive signs on controls are often not an indication of model misspecification, but rather simply an indication that the marginal relationships the analysis is identifying are not readily apparent through casual observation. Relationships for factors that tend to move together (i.e., that are collinear) will tend to be more at odds with simple intuition. The defect is in our intuition, not in the model. Common movement (i.e., collinearity) among controls is not usually a problem, and raising it as an issue does not constitute a valid attack on a statistically significant conspiracy effect.

<sup>1</sup> See Choi, William and Mark Dwyer, “Is a ‘Statistically Significant’ Result Significant?” Inside Econ One, Spring 2004 for a discussion of precautions related to statistical significance.

<sup>2</sup> See for example, Wooldridge, Jeffrey M. *Introductory Econometrics: A Modern Approach*, South-Western, 2000, pp. 131-134, 735-736.

<sup>3</sup> In this example, as often in practice, the available explanatory variables are proxies for economic fundamentals. This proxy nature of many model controls further weakens our ability to assess their economic significance. For an introduction to proxy variables see Wooldridge, *op. cit.*, pp. 284-291.

<sup>4</sup> These high, positive correlations imply positive relationships in the single-city, simple regressions described above. These coefficients range from 0.874 to 1.013, again indicating that on average, gasoline prices across the country tend to be positively related to Los Angeles prices.

<sup>5</sup> In this model, prices in control cities are from the prior week. The price in Los Angeles from the prior week is also a control.

This example is intended to capture some of the most salient features of such a specification for illustrative purposes.

<sup>6</sup> See for example, Greene, William H., *Econometric Analysis*, 4th ed., Prentice Hall, 2000, p. 256.

<sup>7</sup> In the Los Angeles price conspiracy example above, collinearity would only be a problem if prices in other cities only happened to shift significantly in tandem with the onset and the end of the hypothetical conspiracy. This would call into question the presumption that these cities were untainted by the conspiracy.

<sup>8</sup> See Wooldridge, *op. cit.*, pp. 95-97.

<sup>9</sup> See for example Ramanathan, Ramu, *Introductory Econometrics with Applications*, 4th ed. Dryden, 1998, p. 238 as an advocate of such a process, or Greene, *op. cit.*, pp. 341-342. For an introduction to model selection procedures, see Ramanathan *op. cit.*, pp. 164-166.

<sup>10</sup> See, for example, Wooldridge, *op. cit.*, p. 102, or Greene, *op. cit.* p. 338.

<sup>11</sup> As with the previous empirical example, this model aims to be illustrative rather than to represent a complete pricing specification.