Exhibit 1

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

OAKLAND DIVISION

THE APPLE iPOD iTUNES) Lead Case No. C 05-00037
ANTI-TRUST LITIGATION)
)
_____________________________)
This Document Relates To:)
ALL ACTIONS)
)
_____________________________)

CONFIDENTIAL - ATTORNEYS' EYES ONLY

VIDEOTAPED DEPOSITION OF KEVIN M. MURPHY, PH.D.

VOLUME II

January 08, 2014

Phoenix, Arizona

Reported By:
Cathy A. Miccolis
RPR, CRR, CSR No. 50068
Job No. 10009198

A P P E A R A N C E S

For the Plaintiffs: Bonny Sweeney, Esq.
ROBBINS GELLER RUDMAN & DOWD, LLP
655 West Broadway
Suite 1900
San Diego, CA 92101
619.231.1058
bonnys@rgrdlaw.com

For the Defendant Apple, Inc.:
David C. Kiernan, Esq.
JONES DAY
555 California Street
26th Floor
San Francisco, CA 94104
415.626.3939
dkiernan@jonesday.com

Also Present: Thomas C. Tracy, videographer

I N D E X

E X H I B I T S

THE VIDEOTAPED DEPOSITION OF KEVIN M. MURPHY, Ph.D., VOLUME II, was continued on January 8, 2014, commencing at 9:11 a.m. at the offices of BONNETT, FAIRBOURN, FRIEDMAN & BALINT, P.C., 2325 East Camelback Road, Suite 300, Phoenix, Arizona, before CATHY MICCOLIS, a Certified Reporter in the State of Arizona.

THE VIDEOGRAPHER: The time on the record is 9:11 a.m. Today's date is January 8, 2014. My name is Tom Tracy of Aptus Court Reporting. The court reporter is Cathy Miccolis of Aptus Court Reporting located at 600 West Broadway, Suite 300, San Diego, California 92101.

This begins the videotaped deposition of Kevin Murphy, Volume II, testifying in the matter of the Apple iPod iTunes Trust (sic) Litigation, pending in the District Court of California, Oakland Division, Case Number C 05-00037 YGR. This deposition is taking place at 2325 East Camelback, Suite 300, Phoenix, Arizona 85016.

Will counsel please identify themselves, starting with the plaintiffs' counsel.

MS. SWEENEY: Bonny Sweeney for the plaintiffs.

MR. KIERNAN: David Kiernan for Defendant Apple, and Scott Murray, in-house counsel from Apple, may be on the phone.

Scott, are you on the phone?

Okay. No answer.

THE VIDEOGRAPHER: Thank you, Counsel. The court reporter may swear in the witness so we can proceed.

KEVIN M. MURPHY, Ph.D.,

having been first duly sworn to tell the truth, the whole truth, and nothing but the truth, was examined and testified as follows:

EXAMINATION

BY MS. SWEENEY:

Q. Good morning, Professor Murphy. We have met before. My name is Bonny Sweeney. I'm going to be taking your deposition again today in the Apple case.

MS. SWEENEY: Before we get started, I just wanted to ask counsel for Apple a question. So I just want to make sure, you have an open telephone line. I want to know anyone who is on that line.

MR. KIERNAN: There is no one else on the line.

MS. SWEENEY: Okay.

MR. KIERNAN: It's just an open line for Scott Murray, counsel from Apple, and if he joins, he will announce himself.

MS. SWEENEY: Will there be any indication as to when and who joins?

MR. KIERNAN: There is a -- it makes a beeping sound.

MS. SWEENEY: Okay.

BY MS. SWEENEY:

Q. So Professor Murphy, you remember that I deposed you once before in this matter; correct?

A. Yes, I do.

Q. And that was in November of 2013?

A. Sounds about right, although I couldn't tell you for sure.

Q. And since that deposition what work have you done on the Apple case?

A. I have done the work that's reflected in the declaration that I gave here. I have continued to read through the declarations prepared by plaintiffs' experts, both Professor Noll and Professor Wooldridge, and evaluate those reports and the claims that are made therein.

Q. Okay. When you say you've done the work that's reflected in the declaration, are you referring to the supplemental report that you submitted together with Professor Topel dated December 20, 2013?

A. Yes, because that was produced after our earlier deposition, so when you asked me what we had worked on since then, that was a component that I definitely had worked on since that date.

Q. Okay. And for the -- and we will just refer to that as the supplemental Murphy and Topel report, if that's okay with you.

A. That's fine.

Q. So with respect to the supplemental Murphy and Topel report, did both you and Professor Topel write the report?

A. Yes, we worked on it together.

Q. And how does that work when two people are writing a single document in your case, how does that work?

A. We first talk about what it is we want to do and what -- you know, what we are, what we are presenting in the report. We then set out to write that up. Obviously one person at a time writes. We don't sit next to each other and write. So one of us will take a different part and work on writing that part up. And then the other one will review it, make changes, edits, whatever, probably discuss it further back and forth between us with writing and discussing material that's in the report.

Q. And you said that "each of us would take different parts." What part or parts did you have primary responsibility for?

A. You know, I don't really recall. I don't

recall whether I had primary responsibility for one or the other. I know I worked on the whole thing, so I don't even remember which one I started with.

Q. Did anyone assist you and Professor Topel in writing the supplemental report?

A. Yes, we had some assistance from Anita Garten who we work with regularly. And certainly in preparation in terms of the tables and other things that went in there, other people, Ricardo Cossa would be a primary person who worked on doing a lot of the programming and statistical work based on our direction that's in the report.

Q. Anyone else?

A. There would be others. I don't recall all the people who worked on the statistical end of things. In terms of drafting the report it would be me, Professor Topel, Anita Garten. I believe other people looked at it and gave us their thoughts. That would be Ricardo Cossa and Bin Chen I believe would be the other one and possibly Naraj.

Q. And when you and Professor Topel sat down to put together the supplemental report, what were you addressing in that report?

A. We -- I think we state right at the beginning of the report, we address some of the issues that

Professor Noll brought up in his -- I don't know exactly what the title of his rebuttal or his most recent report that was filed after our earlier reports. So we based our discussion there on the claims made by Professor Noll in his report. And we focused on the topics he discussed in his supplemental report or whatever the specific -- actual name of it is, not back to his original report, and we make reference to his original report only when it's needed for context. I think this is described directly in the report that we filed.

Q. And in preparing the supplemental report, did you or staff at CRA take any additional statistical analysis?

MR. KIERNAN: Object to form.

THE WITNESS: Yes, I believe there is statistical analyses presented in that report. Yes. So we did do statistical analysis for purposes of that report.

BY MS. SWEENEY:

Q. Can you describe the analyses that you did?

A. They are contained in the report. It mostly consisted of variations of the regression model that Professor Noll estimated, many of the variations we had done before. Although since he had changed the specification in some ways, it was a little different.

But the basic take of what we did there in terms of redoing his regression analysis was very similar to what we had done before. It just simply worked with his latest specification rather than the specification that was in his earlier report. I think the substance of what we discussed there is very similar. We went on and because he had changed the way he did his weighting and calculation of his standard errors, we went and did some new work using the same methodology to illustrate that there was still a correlation problem with his residuals from his regression. That would be different than what we had in the earlier report, not in the sense that the methodology is different, is because he had done something different. The specifics of the results would be somewhat different, although the basic conclusion would be the same that he continues to have a rather severe problem with correlation between different observations in his dataset that causes him to greatly understate the standard errors of his estimates.

Q. So when you say you conducted some new work using the same methodology that you had used in your previous reports, and when I say reports here I'm referring to the earlier Murphy and Topel reports, am I correct in sort of rephrasing to say that you didn't conduct any new tests, but you just used the different

data that were in the Noll rebuttal report?

A. No, I think we did -- I think we did some things that were different. I think the basic thrust of what we did was very similar. I believe we did some different things in this, in this report than we did in the other. They are not inconsistent with what we did before, but we did add some to that. But the biggest difference, particularly in terms of the regression models, would be the basis for those, would have been his new regression specification rather than the old one.

Q. So what are the different things that you did?

A. You know, I don't recall each one of them individually so there may be some that I leave out. For example, I know we did some F-tests on various blocks of coefficients in this draft that we had not done in the earlier one. We did I think two different versions of clustering in this particular report that we hadn't -- we had done just the single version before. We also examined the correlation of the residuals somewhat further from what we had done in the previous example. So we had done a few things different. I don't recall all of them off the top of my head.

Q. So you said you did F-tests on blocks of coefficients. You hadn't done any F-tests prior to this supplemental report?

A. I think -- I don't know if we did some F-tests or not. I know we did some specific ones here I believe than we did before.

Q. And are those reflected in the report?

A. Yes, they are reflected in I believe Table 5. It's JT-5 or something like that is the table that has those.

Q. And then you said you did two different versions of clustering, whereas in your prior report you'd only done one different version of clustering. What do you mean by that?

A. In this one we did I believe clustering by family as opposed to by family quarter. I think if I recall correctly, that's -- I don't want to -- it's either JT-3 or JT-4. I can't remember which one, but we did that in this report.

Q. And did you also do clustering by family quarter?

A. Yes. That's a primary thing, but that's what we had done before, so that wouldn't be different. That was how we had done it previously.

Q. And when you say "family," what are you referring to?

A. It's the iPod family, which is a categorization that Apple uses to distinguish different models of their iPod, so it might be like an iPod mini, second generation, given quality level, say good, better or best, one of those.

Q. And then you said that you also examined correlation of residuals in a way that you hadn't before. Can you explain that a little more?

A. I don't recall everything that we put -- that's in the report, so I just know that we redid that analysis based on his new results, and I'm not sure everything we did there is exactly the same as what we did before. I'd have to go back and look at the report to tell you.

Q. You also testified that you evaluated the plaintiffs' experts' declarations, including the Wooldridge report. Have you done any subsequent statistical analysis since reviewing the Wooldridge report?

A. Yes, I have. I have looked at the data to evaluate whether his claims are correct or not, and I think the data overwhelmingly say that his claims are incorrect.

Q. And when you say you looked at the data, what data did you look at?

A. The data based on Professor Noll's regression data in his regression model.

Q. And do you have -- do you anticipate doing any

further writing of reports or presenting of analyses that will reflect your conclusions regarding Professor Wooldridge's analysis?

MR. KIERNAN: And I'm going to object and instruct the witness not to answer any, to the extent it reveals any communication or requests from counsel. So I'm instructing you not to answer the question to the extent that it would reveal requests from counsel.

BY MS. SWEENEY:

Q. Are you going to answer the question?

A. I have been instructed not to answer, so...

Q. Well, let's break that up a little bit. So the stipulation in this case says that certain categories of data, information, documents and materials need not be produced, and that includes written correspondence between an expert and the attorneys, notes taken and prepared by or for an expert in connection with the matter, including notes of conversations with the attorneys, but then the stipulation goes on to say, and what I was paraphrasing was in paragraph 3 of the stipulation, goes on to say paragraph 3 -- excuse me, paragraph 5 says, nothing in paragraph 3 however shall be construed to prevent substantive deposition questions with respect to alternative theories, methodologies, variables, data, production of materials or assumptions that the expert may

have considered in preparing his or her report. And paragraph 6 of the stipulation order should not be construed to preclude reasonable questions at deposition going to the expert's compensation, hours expended in preparing his or her report and testimony and frequency and duration of meetings with counsel.

So what I'm getting at is I think that counsel for Apple has construed overly broadly the restrictions of the stipulation. So let me ask you again --

MR. KIERNAN: Bonny?

MS. SWEENEY: Yeah.

MR. KIERNAN: I'm happy for you to ask him about the categories in 5 or 6, but your question was whether or not he has been asked to or that he will submit any additional writing in response to either Drs. Noll's or Wooldridge which is not covered by Categories 5 or 6. So you can ask him about 5 or 6.

MS. SWEENEY: Let me ask you, David, is there anything in the stipulation that prohibits oral questions at deposition about the expert's discussions with the attorneys? I see it talks about documents. I don't see anything that says I can't ask him what the lawyers asked him to do. If you can point that out to me, maybe I'm missing it, but I don't see it.

MR. KIERNAN: Well, I will pull out previous

objections and instructions that you gave Dr. Noll, relying upon --

MS. SWEENEY: Good luck with that. I never stopped him from answering those kind of questions.

MR. KIERNAN: Are you going to continue to interrupt me?

MS. SWEENEY: No. Go ahead. Let's create this record.

MR. KIERNAN: Okay.

-- instructing him not to answer questions about communications between counsel and himself. And Xan Bernay also instructed Dr. Martin not to answer any questions about communications that she had or that you had with Dr. Martin --

MS. SWEENEY: Well --

MR. KIERNAN: -- based on the stipulation.

MS. SWEENEY: Are you done? I'm sorry. I don't want to interrupt you.

MR. KIERNAN: I am done.

MS. SWEENEY: I don't believe that that's a proper interpretation of the stipulation, and to the extent you let that pass, I can't explain that. I'm sure I never instructed Professor Noll not to answer your questions on that basis. And if you can point me to some deposition testimony, I would be very, very surprised because that is just completely false. So if you're instructing the witness not to answer, that is an improper instruction, and I will take it to the court. So let me make sure the record is clear on this.

BY MS. SWEENEY:

Q. Professor Murr- -- strike that.

Professor Murphy, are you on the advice of counsel for Apple refusing to answer my question about communications that you had with Apple's counsel about work that you have performed or will perform in this matter?

A. Yeah, I'm taking the advice of the counsel for Apple. Since I'm not a lawyer, I can't tell you all the interpretation, and I can -- best I can do is go with the advice of counsel.

Q. So you're refusing to answer any question that I ask about instructions given to you by counsel for Apple?

A. If that's what I'm instructed to do by counsel for Apple, that's presumably what I should do.

Q. So if counsel for Apple has instructed you or provided information to you that you've relied on, are you going to refuse to answer questions about that based on the advice of counsel?

MR. KIERNAN: Objection. Object to form.

THE WITNESS: I --

MR. KIERNAN: And that was not the instruction I gave.

THE WITNESS: -- will do my best to answer your questions. If there are specific things that I'm not supposed to answer because they are covered by stipulation, the best I can do to comply with the desires of the court I see is to try to follow those directions to the best I can, and for that I need to rely on counsel.

MR. KIERNAN: And just so the record is clear, I have not made a broad objection as you're stating -- that you're suggesting that Dr. Murphy -- this should go question by question, and if there is a question you want to ask him about whatever you want to ask him, then we will go question by question, and if I feel that it's violating or outside the stipulation, I will assert the objection and instruct the witness not to answer. If I feel that it's not covered by the stipulation, then I won't assert the objection.

MS. SWEENEY: Well, this is an issue on which we strongly disagree. I think that we will just see how it goes, but this is something we will probably have to raise with the Court.

BY MS. SWEENEY:

Q. Professor Murphy, what additional work do you

anticipate doing in this matter?

A. I anticipate continuing my work, evaluating the claims made by Professor Noll and by Professor Wooldridge, that's ongoing, and I can -- I would, I would presume that I would continue to do that in the coming days and weeks and however long it takes.

Q. Are you writing a report to be submitted next week in opposition to plaintiffs' motion to exclude portions of your testimony and Professor Topel's testimony?

A. I don't know whether there is going to be a report submitted or not. I -- I know I'm continuing to do work. And there is a possibility that that would lead to a report. I would understand, but I can't tell you whether there is going to be one submitted or not. That's -- I'm just going to do my work and continue to try to evaluate the claims of plaintiffs' experts.

Q. Did you receive input from Apple's counsel in your report, your supplemental report?

A. What do you mean by received input from them? They gave -- they gave us comments. That's pretty much standard on drafts of the report. I don't know. Usually that's covered by stipulation, the specifics of those comments, but again, you guys are the lawyers, so, yes, we did discuss the report and the contents of the report with

counsel.

Q. Did you adopt the suggestions of Apple's counsel in your supplemental report?

A. Not really.

Q. Did they write any of the report, that is, Apple's counsel?

A. No.

Q. Did you write or did Professor Topel write all of the supplemental report?

A. No, there would have been help from -- I believe Anita Garten helped. Ricardo Cossa maybe did a little bit. I know he provided some comments and edits, but I would say we wrote the report between Bob and myself. Some of the cites and footnotes were filled in by people. We would say cite to a particular document or particular book or chapter. We wouldn't necessarily know the exact title it would be, you know, cite to Angrist and Pischke. Somebody would have to fill it in with the entire Angrist and Pischke cite. So if you consider that writing a report, other people filled in the rest of those footnotes and things like that. But the substance was written by Bob and myself.

Q. How long -- strike that.

How many hours did you spend preparing the supplemental report?

A. I don't recall.
Q. More than 10?
A. Yeah, it would have been more than 10.
Q. More than 100?
A. It would have been less than 100.
Q. Less than 50?
A. I would assume so.
Q. Less than 40?
A. That I -- now we are getting down to -- my level of resolution I think is about that level.
Q. Do you know how much time Professor Topel spent on it?
A. I do not. You'd have to ask him.
Q. Is there anything today that you would like to revise or correct in your supplemental report?
A. Nothing that specifically I'm aware of, no.
Q. Is there anything in your prior report that you want to correct or revise?
A. No, not that I'm aware of.
Q. Have you reviewed the Wooldridge deposition transcript?
A. I skimmed through it. I just got it yesterday, and I was busy at school with things yesterday, so I only had a chance to skim through it.
Q. And did you review the deposition transcript from Professor Noll's deposition?

A. I did. I did, but that was a while back.
Q. Can you summarize for me the opinions that are stated in your supplemental report?
A. Probably not all of them. I mean, I think there is some basic ones. One is I think the first and foremost is Professor Noll has continued to overstate the significance of his results by un -- by failing to account for the strong degree of correlation across transactions, and that's a very substantial error. It results in him calculating standard errors that are roughly 100 times too small. I mean, not always exactly that. Sometimes a lot more; sometimes less. But it's a gross misestimate of the precision of his estimates.
That the -- his reasons for not taking account of the correlation are invalid, and we go through each of those in the report. The residuals at the family by quarter level are demonstrably correlated to a high degree. The fact that he has all the transactions, actually nearly -- or nearly all the transactions, what he calls it's a population rather than a sample doesn't diminish the need to account for the correlation that's present in the data that he has.
That he continues to have omitted variables in his regression, that important determinants of iPod pricing that he leaves out which biases his results and causes him to misestimate the impact of any of the iTunes 7.0 release. He continues to use the wrong but-for world and the wrong specification of the impact of iTunes 7 on iPod prices.

The basic methodology he continues to use is greatly flawed because of the fact that again his estimates are not isolating the impact of the challenged conduct, that simultaneous with the release of iTunes 7 other things in the marketplace changed, including the models being offered by Apple, the characteristics of those products, and rendering his analysis sort of invalid conceptually. Those are the ones I remember. There might be some more conclusions, but those are the basic ones.
Q. And are the bases for all of the opinions that you've just summarized set forth in your supplemental report?
A. Yes. I think we tried to summarize the bases for those opinions at the time we wrote the report, yes.
Q. Now, you said that Professor Noll has continued to overstate the significance of his results, and you say that there is a strong degree of correlation. Did you rely upon any economic literature other than the literature you cited in your first report to support your conclusions regarding the clustering?

A. I don't recall what all we cited in our first report. There is certainly a broad economic literature that supports what we put forward there. I would say what we cite here is certainly you could find other cites that would echo the same points. The articles we have cited, you know, sort of Cameron and Miller and Donald and Lang and Angrist and Pischke and the Hansen, Chris Hansen's work, I mean a number of things that we have referred to and looked at all support the same thing.
If you want to see more, I think Cameron and Miller have a pretty extensive discussion of the underlying further literature that you might want to look at if you wanted more literature, but we didn't cite those directly.
Q. Do you consider yourself an expert on clustering issues?
A. I would consider myself an expert at applied econometrics generally and in particular methods to be used for these type of data. I have spent my career analyzing aggregate -- more aggregated phenomena using microdata, so this is an area I'm very familiar with. So it's something that I would say, yeah, I would consider myself very well versed in.
Q. Have you ever written any articles that have been published specifically relating to clustering?

A. I don't think I published any specific articles on clustering. I have certainly talked about how you use data of this type where you have market-level phenomena you're interested in and micro-level data used to estimate it.

I did a lot of well-cited work on the wage structure, which we made exactly the point that we are talking about here, that even though you might have hundreds of thousands of observations from underlying datasets, when it comes to estimating the determinants of market prices, you have much less data because in fact you only have data on a limited number of market equilibrium. And that's exactly the same point that's being made here. And my work in that area goes back probably almost 30 years.

Q. Now, why don't we go ahead and mark the supplemental report as Murphy Exhibit 6, because there were five exhibits marked in your prior deposition.

(Exhibit 6 marked.)

Q. Now, you said that you and Professor Topel jointly wrote the report; correct?

A. That's correct.

Q. For those parts that you didn't yourself write, do you agree with them?

A. Yeah, I mean, I worked on every part of the

report. I mean, we didn't like write them and then glue them together and send them in. We worked on them back and forth. We talked about it before we ever wrote anything. Even if one of us wrote the paragraph the first time, the other one went through it, and we ended up with something we were both satisfied with.

Q. So you agree with all the statements in the report; is that correct?

A. To the best of my knowledge, yes.

Q. In paragraph 5.a., you say that Professor Noll simply ignores the clustering problem. In your opinion did Professor (sic) in his rebuttal --

MR. KIERNAN: Page 2.

BY MS. SWEENEY:

Q. -- report ignore the clustering problem?

[redacted]

Q. Professor Noll disagrees with you that there is a clustering problem; isn't that correct?

A. Yes, but he disagrees by just denying it, not by presenting evidence that there is no correlation. He doesn't -- he doesn't calculate the correlation and say, look, the correlation is low. He never -- he never does that. I mean, we present a graphical representation of just how highly correlated the residuals are and uses -- you know, he just asserts things without really doing any analysis.

Q. That's interesting.

Did you perform any cluster analyses that are not presented in your exhibits?

A. What does that mean?

Q. Well, you and your staff at CRA performed some analyses on the clusters; correct?

A. We have continued to work on the clustering analysis. Well, it's actually not really the clustering analysis. Analysis of Professor Noll's model and analysis of the error terms in that model and their properties. So we have continued to work on that subsequent to the report. Is that what you're asking?

Q. No. What I want to know is, did you perform any analyses that relate to your opinion regarding clustering that are not presented in the tables in your

supplemental report?

A. Are you referring to analyses done before the report or after the report?

Q. Let's start with before the report.

A. Yeah, I'm sure we did additional analyses that we are not relying on. I don't -- there is nothing that would be inconsistent with what we have done there. There are often many ways to do things, and when it comes time to write a report, you settle on one that you think accurately what you did. But nothing that I'm relying on, no.

Q. And is that true also for your initial report, that is, that you conducted certain analyses relating to your clustering opinions that are not reflected in the tables and exhibits attached to your report?

MR. KIERNAN: Objection; argumentative.

THE WITNESS: Nothing that I would be relying on.

BY MS. SWEENEY:

Q. But nonetheless, is it your testimony, let's start with the supplemental report, that those analyses that are not reflected in your exhibits in which you're not relying on are nonetheless consistent with your opinions?

A. Yes, I would say the analysis we have done is [redacted] And we settled on quarter because we thought it represented a reasonable compromise between taking account of that correlation while still allowing for some independence over time. I think that's the best justification you have for focusing on quarters. If somebody wanted to aggregate further, I think that could be reasonable. I think aggregating less is going in the wrong direction. I think I would continue to stand by that. And certainly doing what Professor Noll did because it neither aggregates over time nor across transactions at a point in time doesn't make any sense.

BY MS. SWEENEY:

Q. So you think it's going in the wrong direction to cluster at any shorter time interval, but you don't know in fact whether your staff at CRA conducted the clustering analysis by week; correct?

MR. KIERNAN: Object to form.

THE WITNESS: I don't believe they did because I didn't ask them to do that because I wouldn't have thought going in that direction made much sense. It's not something that would tell -- economics or the data or the

documents or anything in this case would tell me to move in that direction. So that's not something I would have instructed them to do, at least as I recall. I can't tell you what all -- anything they did, but not that I know of. It's the best answer I can give.

BY MS. SWEENEY:

Q. You testified a little bit ago about the difference that Professor Noll asserted between samples of data and the population of data. Are you familiar with any other instance in which the kind of adjustment that you performed with respect to the standard errors was used to adjust standard errors for clustering when the dataset was a population rather than a sample?

A. Yeah. I mean, for example, in Professor Wooldridge's work that he did on unilateral divorce, the sample -- the data on divorce there are population. They are the divorces that occurred in each state from vital statistics, which is a population, not a, not a sample. The --

Q. Anything else?

A. Yeah. I mean, in like price-fixing cases and things like that where you have all the transactions for a company, you don't -- and you're doing a price analysis, you typically don't treat all the prices as independent because they are determined by common factors. You would

allow for clustering.

Or when people analyze -- go back to the original work of Fama and McBeth. They have -- you know, they have all the stock market data. They have all the closing prices on a daily basis for all the stocks in their universe. They don't have a sample of dates. They got every date, and they have got, you know, all the stocks in their universe. They didn't draw -- they didn't draw a random sample. They had that. And Fama and McBeth, they essentially did original work actually kind of related to clustering. They used a different methodology, but they were concerned with exactly that same phenomena.

Q. Anything else?

A. I mean, there is bunches of them, but there is nothing, there is nothing unusual about that, about doing it.

I think Professor Wooldridge in his textbook talks about the Michigan teachers and clustering in the context of --

(A brief interruption.)

A. So he has one on Michigan teachers I believe in his textbook where again he is dealing essentially with population data. So yeah, it's not hard to find examples of that.

Q. So I asked you whether you tested weekly -- I asked about weekly. Did I ask about monthly clusters? Did you conduct that analysis?

A. You know, I don't recall whether we did or not. It's not something I remember. We might have done that as a sensitivity check back when. It's not something I have done recently. Might have done that before the first report, but I don't recall one way or the other.

Q. But if you --

A. It's not relying on it for sure.

Q. If you did do that, it's not reflected in the report or its exhibits or attachments; correct?

A. Yeah, because I wasn't relying on that. I don't recall doing it specifically, but I don't want to say something that's not true, so best I can say is I don't recall.

Q. So you don't know whether if you created clusters at the weekly level, for example, the results would be statistically significant?

A. Yeah, but --

MR. KIERNAN: Object to form.

THE WITNESS: But if you do that and find the standard errors fall significantly, it's mostly because of the fact that you have inappropriately ignored correlation that exists in the data. So that's what I'm saying. It's that if you wanted to. You could do lots of different things. But the whole point is his analysis is not valid. He assumes that they are independent regardless of which one you did.

BY MS. SWEENEY:

Q. If you have a situation where there are no clusters and you do this kind of clustering analysis, is that harmless error in your view?

A. You know, I'm going to say, I think it's, I think it's -- if you suspect there is correlation, which we have strong reason to do in this case, economics tells us there will be correlation, the data tells us there is correlation, the documents in this case tells us there is going to be correlation, there is just so overwhelming prior that there will be correlation, then I think it's proper to allow for that correlation. I think that is the only prudent approach. To just bury your head in the sand and ignore the fact that they are correlated as Professor Noll does doesn't make any sense.

Q. I'm going to move to strike as nonresponsive, and I'm going to ask the court reporter to read my question back.

MR. KIERNAN: I oppose the motion.

(Record read.)

THE WITNESS: I don't want to -- here is what I

wanted to say is, it's -- the right way to proceed is to allow for the correlation if you think it's going to be there. You wouldn't want to allow for correlation if you had no reason to believe it were there. That doesn't make a whole lot of sense, but when economics tells you there is going to be correlation, when the facts of the case tells you there is going to be correlation, you should allow for that correlation.

Now, is it harmless? I should say it's the opposite of harmless. It's necessary; it's needed. And there is no reason to believe there isn't correlation in this case. You would expect that there would be correlation.

BY MS. SWEENEY:

Q. Okay. Well, listen to my question. If you have a case where there is no reason to believe there is correlation and you conduct this kind of clustering, is that harmless error in your view?

MR. KIERNAN: Objection; asked and answered.

THE WITNESS: I would say it will not -- if we do it -- I don't want to apply it to an abstract world other than this one. In the context like this given the number of clusters that we have, which is upwards of 300 in one case and upwards of 400 in the other case, and given the nature of the data that we have here, I would

say there is very little downside to allowing for that clustering, and the reason is it doesn't create a bias. It doesn't bias your standard errors upward. If anything, it's going to tend to make your standard errors a little bit smaller is what the literature suggests. It's not generally going to lead to an upward bias in your standard errors.

BY MS. SWEENEY:

Q. Now, you added a number of characteristics to Professor Noll's regression; correct?

A. Yes, this -- that's not new to this report. We had done that in the previous report as well.

Q. And are the -- where in your report are those characteristics listed?

A. Well, they are listed in the regression tables themselves.

Q. Okay. Point to me a particular table that --

A. Like just go to -- well, might as well just start at the beginning. Go to Table 1. They are listed at the bottom of the table. Well, not the very bottom, but the bottom of the coefficient portion. See where there is like a white space in the first three columns and then there is numbers in the last two columns? Those would be the additional characteristics.

Q. So the first one is HP_OEM?

A. Correct. And it runs down to log_recharge_hours.

Q. What is HP_OEM?

A. That was the specific iPods that were I believe HP-branded at one point in time, so they were a little bit different. So that allows for it was an OEM iPod as opposed to an Apple iPod. That sort of refers to some specific models.

Q. What about -- what is USB?

A. USB is whether the Ap- -- whether the iPod is USB-compatible. Some of the early generation iPods were not USB-compatible. They were FireWire based.

Q. And so the next one, FireWire, is that sort of the opposite then of USB? Either it has USB or it has FireWire?

A. No. There was some that I think had both, which is why there -- both effects can be there. The majority either had one or the other as I recall, and I don't remember the fraction that had both. There were some particular models as I recall that had both. That's why they are both in there.

Q. So do you know whether all iPods sold to class members during the class period had USB capability?

A. I'd have to go back and check. I don't recall. But that doesn't affect whether you'd want a control for

(Concluded at 12:07 p.m.)

DECLARATION UNDER PENALTY OF PERJURY

Case Name: The Apple iPod iTunes Anti-Trust Litigation
Date of Deposition: 1/8/2014
Job No: 10009199

I, KEVIN M. MURPHY, Ph.D., the witness herein, declare under penalty of perjury that I have read the foregoing in its entirety; and that the testimony contained therein, as corrected by me, is a true and accurate transcription of my testimony elicited at said time and place.

Executed this ____ day of ___________________, 2014, at ____________________________.

_________________________ ______________
KEVIN M. MURPHY, Ph.D. Date

DEPOSITION ERRATA SHEET

Page No. ____ Line No. ____
Change: ____
Reason for change: _____________________________________
Page No. ____ Line No. ____
Change: ____
Reason for change: _____________________________________
Page No. ____ Line No. ____
Change: ____
Reason for change: _____________________________________
Page No. ____ Line No. ____
Change: ____
Reason for change: _____________________________________
Page No. ____ Line No. ____
Change: ____
Reason for change: _____________________________________
Page No. ____ Line No. ____
Change: ____
Reason for change: _____________________________________
Page No. ____ Line No. ____
Change: ____
Reason for change: _____________________________________

_________________________ _____________________
KEVIN M. MURPHY, PH.D. Dated

STATE OF ARIZONA)
) ss.
COUNTY OF MARICOPA)

BE IT KNOWN that the foregoing deposition was taken before me, Cathy A. Miccolis, RPR, a Certified Reporter, Certificate #50068, for the State of Arizona, and by virtue thereof authorized to administer an oath; that the witness before testifying was duly sworn by me to testify to the whole truth; that the questions propounded to the witness and the answers of the witness thereto were taken down by me in shorthand and thereafter reduced to print by computer-aided transcription under my direction; that pursuant to request, notification was provided that the deposition is available for review and signature; that the transcript consisting of pages 233 through 348 is a full, true and accurate transcript of all proceedings and testimony had and adduced upon the taking of said deposition, all done to the best of my skill and ability.

I FURTHER CERTIFY that I am in no way related to nor employed by any of the parties hereto nor am I in any way interested in the outcome hereof.

DATED at Phoenix, Arizona, January 9, 2014.

Cathy A Miccolis

Cathy A. Miccolis, RPR, CRR
Certified Reporter #50068

Exhibit 2

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
OAKLAND DIVISION

THE APPLE iPOD iTUNES) Lead Case No. C 05-00037
ANTI-TRUST LITIGATION)
)
____________________________)
This Document Relates To:)
ALL ACTIONS)
)
____________________________)

CONFIDENTIAL - ATTORNEYS' EYES ONLY
VIDEOTAPED DEPOSITION OF ROBERT H. TOPEL, Ph.D.
VOLUME II
January 08, 2014
Phoenix, Arizona

Reported By:
Cathy A. Miccolis
RPR, CRR, CSR No. 50068
Job No. 10009199

A P P E A R A N C E S

For the Plaintiffs: Bonny Sweeney, Esq.
ROBBINS GELLER RUDMAN & DOWD, LLP
655 West Broadway
Suite 1900
San Diego, CA 92101
619.231.1058
bonnys@rgrdlaw.com

For the Defendant Apple, Inc.:
David C. Kiernan, Esq.
JONES DAY
555 California Street
26th Floor
San Francisco, CA 94104
415.626.3939
dkiernan@jonesday.com

Also Present: Thomas C. Tracy, videographer

I N D E X

E X H I B I T S

THE VIDEOTAPED DEPOSITION OF ROBERT TOPEL, Ph.D., VOLUME II, was continued on January 8, 2014, commencing at 12:56 p.m. at the offices of BONNETT, FAIRBOURN, FRIEDMAN & BALINT, P.C., 2325 East Camelback Road, Suite 300, Phoenix, Arizona, before CATHY MICCOLIS, a Certified Reporter in the State of Arizona.

THE VIDEOGRAPHER: We are now on the record. The time is approximately 12:56 p.m. Today's date is January 8, 2014. My name is Tom Tracy of Aptus Court Reporting. The court reporter is Cathy Miccolis of Aptus Court Reporting, located at 600 West Broadway, Suite 300, San Diego, California 92101.
This begins the videotaped deposition of Robert Topel, Volume II, testifying in the matter of the Apple iPod iTunes Antitrust Litigation pending in the District Court of California, Division of Oakland, Case Number C 05-00037 YGR, taken at 2325 East Camelback, Suite 300, Phoenix, Arizona 85016.
Counsel, will you please identify yourself and whom you represent for the record at this time, starting with the plaintiffs' counsel.
MS. SWEENEY: Bonny Sweeney for the plaintiffs.
MR. KIERNAN: David Kiernan for Apple.
THE VIDEOGRAPHER: Thank you, Counsel. The

Q. Yeah.

[redacted]

Q. Okay. Another question, I asked you if you'd looked at the percentages of iPod buyers who were first-time versus repeat buyers during other periods, and you mentioned your recollection. Did you look at all at the percentages in the period more than five months after September 2005?

THE WITNESS: Could you read back the question?

(Record read.)

THE WITNESS: Yes.

BY MS. SWEENEY:

Q. And what did you find?

A. I found that more than five months after that I found that 14 percent number.

MR. KIERNAN: Can you take maybe a short break? This thing is forcing a shutdown.

MS. SWEENEY: Let's take a break. I wouldn't mind a break anyway.

THE VIDEOGRAPHER: We are going off the record. The time is approximately 2:07 p.m.

(Recess taken at 2:07 p.m.; resumed at 2:16 p.m.)

THE VIDEOGRAPHER: We are going back on the record. The time is approximately 2:16 p.m.

BY MS. SWEENEY:

Q. You said that you read Professor Wooldridge's declaration?

A. Yes.

Q. And were you familiar with his work before reading the declaration?

A. I had probably read something that he had done before the declaration, yeah.

Q. And do you consider yourself an expert in cluster samples?

MR. KIERNAN: Object to form.

THE WITNESS: Yes, as an applied econometrician I do.

BY MS. SWEENEY:

Q. Have you published any articles or books about cluster samples?

A. No. It's mainly in my teaching and using it in my research and using it in my empirical analysis.

Q. And have you in your own work used a cluster adjustment in the case where you have a whole population instead of a sample?

A. Yes.

Q. And are those just the instances that you described in your last deposition?

A. Two months. No.

Q. Maybe I didn't -- I don't think my question was clear. I think you testified about those instances in your last deposition, and I'm just asking whether you did anything in addition to the ones that you've already described.

A. Well, I don't recall what I described in my last deposition, but it's certainly possible that in another matter I clustered on a characteristic.

Q. Okay. What would that be?

A. Well, I have to be careful. I think this is filed under seal, so I can't describe all the details of the case, but it involved earnings and whether certain things had statistically significant impacts in an earnings regression, and oddly enough, the other side thought that I had not taken into account the correlation in the residuals for people from -- for group people, and so they had a statistician say that I should, and in fact I had. So they thought I had made the mistake that Professor Noll made and I hadn't.

Q. And that you said was an entire population of data?

A. Yes.

Q. But I know you can't tell me details about the case, but it involved wages?

A. Compensation, yeah.

Q. Compensation. So is it fair to say that you disagree with Professor Wooldridge that clustering is not appropriate either in the case of a randomly drawn sample or a total population?

MR. KIERNAN: Object to form.

THE WITNESS: That's fair to say, yes.

BY MS. SWEENEY:

Q. And can you identify for me the bases for that statement, for your opinion?

A. Well, sure. For one thing when Professor Wooldridge had a population in at least two instances where he had a population, he clustered his -- he clustered. He based his opinion that -- on the fact that common factors would be affecting people in different groups in a similar way, and so he clustered on those groups, and so now he is -- that stuff is in his written teaching and in his research, and now he has turned 180 degrees in saying he wouldn't do that because he is testifying in this case. I just think it's very odd and totally inconsistent.

Q. I'm going to move to strike as nonresponsive.

MR. KIERNAN: Oppose the motion.

BY MS. SWEENEY:

Q. Is it your view that one must cluster in every

DECLARATION UNDER PENALTY OF PERJURY

Case Name: The Apple iPod iTunes Anti-Trust Litigation

Date of Deposition: 1/8/2014

Job No: 10009199

I, ROBERT TOPEL, Ph.D., the witness herein, declare under penalty of perjury that I have read the foregoing in its entirety; and that the testimony contained therein, as corrected by me, is a true and accurate transcription of my testimony elicited at said time and place.

Executed this ____ day of __________________, 2014, at ______________________________.

______________________ __________________

ROBERT TOPEL, Ph.D. Date

DEPOSITION ERRATA SHEET

Page No. ____ Line No. ____

Change: ______________________________________

Reason for change: ______________________________

Page No. ____ Line No. ____

Change: ______________________________________

Reason for change: ______________________________

Page No. ____ Line No. ____

Change: ______________________________________

Reason for change: ______________________________

Page No. ____ Line No. ____

Change: ______________________________________

Reason for change: ______________________________

Page No. ____ Line No. ____

Change: ______________________________________

Reason for change: ______________________________

Page No. ____ Line No. ____

Change: ______________________________________

Reason for change: ______________________________

Page No. ____ Line No. ____

Change: ______________________________________

Reason for change: ______________________________

________________________ ____________________

ROBERT TOPEL, Ph.D. Dated

STATE OF ARIZONA)
) ss.
COUNTY OF MARICOPA)

BE IT KNOWN that the foregoing deposition was taken before me, Cathy A. Miccolis, RPR, a Certified Reporter, Certificate #50068, for the State of Arizona, and by virtue thereof authorized to administer an oath; that the witness before testifying was duly sworn by me to testify to the whole truth; that the questions propounded to the witness and the answers of the witness thereto were taken down by me in shorthand and thereafter reduced to print by computer-aided transcription under my direction; that pursuant to request, notification was provided that the deposition is available for review and signature; that the transcript consisting of pages 194 through 264 is a full, true and accurate transcript of all proceedings and testimony had and adduced upon the taking of said deposition, all done to the best of my skill and ability.

I FURTHER CERTIFY that I am in no way related to nor employed by any of the parties hereto nor am I in any way interested in the outcome hereof.

DATED at Phoenix, Arizona, January 9, 2014.

Cathy A. Miccolis, RPR, CRR
Certified Reporter #50068

Exhibit 3

Proving Antitrust Damages

Legal and Economic Issues

Second Edition

ABA SECTION OF ANTITRUST LAW

ABA AMERICAN BAR ASSOCIATION
Defending Liberty
Pursuing Justice

This volume should be officially cited as:

ABA SECTION OF ANTITRUST LAW,
PROVING ANTITRUST DAMAGES: LEGAL AND ECONOMIC ISSUES
(2D ED. 2010)

Cover design by ABA Publishing.

Printed in the United States of America.

ISBN: 978-1-60442-878-0

Discounts are available for books ordered in bulk. Special consideration is given to state bars, CLE programs, and other bar-related organizations. Inquire at ABA Publishing, American Bar Association, 321 N. Clark St., Chicago, Illinois 60610.

14 13 12 11 10   5 4 3 2 1

www.ababooks.org

CHAPTER 6

ECONOMETRICS AND REGRESSION ANALYSIS

A. Introduction to Econometrics and Regression Analysis

Damages analyses require estimates of the impact of an allegedly anticompetitive act on specific market outcomes. However, market outcomes often result from a complex interaction among a large number of factors. For example, once liability has been established in an antitrust price-fixing case, the questions are (1) whether the alleged conspiracy affected prices (the fact of injury) and, if so, (2) the extent to which prices have increased as a result of the alleged conspiracy (the amount of injury).[1] Answering these questions is complicated by the fact that the prices prevailing in the market likely were affected by a wide variety of other demand and supply factors unrelated to the alleged conspiracy. Isolating and measuring the effect of the alleged conspiracy on price requires properly accounting for these other factors. Otherwise, one might confuse the effects of one or more of these other factors with the effects of the alleged conspiracy. Econometric and regression analyses are particularly useful in separating the impact of an alleged anticompetitive act on market outcomes (such as pricing) from the impact of other influences.[2]

Econometric techniques are statistical methods designed to uncover the relationship between a dependent variable (the object of the analysis, such as prices) and one or more explanatory variables (the potential influences on the dependent variable, such as price fixing, costs, or demand factors). In particular, econometric techniques can provide an estimate of the partial effect of each of the explanatory variables on the dependent variable. An explanatory variable's partial effect is the change in the dependent variable that would result from a change in that explanatory variable, holding all of the other explanatory variables

---

1. *See* John E. Lopatka, *Antitrust Injury and Causation*, *in* III ABA SECTION OF ANTITRUST LAW, ISSUES IN COMPETITION LAW & POLICY 2299, 2311 (2008); Theon van Dijk & Frank Verboven, *Quantification of Damages*, *in* III ABA SECTION OF ANTITRUST LAW, ISSUES IN COMPETITION LAW & POLICY 2331, 2334 (2008).
2. *See* Jonathan B. Baker & Daniel L. Rubinfeld, *Empirical Methods in Antitrust: Review and Critique*, 1 AMER. L. & ECON. REV. 386, 388 (1999).

constant. Thus, when correctly implemented, econometric techniques can isolate and measure the effect of a single explanatory factor on the economic outcomes that are relevant when estimating damages.[3]

Even sophisticated econometric techniques are of no use if the partial effect of interest is not identified (i.e., distinguishable).[4] Identification is often a function of the available data. In many situations, identification requires that the existence (or the extent) of the alleged unlawful conduct varied over the observations in the data in a way that is not perfectly mirrored by ("correlated" with) other explanatory variables. If such data exist and there are adequate controls for the effects of other economic factors, identification generally will be possible.[5]

Assuming liability and given the construction of a reasonable explanatory variable to represent the anticompetitive act, properly applied econometric techniques can provide a reliable estimate of damages based on a statistical estimate of the partial effect of the anticompetitive act on the dependent variable of interest, such as price. These techniques also can provide useful information about the important damage-related issue of causation, which is linking the assumed anticompetitive acts to adverse outcomes for competition and for plaintiffs. In particular, econometric analyses can test whether the data are consistent or inconsistent with the proposition that the relevant explanatory variable "caused" changes in the dependent variable.[6] However, econometric analysis cannot demonstrate causality in the strictest sense as can a randomized experimental design, such as that used in laboratory experiments.[7]

One frequently used econometric methodology is the before-during approach. It can be used when data exist from both the period before an alleged anticompetitive act (e.g., a price-fixing conspiracy) and the period of the alleged anticompetitive act. This approach identifies the effect of an alleged anticompetitive act by comparing, for example, prices in the "during" period (the affected period) to prices in the "before" period (the control period), while also controlling for differences in other economic factors between the during and before periods. "Before-during" is used here to distinguish it from a "before-during-after" approach. This latter approach can be used when data exist

---

3. *See* JEFFREY M. WOOLDRIDGE, ECONOMETRIC ANALYSIS OF CROSS SECTION AND PANEL DATA 3-4 (2002).
4. *Id.* at 52-53.
5. *Id.*
6. *Id.* at 3-4. *See also* Chapter III.C.
7. *See* WOOLDRIDGE, *supra* note 3, at 3-4.

not only from before and during, but also from after, the alleged unlawful conduct.[8]

Another frequently employed methodology uses data from a *benchmark* or *yardstick* market,[9] where the alleged unlawful conduct did not occur, as a control group.[10] A benchmark market may be a different area than where the unlawful conduct occurred, a different product, or a different end use market. The idea again is that one can compare the prices or profit margins in the market of interest (the *treatment* market) to prices in the benchmark (control) market, accounting for other economic factors that might have differed between the two markets. This approach may be particularly useful when the data available from the time before the anticompetitive act are limited or unreliable.[11]

Sometimes the benchmark and before-during approaches can be combined into what is called the *difference-in-differences* approach.[12] Under this approach, the before period is used to calibrate the difference between the variable of interest (e.g., price) in the treatment and benchmark markets in the absence of the alleged unlawful conduct. To identify the partial effect of the alleged unlawful conduct, this approach then subtracts the difference between the variable (price) in the market at issue and in the benchmark market in the during period from that difference in the before period (i.e., the difference in the differences).[13] For example, suppose prices were $10 higher on average in the affected market than in the benchmark market before the alleged conduct, but that differential increased to $15 during the alleged conduct, controlling for

8. *See* van Dijk & Verboven, *supra* note 1, at 2335.
9. Unless otherwise noted, our use of the term "market" is meant to denote an economic market, and not necessarily an antitrust relevant market.
10. For a discussion of benchmark approaches see, e.g., PHILLIP E. AREEDA ET AL., ANTITRUST LAW: AN ANALYSIS OF ANTITRUST PRINCIPLES AND THEIR APPLICATION 483 (2d ed. 2005); Robert H. Lande & James Langenfeld, *The Perfect Caper? Private Damages and the Microsoft Case*, 69 GEO. WASH. L. REV. 902, 903, 907, 908-09 (2001). A benchmark approach or before-during-after approach is similar to a natural experiment, in that both rely on a control group or event that is economically similar to the market being studied, but is unaffected by the alleged anticompetitive act. *See* Mary Coleman & James Langenfeld, *Natural Experiments*, *in* I ABA SECTION OF ANTITRUST LAW, ISSUES IN COMPETITION LAW & POLICY 743 (2008).
11. For a discussion of benchmark analyses, see part F of this chapter.
12. *See* WOOLDRIDGE, *supra* note 3, at 130.
13. *Id.* One might still want to control for other economic factors that might have changed between the two periods in the two markets. *Id.*

changes in other economic factors. In this example, an estimate of the effect of the alleged conduct would be a $5 price increase.

When undertaking an econometric analysis, it is important to recognize that data need not be perfect to be reliable. For example, it is not necessary to observe data on every factor that affected the dependent variable. Indeed, if the available data were that complete, econometric methods may not be needed.[14] Nor is it necessary to have perfectly measured variables.[15] Nevertheless, the statistical precision of the econometric results generally will be greater the more accurate and plentiful the available data, and the required econometric techniques generally will be less complex with better quality data.

With sufficiently reliable data and identification, economists have a wide variety of econometric techniques available to them to estimate the partial effect of the alleged unlawful conduct in an antitrust case.[16] The appropriate technique to use in a given circumstance depends on the type of available data, the properties of the underlying economic variables, and the specification of the econometric model used.

This chapter covers a wide range of issues. Not all of the tests or analyses discussed here need to be performed in every damages analysis, and the ones that are important for a particular case will depend on the specific facts of the case.

B. Legal Requirements

The legal requirements for regression analysis fall under the rules for testimony by experts. Under Federal Rule of Evidence 702:

> If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training,

---

14. *See* G.S. MADDALA, INTRODUCTION TO ECONOMETRICS 5 (3d ed. 2001).
15. See Jerry A. Hausman, *Mismeasured Variables: Problems From the Right and Problems From the Left*, 15 J. ECON. PERSPECTIVES 57 (2001), for a discussion of measurement error, and WOOLDRIDGE, *supra* note 3, at 63-64, for a discussion of the use of "proxy" variables.
16. Such techniques include ordinary least squares regression, nonlinear least squares regression, maximum likelihood, instrumental variables estimators, generalized method of moments estimators, and various semi-parametric and nonparametric estimators. For a description of a wide range of econometric techniques, see generally WOOLDRIDGE, *supra* note 3, and WILLIAM GREENE, ECONOMETRIC ANALYSIS (4th ed. 2000).

> **or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.**[17]

Regression analyses have met these requirements many times in litigation for a wide range of issues, including the estimation of antitrust damages.[18] A competent expert should be able to tell the litigant whether sufficient facts and data are available for econometric analysis, explain the types of econometric analyses that are applicable, and reliably apply these econometric techniques.

As suggested above, a key advantage of using econometric analysis is that it allows the expert to separate the impact of an anticompetitive act from other economic factors that affect the variable of interest, such as sales volumes and prices. In general, econometric results will be more reliable as the amount and quality of data increase. If sufficient data are available, econometric analysis has been found necessary for achieving the minimum scientific standard for establishing lost sales and price changes. For example, in *Zenith Electronics Corp. v. WH-TV Broadcasting Corp.*,[19] expert opinion and internal forecasts for sales growth were excluded because data to estimate sales growth via regression analysis were available and not used. But using regression analysis does not guarantee that the analysis will be accepted. The regression must be in a form that assists in determining a material fact, such as the amount of lost sales or the size of price changes.[20] The analysis also must be based on data "reasonably relied upon by experts in the field."[21] Because experts typically verify data for accuracy in consulting work or academic research, experts presenting regression

17. FED. R. EVID. 702.
18. *See, e.g.*, Conwood Co. v. U.S. Tobacco Co., 290 F.3d 768 (6th Cir. 2002); Tuscaloosa v. Harcros Chems., 158 F.3d 548 (11th Cir. 1998); Allapattah Servs. v. Exxon, 61 F. Supp. 2d 1335 (S.D. Fla. 1999) (order denying motions to preclude and to strike expert testimony).
19. 395 F.3d 416 (7th Cir. 2005).
20. *See* Daniel Rubinfeld, *Reference Guide on Multiple Regression, in* FEDERAL JUDICIAL CENTER, REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 186 (2d ed. 2000), *available at* http://www.fjc.gov/public/pdf.nsf/lookup/sciman03.pdf/$file/sciman03.pdf.
21. FED. R. EVID. 703.

analyses in court need to conduct similar verifications of the data that they use.[22]

The principle for reliability encompasses many factors entering the regression analysis, and a competent expert should conduct reliability checks to ensure that the results survive the rigors of litigation. Part D of this chapter discusses many of the issues that arise in implementing reliably econometric techniques and econometric tests that a testifying expert should perform when appropriate. Moreover, econometric results typically should not change materially with minor changes to the data (e.g., deleting a few observations).[23]

C. Causation and Quantification

Courts typically require an analysis of damages that not only quantifies the amount of damages, but also demonstrates that those damages are causally linked to the allegedly anticompetitive acts.[24] In a price-fixing case, for example, there must be an analysis that provides evidence of a clear link between the agreement to fix prices and an increase in prices that is not explained by other factors (i.e., prices would not have been as high but for the agreement).[25]

Properly applied econometric techniques can provide both reliable estimates of the magnitude of damages and useful information about causation. As discussed above, assuming liability and the construction of a reasonable explanatory variable representing the anticompetitive act at issue, econometric tests can be used to determine whether the data are consistent or inconsistent with the anticompetitive act having caused changes in the dependent variable.[26] Most economic data, however, come from observations, not controlled experiments. Therefore, the economist typically has no control over how the data were generated.

The fact that economic data typically do not come from controlled experiments may present certain challenges to econometric analysis. To see this point, consider an investigator conducting a controlled laboratory experiment. Such an investigator can set the values of the factor of interest and apply different levels to different test subjects. For example, a sample of subjects in a pharmaceutical clinical trial can be assigned

22. *See, e.g.*, Cooper v. Travelers Indem. Co., 113 F. App'x 198 (9th Cir. 2004).
23. *See* Rubinfeld, *supra* note 20, at 199.
24. *See, e.g.*, Lopatka, *supra* note 1, at 2299.
25. *See* van Dijk & Verboven, *supra* note 1, at 2335.
26. *See* WOOLDRIDGE, *supra* note 3, at 3-4.

randomly to a "test" group and a "control" group, with the factor of interest (the pharmaceutical "treatment") applied only to the test group. Given this experimental design, if after treatment, the dependent variable changes in the test group in a way that is not found in the control group, then the treatment very likely caused that change and the impact of the treatment can be estimated.[27] Economists rarely have this luxury. Therefore, although econometric techniques applied to observational data can provide evidence consistent with causality and quantification of impact, even sometimes using "control groups" to create natural experiments, econometric analysis typically provides a weaker form of evidence of causality than a randomized experimental design.[28] Nevertheless, (1) a properly specified econometric model showing that an explanatory variable has a statistically significant partial effect on the dependent variable, holding constant other factors, and (2) a sound economic theory explaining why one would expect the explanatory variable to have a causal effect, together provide evidence consistent with the existence of a causal relationship and an estimate of the magnitude of the effect.[29]

For example, assume a properly specified econometric analysis finds a measure of the partial effect of the allegedly anticompetitive act on the dependent variable of interest to be statistically significant (discussed

---

27. The definition of causality has been the subject of philosophical debate. *See* John DiNardo, *Interesting Questions in* Freakonomics, 45 J. ECON. LIT. 973 (2007). However, an adequate definition of causality for the purposes of this chapter is that factor A causes outcome B if the random application of factor A in the population would result in outcome B for the treated segment of the population. *Id.*

28. *Id.*

29. *See* Rubinfeld, *supra* note 20, at 184-85. An econometric test called the *Granger Causality Test* is designed to determine whether changes in one variable occur before changes in another variable. While this property can potentially supply some useful information, it is not the same as causality in the sense in which this chapter uses the term. For example, as PETER KENNEDY, A GUIDE TO ECONOMETRICS 64 (6th ed. 2008), notes, Christmas cards "Granger-cause" Christmas, but obviously do not "cause" Christmas. *See also* Jerry A. Hausman, *Specification and Estimation of Simultaneous Equation Models*, *in* 1 HANDBOOK OF ECONOMETRICS 391, 435-436 (Z. Griliches & M.D. Intriligator eds., 1983).

below) and of the expected sign.[30] This result would be consistent with the act resulting in damages, as long as there is a clear economic explanation of the linkage between the act and the measure of the partial effect. Econometric analysis also can provide an estimate of magnitude of the impact to estimate the amount of damages. A finding that the estimated impact is not statistically different than zero would be consistent with a lack of causation and zero damages.

D. Regression, Hypothesis Testing, and Specification Testing

1. Simple Regression

As discussed above, regression analysis in the context of damage estimation is a statistical method used to analyze the relationship between a dependent variable and a set of explanatory variables based on a sample of observational data. The dependent variable is typically an economic outcome important in estimating any damages sustained by the plaintiff. Examples of such outcomes include price, profit margin, and quantity. The explanatory variables are typically the major economic factors that may explain variation in the dependent variable.[31]

Econometric analyses that reliably estimate the impact of the major influences on the dependent variable begin with econometric modeling. Econometric modeling assumes that the data are generated by underlying economic forces. The econometric model should be designed to be consistent with and incorporate these underlying economic forces.[32] In building the econometric model, the economist makes initial modeling or specification choices by employing economic theory and reasoning. This modeling process identifies the likely major influences on the dependent variable and the potential interactions between those influences. Within the model, the impact of a given economic factor on the dependent variable is typically summarized by a parameter or coefficient that can be estimated from the available data.[33] The modeling choices are very important. Deviations of the econometric model specification from the underlying economic process generating the data can lead to specification error, which make the resulting parameter estimates

30. For example, the "expected sign" would be positive if the allegedly competitive act was a price-fixing conspiracy and the dependent variable of interest was price.
31. *See* KENNEDY, *supra* note 29, at 75.
32. *Id.*
33. *See* KENNEDY, *supra* note 29, at 3-4.

unreliable—even if they are statistically significant.[34] Fortunately, the validity of many specification choices can be tested statistically to ensure that a model is correctly specified. Throughout this chapter, we discuss the rationale for some specification choices and how the validity of specification choices can be statistically tested.

To understand the basics of regression analysis, we begin with the relationship between one dependent and one independent variable, known as a simple regression analysis. Assume that prices are determined by cost, plus some fixed mark-up for profits. Most economic models hypothesize a direct positive relationship between cost and price.[35] Although most economic models do not predict a fixed mark-up for profits and they typically identify a number of other factors that influence price, for purposes of exposition we assume this simplified model. The deterministic or mathematical relationship between cost and price can be described by the simple algebraic equation of:

$$PRICE = \beta_0 + \beta_1 COST\ .$$

Here, the *regression coefficient* β_1 indicates how much price will increase with cost.[36] For example, if price (the dependent variable) increases by \$.90 for every dollar that cost (the exogenous variable) increases, then $\beta_1 = .9$. A second regression coefficient β_0 is a *constant term* reflecting a fixed mark up over changes in costs, say \$1. Price will equal the sum of these two effects in this equation. That is if $COST$ = \$10, then $PRICE$ = \$1 + (.9)x(\$10) = \$10. If $COST$ = \$12, then $PRICE$ = \$1 + (.9)x(\$12) = \$11.80. This relationship is illustrated in Figure 1.

34. *See* WOOLDRIDGE, *supra* note 3, at 50-51.
35. Jonathan B. Baker & Timothy F. Bresnahan, *Economic Evidence in Antitrust*, *in* HANDBOOK OF ANTITRUST ECONOMICS 1, 16 (Paolo Buccirossi ed., 2008).
36. *See* MADDALA, *supra* note 14, at 59-64.

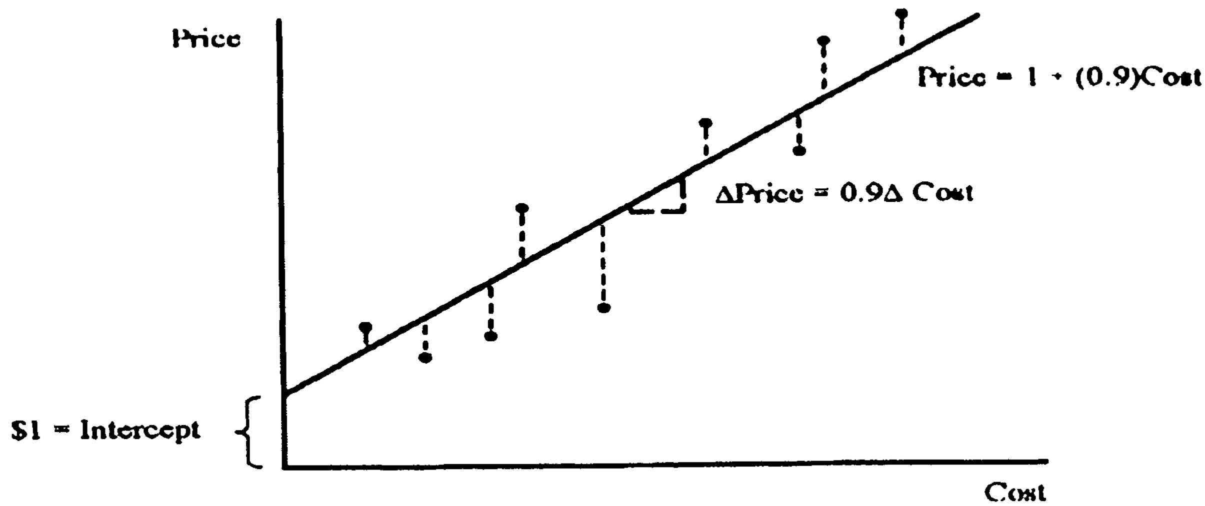


Figure 1

Relationships may be stochastic rather than deterministic. A stochastic relationship is one that has a random component, a component that is not generated from a describable systematic process.[37] A stochastic relationship does not yield a single unique value of the dependent variable (e.g., *PRICE*) for a given value of the explanatory variable (e.g., *COST*). In other words, the dependent variable is influenced, but not uniquely determined, by the explanatory variable. The stochastic factors not accounted for by the explanatory variable are summarized in the *error term*, represented by the variable ε in the equation that defines the dependent variable.[38] The error term is the difference between the observed dependent variable and the portion of that value that is explained by the explanatory variable. In a simple regression model, this stochastic relationship can be expressed as:

$$PRICE = \beta_0 + \beta_1 COST + \varepsilon .$$

Because the error term is not observed by the econometrician, it amounts to statistical "noise" that partially obscures the "true" underlying relationship between the dependent variable and the included

37. *See* KENNEDY, *supra* note 29, at 3.
38. *Id.*

explanatory variable.[39] Specifically, the statistical noise prevents us from being able to measure the regression coefficients β_0 and β_1 from the data exactly. However, regression analysis provides a way to cut through the statistical noise and obtain *estimates* of each of the regression coefficients, although these estimates will not exactly equal the true underlying coefficients (they are, after all, estimates). Intuitively, one may think of the regression estimate of β_1 as being the average observed impact of cost on price based on the data analyzed, recognizing it is not an exact measure.[40] Under certain conditions the coefficient estimates will have the desirable property of being *consistent*, which means they converge to the true coefficient values as the size of the data sample grows.[41]

One method frequently used for estimating the coefficients that summarize the relationship between the explanatory variables and the dependent variable is *ordinary least squares* (OLS).[42] This statistical technique minimizes the squared deviations from a line passing through the data, creating the line that fits the pattern of the data as closely as possible. Specifically, it minimizes the sum of the squares of the vertical distances of the observed values to obtain a line that best fits the data, as illustrated in Figure 1.[43] This technique has the desirable property of being consistent under general conditions.[44]

---

39. The error terms may be caused by unpredictable randomness in the behavior of consumers or businesses, the effect of omitted explanatory variables, or measurement error in the dependent variable. *Id.*
40. The coefficient is often assumed to be the same for each unit of observation (e.g., customer or firm). However, an economist may want to test this assumption if there is reason to believe it does not hold. See part D.7. of this chapter for further discussion.
41. *See* WOOLDRIDGE, *supra* note 3, at 52-54. Coefficient estimates from a regression will be consistent if the error term and the explanatory variables are uncorrelated. *Id.*
42. Other estimation techniques use different criteria to determine the coefficient estimates. Examples of other techniques include generalized method of moments and maximum likelihood estimation. *See* WOOLDRIDGE, *supra* note 3, ch. 12-14.
43. *See* KENNEDY, *supra* note 29, at 48.
44. *See* WOOLDRIDGE, *supra* note 3, at 52-54. OLS moreover will be *unbiased*—correct on average in small samples—under the condition that the conditional expectation of the error term given the explanatory variables is zero. *See* GREENE, *supra* note 16, at 245. Because the concepts of consistency and unbiasedness have similar meanings for

The relationship in this example assumes that the explanatory variable has a linear (i.e., straight line) relationship with the dependent variable, in that a given change in the dollar value of cost translates to a fixed dollar change in price. However, economic theory, statistical evidence, or information from market behavior may indicate that a different functional form for the regression equation may be more appropriate. For example, in certain instances a given percentage increase in costs will yield a proportionate percent increase in price—rather than a fixed dollar relationship. In this case, a regression equation similar to the one described above can be estimated, but the price and cost data would be transformed into logarithms and the cost coefficient would measure the impact of a percentage change in costs on the percentage change in price.[45] In general, there should be an economic or statistical basis for the functional form of the regression equation that is estimated.[46]

2. *Multiple Regression*

As discussed above, one of most compelling reasons for using regression analysis in a study of damages is that multiple regression analysis can help separate the partial effects of various influences on a dependent variable. Multiple regression analysis is a particularly important econometric tool. There typically will be several major influences on a dependent variable, and the econometric model must take these various influences into account in a manner consistent with economic theory and reasoning.

For example, multiple regression analysis might be used to measure the extent to which the price customers paid for a product (PRICE), the dependent variable, is related to three explanatory variables rather than one: (1) the cost of production (COST), (2) the level of industrial production in the downstream industry (DEMAND), and (3) whether the customers purchased the product during the period of an alleged conspiracy (PERIOD).[47] DEMAND is included because, in addition to

---

purposes of this chapter, the terms "consistent" and "unbiased" are used interchangeably in this chapter.

45. *See* KENNEDY, *supra* note 29, at 106.

46. *Id.* at 102.

47. The relevant dependent variable will be determined by the nature of the antitrust allegations, the theory of causation, and specific facts of the case. For example, consider a monopolization case alleging anticompetitive exclusionary conduct through the monopolist knowingly

cost, most economic models predict that the level of demand will affect price. As industry demand increases, prices might be expected to increase. PERIOD is included to account for the alleged anticompetitive act. The allegation in most price-fixing conspiracies is that prices are higher during the conspiracy period, all else equal.[48]

The relationship between the dependent variable and the explanatory variables might take the following form:

$$PRICE = \beta_0 + \beta_1 COST + \beta_2 DEMAND + \beta_3 PERIOD.$$

β_0, β_1, β_2, and β_3 are the regression coefficients associated with the partial effect of each explanatory variable on PRICE (β_0 is, as before, the constant term, β_1 is the coefficient associated with COST, β_2 is the coefficient associated with DEMAND, and β_3 is the coefficient associated with PERIOD).[49]

The coefficient associated with each explanatory variable indicates the amount by which the dependent variable changes when the associated explanatory variable changes, holding the other explanatory variables constant.[50] Again, suppose β_1, the coefficient associated with COST, is equal to 0.90. In a multiple regression context this implies that the customer's price would be higher by \$0.90—holding DEMAND and PERIOD constant. Further assume that β_2 equals \$0.10, so that

---

asserting invalid patents against its existing competitors. If before the resolution of the patent cases customers reacted to the pending litigation by switching their purchases to the monopolist (or failing to switch away from the monopolist), the damages to the competitors could in part be the profits on the sales they lost as a result. Given sufficiently reliable data, an econometric analysis of the amount of switching may be relevant both to causation issues and to the quantification of damages. Also note that a regression model used in practice often will have more than three explanatory variables. A relatively simple model is used here for expository purposes.

48. See Chapter 7 for an overview of the legal and economic issues involved in estimating damages associated with anticompetitive price increases or overcharges. As discussed below, the use of a single "dummy" variable for the conspiracy period may not be appropriate. *See* part E of this chapter.

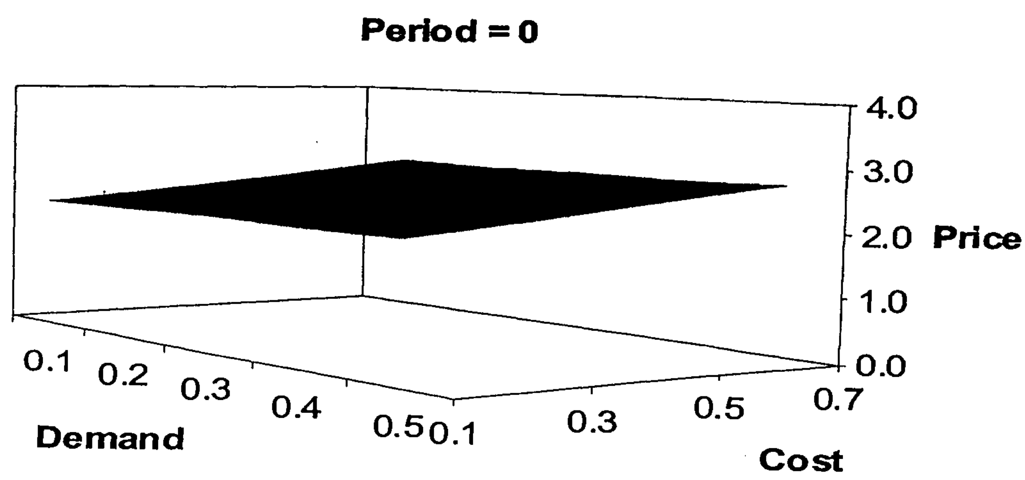
49. *See* MADDALA, *supra* note 14, at 128-34.

50. *See* WOOLDRIDGE, *supra* note 3, at 14-16.

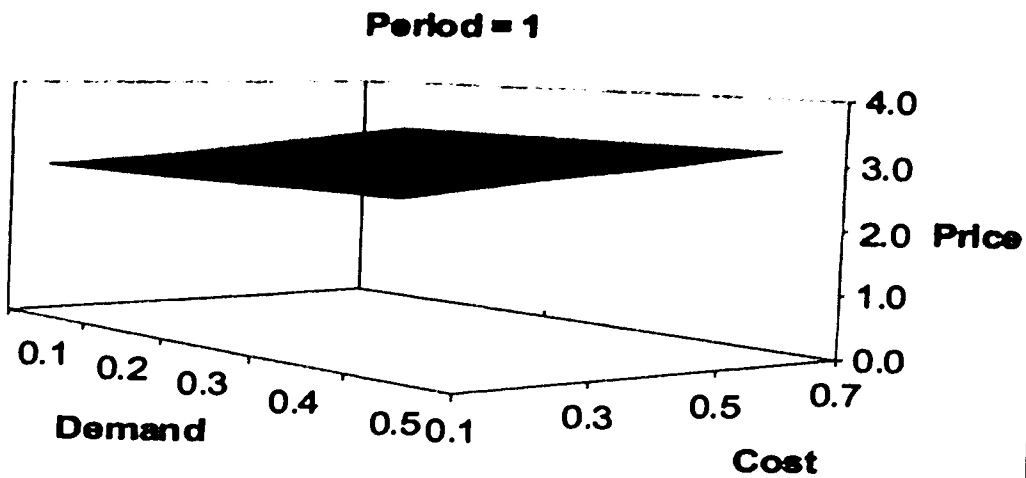
increasing DEMAND by 1 unit would increase price by $0.10, holding COST and PERIOD constant. PERIOD takes on the value of zero in the before period and one in the during period. Similar to the other coefficients, β_3 measures the impact of PERIOD on price, holding COST and DEMAND constant. The example assumes this impact is $0.50 for the during period.

Figure 2 graphically represents how price is related to DEMAND, COST, and PERIOD in three dimensions using the parameters above. The figure is in two parts, one for PERIOD = 0 (the before period) and one for PERIOD = 1 (the during period). As drawn in Figure 2, for each period, the relationship between price, demand, and cost forms a plane. For each point on the plane, increasing cost by $1 increases the PRICE by $0.90 (holding DEMAND and PERIOD constant) and PRICE increases by $0.10 for each increase of one in DEMAND (holding COST and PERIOD constant). The effects of DEMAND and COST are the same whether PERIOD is equal to 0 or equal to 1. The effect of PERIOD is to raise the plane by $0.50 as illustrated by the plane shifting upwards in the PERIOD 1 graph compared to PERIOD 0 graph.

Figure 2: Graphs of Price Equation



PRICE = 2 + 0.1x*DEMAND* + 0.9x*COST* + 0.5x*PERIOD*



PRICE = 2 + 0.1x*DEMAND* + 0.9x*COST* + 0.5x*PERIOD*

As a practical matter, the set of explanatory variables included in a regression model never accounts for all of the factors that affect the dependent variable.[51] When there are many factors at work, there generally will be some factors that the econometrician can identify, measure, and include as explanatory variables, and some factors that cannot be identified and measured and thus cannot be included as explanatory variables. The relationship between the dependent variable, the explanatory variables, and the error term in the example would be expressed as follows:

$$PRICE = \beta_0 + \beta_1 COST + \beta_2 DEMAND + \beta_3 PERIOD + \varepsilon .$$

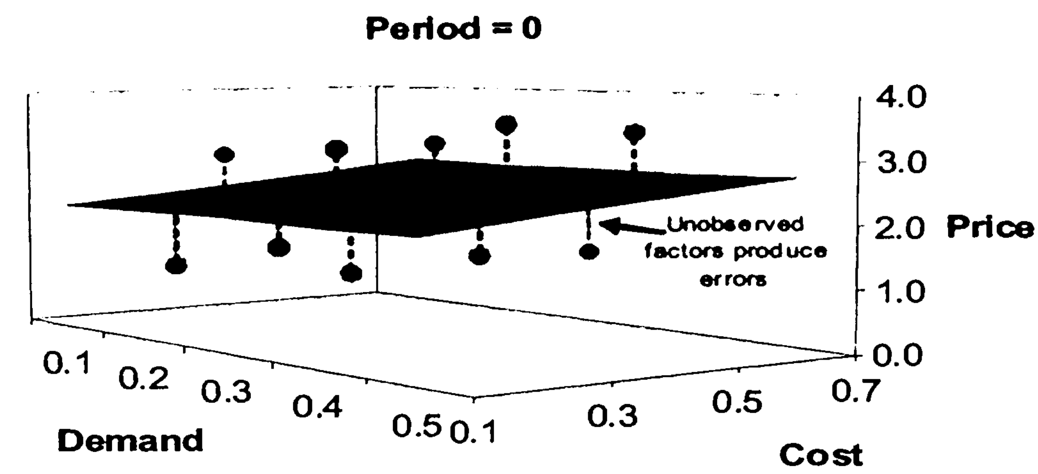
In this example, the unobserved factors that appear in the error term might include particular aspects about the supply and demand conditions facing the customer, such as the end uses to which the customer puts the product, the availability of substitutes for the product (which may differ across customers), the production level of the particular industry in which the customer participates (since the included demand variable

51. *See* KENNEDY, *supra* note 29, at 3.

DEMAND may not be industry-specific), and the customer's negotiating skill if prices are individually negotiated.[52]

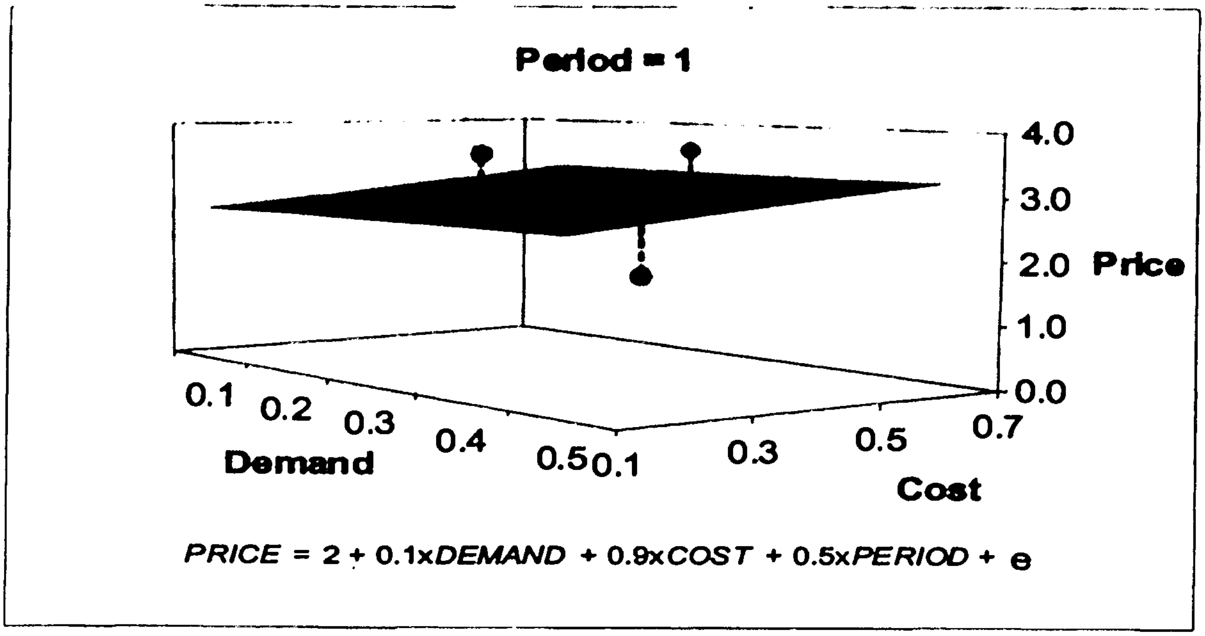
Figure 3 shows the planes discussed above within the actual observations used in the regression, similar to the simple regression discussed above. The coefficients of the equation are estimated by minimizing the squared errors between the data points and the plane in order to get the "best fit" of the plane to the available data.

Figure 3: Estimating Price Equation From Actual Observations



$PRICE = 2 + 0.1 \times DEMAND + 0.9 \times COST + 0.5 \times PERIOD + e$

52. *See, e.g.*, John H. Johnson & Gregory K. Leonard, *Economics and the Rigorous Analysis of Class Certification in Antitrust Cases*, 3 J. COMP. L. & ECON. 341, 346-47 (2007).



3. *Hypothesis Testing*

Hypothesis testing is the use of statistics to assess whether the data are consistent with a specified hypothesis.[53] Typically the hypothesis will be a statement about a coefficient. For example, the hypothesis may be that the coefficient on the PERIOD variable is equal to zero. The hypothesis test will consider whether the estimate of the PERIOD coefficient obtained from the data is consistent with what would be expected if in fact the PERIOD coefficient were zero. As mentioned above, the estimated coefficient for PERIOD likely would not be exactly equal to zero even if the true underlying coefficient were zero. However, the estimated coefficient should not be "far" from zero if the hypothesis were true.

Testing a hypothesis about a coefficient requires a measure of the *statistical precision* of the estimate of the coefficient. If there is a great deal of statistical noise so that a coefficient estimate is highly imprecise, it would provide relatively little information about the value of the true underlying coefficient.[54] As an example, an opinion poll based on a very small sample of respondents generally would not be very precise and thus would not provide a very useful estimate of what percentage of the

53. *See* GREENE, *supra* note 16, at 147.
54. *See* KENNEDY, *supra* note 29, at 67-68.

overall population held the opinion. A larger sample of respondents would produce a more precise estimate, but getting information from the entire population is not necessary to get a reasonably accurate estimate.

To measure the precision of a statistical estimate, econometricians and statisticians typically use what is called a *standard error*.[55] The standard error is best defined by explaining how it is used to create a *95 percent confidence interval*. A 95 percent confidence interval is approximately equal to the range defined by the coefficient estimate plus and minus two times the standard error. One can be "95 percent confident" that the true underlying coefficient lies within the appropriately defined 95 percent confidence interval. Economists often use 95 percent confidence intervals, but they also sometimes use intervals defined by lower or higher levels of confidence, such as 90 percent or 99 percent.[56]

When the coefficient estimate is relatively imprecise, the 95 percent confidence interval must be relatively wide to ensure that it includes the true value of the coefficient with 95 percent confidence. Conversely, when the coefficient estimate is relatively precise, the 95 percent confidence interval can be narrower.[57] Returning to the opinion poll example, pollsters often report the 95 percent confidence interval, e.g., "45 percent of respondents supported the proposal with a margin of error of +/- 5 percent." That is, the pollsters are 95 percent confident that the percentage of the population supporting the proposal is between 40 percent and 50 percent. The margin of error shrinks, all else equal, as the sample size of the poll increases and the precision of the poll increases accordingly. With a margin of error of +/-1 percent, the 95 percent confidence interval for the percentage of the population supporting the proposal would be narrower, from 44 percent to 46 percent.

Standard errors also can be used to conduct hypothesis tests regarding coefficients. Suppose the economist wishes to test the hypothesis that the coefficient on the PERIOD variable is equal to zero. To test this hypothesis, the economist would calculate the ratio of the coefficient estimate to its standard error, or "$\hat{\beta}/se(\hat{\beta})$" where $\hat{\beta}$ is the

55. *See* WOOLDRIDGE, *supra* note 3, at 41.
56. *See* GREENE, *supra* note 16, at 146.
57. *See* KENNEDY, *supra* note 29, at 54.

coefficient estimate and $se(\hat{\beta})$ is the standard error of the coefficient estimate. This ratio is called a *t-statistic*.[58]

If the hypothesis is correct and the true underlying coefficient is in fact zero, then the t-statistic should not be very far from zero. If the t-statistic turns out to be far from zero, it would cast doubt on the truth of the hypothesis. How do we determine whether the t-statistic is "far" from zero? We can calculate the probability that the t-statistic achieves a certain value, if the hypothesis were true. For example, if the hypothesis were true, there is about a 90 percent probability that the t-statistic will fall between 1.7 and -1.7 and about a 95 percent probability that the t-statistic will fall between 2 and -2. Thus, if the hypothesis were true, there would be only a 5 percent probability that the t-statistic we observe would be either greater than 2 or less than -2. Accordingly, if we observe a t-statistic greater than 2 or less than -2, the data would appear to be inconsistent with the hypothesis (because such an outcome is quite unlikely if the hypothesis were in fact true).

Indeed, if the absolute value of the t-statistic that the economist calculates exceeds two, then the hypothesis that the true underlying coefficient equals zero typically would be said to be *rejected* at the 5 percent significance level and the result typically would be termed *statistically significant*.[59] This result often is also expressed by saying that the coefficient is "statistically significantly different from zero (at the 5 percent level of significance)." The 5 percent level of significance (and the corresponding 95 percent confidence interval) is often used by economists and statisticians when conducting hypothesis tests, but other levels of significance, such as 1 percent or 10 percent, are also sometimes used.

As an example of these techniques of statistical inference, suppose that the coefficient estimate on the PERIOD variable in the price regression was 0.50, which would imply that prices were \$0.50 higher during the alleged conspiracy period as compared to outside that period, holding constant the variables COST and DEMAND (and assuming correct model specification). Suppose further that the standard error of the coefficient estimate on PERIOD is 0.20. In this case, the 95 percent confidence interval would be approximately \$0.10 to \$0.90—one can be 95 percent confident that the true underlying coefficient on PERIOD lies

58. *See* GREENE, *supra* note 16, at 249.
59. *Id.*

within this interval, assuming the validity and reliability of the econometric model and approach.

Similarly, the standard error can be used to test the hypothesis that the coefficient on the PERIOD value is zero, which implies that price was not higher during the alleged conspiracy period (holding constant COST and DEMAND). This test would be conducted by calculating the t-statistic: 0.50/0.20 = 2.50. Since the calculated t-statistic of 2.50 is greater than two, the coefficient on PERIOD would be said to be statistically significantly different from zero at the 5 percent level of significance, and the hypothesis that the price was no higher during the alleged conspiracy period would be rejected.

4. *Estimation of the Standard Errors*

Correct statistical inference requires not only good estimates of the coefficients of the model, but also good estimates of the standard errors of these coefficient estimates.[60] For example, a t-test of whether a coefficient β is zero is conducted by forming the t-statistic "$\hat{\beta}/se(\hat{\beta})$," discussed in the last section. This t-statistic can be invalid and lead to incorrect statistical inference if the standard error of the coefficient estimate, $se(\hat{\beta})$, is itself inconsistently estimated.

Consistent estimation of the standard errors requires that the properties of the error term of the regression be properly taken into account. For example, standard errors frequently are estimated assuming the error term of the regression is *independently and identically distributed*. That is, the error for each observation reflecting the impact of unmeasured factors is assumed to be from the same distribution (or pattern) of possible errors, and each error is statistically independent of the others (they are not correlated to each other).[61] If in fact the error terms are *correlated* with each other (i.e., not independent) or not identically distributed, then the resulting standard error estimates generally will be inconsistent.[62]

Correlation among the errors of different observations can arise in various situations. For example, suppose the data sample is a *time series*, i.e., the data were generated by observing the variables (e.g., price, cost,

---

60. *See* WOOLDRIDGE, *supra* note 3, at 57.
61. *Id.* at 54-57.
62. *Id.*

and demand) at various points over time (e.g., on a monthly basis).[63] In such a case, the error in one month might well be related to the errors in adjacent months, since the unobserved economic factors that appear in the error term might themselves exhibit correlation over time. This correlation of errors over time is called *serial correlation*.[64]

As another example, suppose the data sample is a *cross-section/time series*, or *panel data* set, where the variables are observed at various points of time separately for each of a number of units of observation such as individual customers.[65] Each customer's data are a time series. Therefore, the error terms for a given customer may exhibit serial correlation. In addition, each customer may have idiosyncratic factors that affect the price it paid, but are unobserved in the data. These factors would be present in all of the errors across time for that customer, which would be a further cause of correlation among the errors for a given customer. This effect is called an *unobserved individual-specific effect* (where the "individual" refers to the unit of observation, e.g., a customer).[66]

Importantly, the correlation among errors need not be confined to errors that pertain to the same customer. For example, the error terms for all customers within the same time period also may be correlated. Unobserved economic factors may affect all customers' prices at a given point in time and therefore these common factors will appear in the errors of all of the customers in a given time period. Similarly, if these unobserved factors are themselves serially correlated, then the error for one customer in one month will be correlated with the error for another customer in another month. Therefore, there may be correlation among the errors both within and between units of observation in a panel data set.[67]

There can be substantial consequences from estimating the standard errors for the coefficient estimates as if the errors were uncorrelated when they are in fact correlated. With positive correlation between the error terms, the incorrectly estimated standard errors generally will be

---

63. *See* GREENE, *supra* note 16, at 97.
64. *Id.* at 525.
65. *Id.* at 98.
66. *See* WOOLDRIDGE, *supra* note 3, at 248.
67. This problem is widely recognized in the econometrics literature. *See* Brent R. Moulton, *An Illustration of the Pitfall in Estimating the Effects of Aggregate Variables on Micro Units*, 72 REV. ECON. STAT. 334 (1990); Marianne Bertrand et al., *How Much Should We Trust Differences-in-Differences Estimates?*, 119 Q. J. OF ECON. 249 (2004).

biased downward, making the regression coefficients seem to be more precisely estimated than they really are. As a result, a statistical test on the coefficients may yield what appears to be a statistically significant result but is not.[68]

To see why this is so, suppose that the estimate of the coefficient on a price-fixing conspiracy dummy variable is 0.15 and that the standard error, estimated by incorrectly ignoring correlation among the errors, is 0.05. The (incorrect) t-statistic would then be 0.15/0.05 = 3 and the hypothesis that the alleged conspiracy had no effect on prices would be strongly rejected. But, the standard error is too low because it did not account for correlation in the error terms. Suppose that the standard error is re-estimated correctly to account for correlation in the error terms, and this correct standard error is 0.15. Then, the correctly calculated t-statistic is only 0.15/0.15 = 1. Because the correct t-statistic is less than two, the hypothesis that the alleged conspiracy had no effect would not be rejected at conventional levels of statistical significance. If the standard error had not been corrected, the wrong inference would have been made.

Econometricians have a variety of methods for consistently estimating the standard errors when correlation among the errors exists. In a time series context (discussed in more detail below), various non-parametric procedures (procedures that do not impose any functional form on the correlation) have become widely used.[69] In a panel data context (also discussed in more detail below) these procedures may be used, and other methods have been proposed as well.[70] Some of these panel data procedures are easily implemented.[71]

---

68. *Id.*
69. *See* Donald W. Andrews, *Autocorrelation and Heteroskedasticity Consistent Covariance Matrix Estimation*, 59 ECONOMETRICA 817 (1991); Donald W. Andrews & J. Christopher Monahan, *An Improved Autocorrelation and Heteroskedasticity Consistent Covariance Matrix Estimator*, 60 ECONOMETRICA 953 (1992). The resulting standard errors are also consistent in the presence of *heteroskedasticity* (the variance of the error term differs across observations). Many statistical software packages implement the Newey-West procedure, which is one example of such a procedure to obtain autocorrelation and heteroskedasticity consistent standard errors. *See* Whitney Newey & Kenneth West, *A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix*, 55 ECONOMETRICA 703 (1987).
70. *See* Bertrand et al., *supra* note 67.
71. For example, Stata, a popular econometrics software package, includes a "cluster" option for calculating standard errors assuming unspecified

These procedures produce consistent estimates of the standard errors even when there is no correlation among the error terms. In other words, they work well in both situations. Thus, these procedures have become used more generally in practice.[72] When there is good reason to suspect the existence of correlation among the errors, such procedures should be used to avoid making incorrect statistical inferences.[73]

Finally, heteroskedasticity can also create problems in accurately measuring standard errors. Heteroskedasticity occurs when the variance of the error term varies across observations.[74] This condition is another violation of the independent and identically distributed error term assumption that can cause the traditional standard error calculation to be inconsistent. Again a well-known and widely used technique exists for calculating standard errors that are robust to heteroskedasticity (White standard errors).[75] This technique is also easily implemented in many econometric packages.[76]

5. *Choice of Explanatory Variables*

As discussed above, the explanatory variables in an econometric model represent economic factors that influence the dependent variable.[77] An important question is which explanatory variables to include in the model. Answering this question should begin with economic theory combined with qualitative knowledge about the industry. For example, if the dependent variable is price, economic theory suggests that demand drivers, cost factors, and industry capacity, among other things, are potential explanatory variables.[78] Industry knowledge would suggest specific variables that would appropriately represent these factors. If a

---

within-group (cluster) correlation between the error terms. *See* 3 STATA PRESS, BASE REFERENCE MANUAL 81 (2007).

72. *See* GREENE, *supra* note 16, at 465 ("The . . . Newey-West estimator [is] becoming ubiquitous in the econometrics literature").
73. If one is not sure, there are various tests one can run to test the hypothesis of no correlation in the error terms. *See* WOOLDRIDGE, *supra* note 3, at 130, 279, 420-449; GREENE, *supra* note 16, at 538-42.
74. *See* GREENE, *supra* note 16, at 499.
75. Hal White, *A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity*, 48 ECONOMETRICA 817, 838 (1980).
76. For example, Stata includes an option for calculating White standard errors. *See* STATA, *supra* note 71, at 81.
77. *See* part A of this chapter.
78. *See* Baker & Rubinfeld, *supra* note 2, at 391.

product is used as an input by downstream industries, the level of production in those industries might drive the demand for the product.[79] The prices of the inputs used to produce the product could be important to costs.

a. Too Many or Too Few Variables?

The number of explanatory variables suggested by economic theory and industry knowledge often will be large. Is it best to include all of the explanatory variables, or should one try to pare back the number of variables in order to have a simpler model? The downside to including extraneous explanatory variables in a regression is that the coefficients may be less precisely estimated. However, these estimates will still be unbiased.[80] Moreover, the effect on precision of having additional variables will often be small when the sample size is large.

Mistakenly excluding important explanatory variables in an attempt at simplicity, on the other hand, can result in an *omitted variable bias.* Omitted variable bias arises when important explanatory variables that have been omitted from the regression model are correlated with included explanatory variables. Because the omitted variables are in the error term, the result will be a correlation between the included explanatory variables and the error term.[81] This misspecification will bias the resulting coefficient estimates, and make these estimates

---

79. Care needs to be taken that the explanatory variables used in a least squares regression are *exogenous*, or uncorrelated with the error term, to the extent possible. *See* WOOLDRIDGE, *supra* note 3, at 50-51. For example, if the intermediate product in question represents a large share of the downstream industry's costs, the amount of downstream production may be affected by the price of the intermediate good. If the impact of the price of the intermediate good on the sales of the downstream product is substantial, then downstream production could be considered *endogenous* (correlated with the error term) rather than exogenous. *See* Baker & Rubinfeld, *supra* note 2, at n.17. Methods of detecting and dealing with endogeneity are discussed later in this chapter.

80. More specifically, ordinary least squares is still the best linear unbiased estimator as demonstrated by the Gauss-Markov Theorem, one of the more famous theorems in statistics. *See* GREENE, *supra* note 16, at 246. This means that not only is least squares unbiased, but also it is the most efficient (i.e., most precise) among linear unbiased estimators.

81. *See* WOOLDRIDGE, *supra* note 3, at 61-62.

unreliable for damage estimation. This bias does not diminish as sample size increases.[82]

To see why, suppose that in the regression model discussed above, there is a second important demand driver variable, DEMAND2, that was omitted from the regression model. Suppose further that this variable was positively correlated with the alleged conspiracy period variable PERIOD. For example, DEMAND2 might be unusually high during the period of the alleged conspiracy, so DEMAND2 and PERIOD would be highly positively correlated. All else equal, when DEMAND2 was unusually high, PRICE should be unusually high. A regression model that included PERIOD, but did not include DEMAND2 would mistakenly attribute the effects of DEMAND2 (which was omitted from the model) to PERIOD, the variable that represents the alleged anticompetitive act. As a consequence, the regression would lead to the mistaken conclusion that the alleged conspiracy represented by PERIOD was the "cause" of the high prices, when in fact the cause was the unusually high values of DEMAND2 during the alleged conspiracy period. This bias will only occur, however, if DEMAND2 is empirically important for the dependent variable PRICE and correlated with PERIOD.

In general, one should be more concerned with avoiding bias than with improving precision. For these reasons, it is in general better to include more variables than fewer.[83] Nonetheless, no variable should be included unless there is an economic rationale for putting it in the model.

When there is concern that extraneous variables could greatly affect the precision with which coefficients can be estimated, a parsimonious model may have virtue.[84] There are several model selection criteria one can use to compare models containing different sets of explanatory variables.[85] Examples are the *Akaike Information Criterion* (AIC) and the *Schwartz Criterion* (SC).[86] These criteria reward increased explanatory power of the model, and penalize the addition of extraneous variables. When the penalty from the addition of a sufficiently weak explanatory variable exceeds the reward, the criteria will indicate that the variable should not be added. There are other sophisticated econometric methods that attempt explicitly to trade off bias against increased

82. *Id.*
83. *See* KENNEDY, *supra* note 29, at 95.
84. *Id.*
85. *Id.* at 206.
86. *Id.*

precision of the estimates.[87] However, these techniques should still only consider variables that economic theory and industry information identify as likely to be important and should not be used in a mechanical way.

b. Multicollinearity

Multicollinearity occurs when two or more of the explanatory variables are highly correlated with each other ("collinear").[88] When this happens, it may be difficult for the regression to distinguish which of the collinear variables is "important." In that case, the coefficient estimates on the collinear variables may be highly imprecise; that is, they may have large standard errors. It then becomes difficult to make statistical inferences about the true underlying coefficients. Dropping one or more of the collinear variables is sometimes argued to solve the multicollinearity "problem." This proposition is questionable and can lead to a significant bias in the estimates of the coefficients on the retained variables.[89] Dealing with multicollinearity requires recognizing several important points.

First, multicollinearity is a matter of degree. Explanatory variables are almost always correlated with each other to some extent; the question is, is the multicollinearity severe? Since multicollinearity is a characteristic of the sample of data rather than a characteristic of the underlying population, there is no rigorous statistical test for multicollinearity.[90] A number of indicators of multicollinearity have been developed, but many of these have shortcomings.[91]

---

87. Such techniques are often called "shrinkage" estimators. *See* KENNEDY, *supra* note 29, at 198. Because these techniques lead to biased estimators, there is no consensus among economists as to whether they should be used. *Id.* at 201.
88. Correlation coefficients measure how closely two data series are related to each other. When the two series form a perfectly straight line in a positive relationship ($y = \alpha x$), the correlation coefficient will be 1. When the two series form a perfectly straight line in a negative relationship ($y = -\alpha x$), the correlation coefficient will be -1. More generally, when the two series do not form a perfectly straight line but have random variations from a straight line ($y = (+/-)\alpha x + \varepsilon$), the correlation coefficient will be between -1 and 1. *See* GREENE, *supra* note 16, at 75.
89. *See* KENNEDY, *supra* note 29, at 197.
90. DAVID A. BELSLEY ET AL., REGRESSION DIAGNOSTICS: IDENTIFYING INFLUENTIAL DATA AND SOURCES OF COLLINEARITY 95 (1980). Statistical tests allow one to make an inference about a population

Second, multicollinearity among a subset of the explanatory variables may affect the precision with which the coefficients on those explanatory variables are estimated, but not affect the precision with which the coefficients on other explanatory variables are estimated.[92] Accordingly, even severe multicollinearity among a subset of explanatory variables may not affect the coefficient estimates that are of most interest in the analysis. In a price-fixing case, for example, interest typically focuses on the coefficient on the alleged conspiracy variable. Multicollinearity that affects the precision with which the demand variable coefficients can be estimated should be of little concern if it does not affect the precision with which the alleged conspiracy variable coefficient can be estimated. For this reason, a useful indicator for multicollinearity is that proposed by Belsley, Kuh, and Welsch.[93] This indicator focuses on identifying which explanatory variable coefficient estimates are potentially affected by the multicollinearity.

Third, if the precision of the coefficient estimate on the variable of primary interest (e.g., the alleged conspiracy variable) is not substantially affected by the multicollinearity, there is no reason to drop any variables that have a strong economic rationale for inclusion. As discussed above, dropping variables that are truly important could lead to omitted variable bias. A large change in the alleged conspiracy variable coefficient estimate after dropping collinear variables could indicate omitted variable bias, which would render the estimate of the alleged conspiracy variable coefficient unreliable. Therefore, in general the best course of action in that case would be to retain the collinear variables.[94]

Now assume the variable of interest is affected by the multicollinearity. Would it now be valid to drop the other presumably economically important variable(s) with which the variable of interest is collinear? Generally no valid basis exists for choosing to retain the variable of interest as opposed to retaining the variable(s) with which it is collinear. The multicollinearity and the imprecision of estimation that results indicates that it is difficult to separate out the individual effects of

---

characteristic from the sample; but multicollinearity is a sample characteristic, not a population characteristic.

91. *See* KENNEDY, *supra* note 29, at 195.
92. *See* BELSLEY ET AL., *supra* note 90, at 91.
93. *See id.* Other indicators are "Klein's Rule of Thumb" and the "variance inflation factor." *See* GREENE, *supra* note 16, at 258.
94. *See* KENNEDY, *supra* note 29, at 197.

the collinear variables.[95] In that situation, there may be no "solution" other than gathering more data, which often may not be feasible. With severe collinearity between the variable of interest and other economically important explanatory variables, the coefficient on the variable of interest cannot be estimated precisely, and the regression model may therefore be unreliable for estimating damages. Dropping explanatory variables to reduce the multicollinearity without a sound economic reason does not fix the problem.[96]

6. *R-Squared and Other "Goodness of Fit" Measures*

Most regression software packages generate so-called goodness of fit measures. These measures describe how well a statistical model fits a set of data. This discussion concentrates on the most common goodness of fit measure, *R-squared* (or R^2). R-squared is the percentage of the total variation in the dependent variable that is explained by the regression model. A higher R^2 means that the model explains more of the variation in the dependent variable. In the extreme, if the model were a perfect fit, the R^2 would attain its maximum possible value of 1.[97] If the model explained none of the variation in the dependent variable, the R^2 would attain its minimum possible value of 0.

A very small R^2 may raise concern about whether the model is missing important explanatory variables. In such instances, the model may not be capturing the dynamics of the market, or there may be important omitted variables that may bias the estimates.[98] However, one must not ascribe too much importance to R^2, because a high R^2 (close to 1) is not required for a regression model to be reliable. Indeed, if one is

95. Jacques Dreze, Comment, *Nonspecialist Teaching of Econometrics: A Personal Comment and Personalistic Lament*, 2 ECONOMETRIC REVIEWS 291, 296 (1983), states this view quite well when he says that omitting a variable in this situation "amounts to elevating ignorance to arrogance."

96. *Id.*

97. KENNEDY, *supra* note 29, at 14. R-squared can be increased by simply adding variables to a regression, even if they are not statistically significant. To account for this problem, the *adjusted R-squared* is sometimes calculated. Adjusted R-squared increases when a new variable is added to the regression only if that new variable has a t-statistic greater than one. *Id.* at 94.

98. For a discussion, see Rubinfeld, *supra* note 20, at 179, 188, 216-17. However, even with a small R^2, least squares produces unbiased estimates, as long as the conditional expectation of the error term given the included explanatory variables is zero.

interested in estimating the partial effects of the explanatory variables on the dependent variable, R^2 may not be particularly relevant if the omission of other exogenous variables does not create any bias in the estimation of coefficient for the variable of interest.

One must also take care when using R^2 to compare two regression models. Such a comparison can only be given any weight if the two regression models have the same dependent variable. Otherwise, the comparison is one of apples to oranges. For example, consider one model where the dependent variable is average list price and a second model where the dependent variable is the average transactions price (including discounts). Even if the first model explains a larger percentage of the variation in price, the first model is not necessarily "better" than the second model if the regression is intended to estimate the impact of an act on the prices customers pay. Moreover, even when the dependent variable is the same in two regressions, the model selection criteria discussed in the previous section are likely a better basis for comparison than R^2.

R-squared is more relevant in a forecasting context, because a higher R^2 suggests that the model will provide forecasts with smaller confidence intervals. Nonetheless, when building a forecasting model, one must still be careful not to choose explanatory variables just to maximize the R^2.[99] Making selections based on a higher R^2 can lead to "over-fitting" of the model to the data set on which the model was estimated.[100] While the model might fit those data very well, it may actually provide worse forecasts when applied to other situations because the model has not really captured the important underlying economic relationships. Again, when building a forecasting model, the model selection criteria discussed in the previous section should be used instead of R^2.[101]

7. *Misspecification Bias and Specification Tests*

Regression analysis provides unbiased estimates of the regression model coefficients generally only if the regression model is correctly specified, so that the error term (the unexplained component of the dependent variable) is not correlated with the explanatory variables.[102] If this condition does not hold, the regression model may be misspecified and may not produce reliable coefficient estimates. Thus, the model will

99. *Id.* at 219-20.
100. *See* GREENE, *supra* note 16, at 306.
101. *Id.* *See also* part D.5.a. of this chapter.
102. *See* WOOLDRIDGE, *supra* note 3, at 49-54.

not reliably measure the effect of the allegedly anticompetitive act on the variable of interest. Thus, from an economic and statistical point of view, it is important to test whether a regression model is misspecified in order to avoid drawing incorrect or unreliable conclusions.[103]

Misspecification can arise from,[104] among other reasons, the imposition of incorrect restrictions on the regression coefficients, the assumption of an inappropriate functional form, or inclusion of endogenous variables as explanatory variables.[105] Endogenous variables are those that are caused, in whole or in part, by the dependent variable of the model.[106] For example, the quantity sold of a good is determined in part by its price. Thus, quantity is typically an endogenous variable in a regression model where price is the dependent variable.

In many circumstances, these problems can be remedied if detected and if the appropriate data are available.[107] Fortunately, there are well-known and generally accepted econometric tests for the presence of misspecification.[108] These tests often provide an objective and scientific means to determine whether a regression model is reliable.

For example, omitted variable bias can be tested by identifying and including in the regression model additional explanatory variables that economic reasoning and other market information suggest are likely to affect the dependent variable. If these additional explanatory variables turn out to be statistically significant, and the coefficient estimates on the previously included explanatory variables change substantially when the additional variables are added, then the regression model that omitted the additional explanatory variables likely is misspecified and its results are biased and unreliable.[109]

Another potential misspecification is caused by imposing incorrect restrictions on the coefficients of the regression model—that is, by forcing the coefficients to take incorrect values. Forcing incorrect values on some coefficients can bias other coefficients.[110] For example, suppose that the data contain information on two different customers (A and B),

103. *See* GREENE, *supra* note 16, at 701.
104. Another source of bias, the omission of important explanatory variables from the model, is discussed in part D.5.a. of this chapter.
105. *See* KENNEDY, *supra* note 29, ch. 6, 9, 10.
106. *See* WOOLDRIDGE, *supra* note 3, at 50-51.
107. *Id.* at 50, 51, 58.
108. *Id.* at 44; *see also* Jerry A. Hausman, *Specification Tests in Econometrics*, 46 ECONOMETRICA 1251 (1978).
109. *See* GREENE, *supra* note 16, at 337.
110. *Id.* at 281-82.

and that the regression model imposes the restriction that both customers have the same coefficients. Further assume that customer A has a perfect substitute to which it could turn in response to an anticompetitive price increase, while customer B has no substitute. In that case, even if there were a conspiracy, the alleged conspirators would be able to impose an anticompetitive price increase only on customer B. The existence of a substitute for customer A would prevent the alleged conspirators from imposing a supracompetitive price increase on customer A. If the coefficient on PERIOD were allowed to be different for the two customers, customer A would have a zero coefficient (since its price would not be higher during the alleged conspiracy period), while customer B would have a positive coefficient, say 10%, reflecting a 10% overcharge. However, a regression model that forced the PERIOD coefficient to be the same for the two customers likely would result in a positive coefficient estimate that was biased for both customers (too high for customer A and too low for customer B). For example, the estimate of the PERIOD coefficient might turn out to be 5%. Forcing the PERIOD coefficients to be the same for the two customers could lead to the incorrect inference that both customers were harmed by the alleged conspiracy, when in fact customer A was not harmed. This restriction may also significantly affect the accuracy of the estimate of total damages. For example, if customer B had purchased 100 units at $1.10 and customer A purchased 10 units at $1, the true total damages would be ($0.10*100 + $0) = $10. Yet, the results from the regression model in this example would lead to a total damages estimate of approximately ($0.05*100 + $0.05*10) = $5.50, a substantial underestimate. The existence of bias resulting from the imposition of incorrect restrictions can be econometrically tested using standard procedures often used by economists.[111]

8. *Structural Models Versus Reduced Form Models*

As discussed above, an expert must specify an econometric model that lays out the relationship between the dependent variable and the explanatory variables that represent the major economic factors potentially affecting the dependent variable. Econometric models can generally be divided into two broad types: *structural models* and *reduced form models*.[112]

---

111. *Id.* at 271-300.
112. *See* KENNEDY, *supra* note 29, at 171-72.

A structural model consists of equations that represent the basic economic relationships between variables.[113] For example, the structural model for price and quantity in an industry might consist of a demand equation—representing how customer demand for the product is determined—and a pricing equation—representing how price for the product is determined. Price and quantity would be the endogenous variables, since they are determined by the other variables in the model.

Estimating structural models has certain advantages. For example, the parameters in the equations of a structural model generally have a specific economic meaning. For example, the coefficient on price in the demand equation would be related to the elasticity of demand. Nonetheless, structural equations can be difficult to estimate. Thus, analysts often instead estimate a "reduced form."[114] The reduced form is an algebraic rearrangement of an econometric model that has each endogenous variable on the left side of one equation and only exogenous explanatory variables on the right side. One can start with an explicitly specified structural model and then solve it to obtain the reduced form model, or one can directly specify the reduced form model. In this latter case, the endogenous variables price and quantity are expressed in terms of the exogenous variables and the error terms, and the coefficients on the exogenous variables in the reduced form are functions of (depend on) the structural parameters.[115] One disadvantage of reduced forms is that their coefficients are generally a mixture of the structural parameters from the demand equation and the price equation and thus often do not have a direct economic interpretation.[116]

Should one use a structural model or a reduced form model? If the question of interest centers on the value of a structural parameter (e.g., the elasticity of demand), estimation of a structural model is generally a more straightforward way to proceed.[117] However, if the question of interest concerns predicting an outcome variable on the basis of the exogenous variables, estimation of a reduced form model is generally a more straightforward way to proceed.[118] For damages analysis in

113. *Id.*
114. *See* Baker & Rubinfeld, *supra* note 2, at 391-92.
115. *See* KENNEDY, *supra* note 29, at 171-72.
116. *See* Baker & Rubinfeld, *supra* note 2, at 392.
117. *Id.* at 405.
118. *Id.* One can predict the endogenous variables from a structural model by solving the model given the estimated structural parameters. However, this method is generally not as straightforward as estimating the reduced form model directly. *Id.* at 414-16.

antitrust cases, the question of interest often involves predicting the likely "but-for" value of an outcome variable, such as price.[119] Accordingly, reduced form models will often be better suited to antitrust damages analyses. Even if one is going to estimate a reduced form, however, consideration of the basic elements of the appropriate structural model can be very useful in identifying the appropriate explanatory variables to include in the reduced form model since the reduced form model is obtained by solving the structural model for the endogenous variables in terms of the exogenous variables.[120]

In some cases, the specified model may be a hybrid of a reduced form equation and a structural equation in that it includes endogenous variables as explanatory variables. If so, ordinary least squares (regular regression) can yield biased and inconsistent estimates.[121] Instead, "instrumental variables" (IV) techniques will usually be more appropriate because they can still be consistent even when one of the explanatory variables is endogenous.[122] While the absence of good instruments can cause problems for IV methods, the bias that results from ignoring the endogeneity problem and applying OLS could be substantial.[123]

9. *Special Topics with Panel Data*

Data sets can be composed of single observations over a period of time (a *time series*) or of a number of observations in single time period (a *cross section*). In general, data sets consisting of both time series and cross sectional observations are called cross-section time series data—or panel data.[124]

Panel data allow for richer econometric specifications than time series or cross-sectional data. Two often-used econometric specifications for panel data are *fixed effects* and *random effects*.[125] To understand the basics of these specifications, start with the usual regression model

---

119. *Id.* at 389.
120. *Id.* at 391.
121. *See* WOOLDRIDGE, *supra* note 3, at 83.
122. *Id.* ch. 5.
123. *See* Michael P. Murray, *Avoiding Invalid Instruments and Coping With Weak Instruments*, 20 J. ECON. PERSPECTIVES 111 (2006).
124. For a discussion of issues involved in using panel data, see part D.7. of this chapter.
125. *See* WOOLDRIDGE, *supra* note 3, at 284-91.

$$y_{it} = X_{it}\beta + \varepsilon_{it}$$

where i indexes the observations in a given period of time and t indexes different time periods. For example, if there were 100 transactions in each year (corresponding to a purchase by each of 100 customers) over five years, i would run from 1 to 100, t would run from 1 to 5, and there would be 500 total data points.

There may be reasons to believe that there are unobserved factors specific to each customer that affect the dependent variable such as price. For example, suppose some customers are better negotiators than others. Good negotiators are likely to receive better prices on each transaction than poor negotiators. Similarly, suppose some customers have better substitutes for the product in question than other customers. Then, the customers with good substitutes are likely to receive better prices on each transaction than customers with poor substitutes.

When these *unobserved individual-specific effects* are present, the error term of the regression can be modelled by breaking it into two parts.[126] The first part, representing the unobserved individual-specific effect, can be written as μ_i, indexed only by i. It is assumed to be fixed across all purchases for customer i. The second part of the error term can be written as η_{it}. It is indexed by i and t, indicating that it differs from one time period to the next for each customer. With this breakdown of the error term, the regression model can be rewritten as

$$y_{it} = X_{it}\beta + \mu_i + \eta_{it}$$

A regression of y_{it} on X_{it} can be run with a set of indicator (dummy) variables for each unit of observation (customer, in our example).[127] This is possible with panel data (but not a cross-section) because there are multiple observations for each customer over time. The dummy variable for customer i will control for the unobserved individual-specific effect for customer i. This is called a fixed effects

126. *Id.* at 251.
127. *Id.* at 265-79. Here, a dummy variable for customer i will equal one for that customer, and zero for all other customers. *See* MADDALA, *supra* note 14, § 8.2.

regression, where the fixed effect for customer i is equal to the regression coefficient on its unique dummy variable.[128]

The use of fixed effects thus allows the econometrician to control for unobserved factors that do not vary across the observations for an entity, such as an individual in our example. Controlling for these factors is important because the unobserved individual-specific factors may be correlated with the explanatory variables. In that case, a regression of y_{it} on X_{it} without including the fixed effects would yield biased and unreliable results. Without controlling for the fixed effects, the regression in effect would suffer from omitted variable bias. Thus, a fixed effects regression has the benefit of providing reliable results, even if the unobserved individual-specific effect component of the error term is correlated with the explanatory variables.[129]

For an example of how failing to control for fixed effects can bias a study's results, consider a study of the effects of education on individuals' earnings. An individual's motivation and inherent ability are difficult to measure fully with observable variables. In an earnings regression model, motivation and inherent ability therefore will appear in the error term as an individual-specific effect that will increase the individual's earnings over his or her lifetime. Yet, these same factors likely also increase the individual's education. Thus, education and the error term likely will be correlated. Without including fixed effects for the individuals in the regression, a regression of earnings on education (and other explanatory variables) may well produce biased results, since the coefficient on earnings would then also reflect the effects of motivation and inherent ability.[130]

If the unobserved individual-specific effects are not correlated with the explanatory variables, fixed effects still yields reliable results, but a better estimator may be available. In particular, the random effects estimator may provide more statistically precise results than fixed effects. But, random effects also will be biased and unreliable if it turns out that the unobserved individual-specific effects are correlated with the explanatory variables. Before relying on random effects results, therefore, it is important to test the validity of the assumption of no correlation between the unobserved individual-specific effect and the explanatory variables.[131]

---

128. *See* WOOLDRIDGE, *supra* note 3, at 265-79.
129. *Id.* at 268.
130. *Id.* at 61.
131. *See* Hausman, *supra* note 108.

The *Hausman Specification Test* can be used for this purpose.[132] Under this test, the fixed effects and random effects are compared. If the zero correlation assumption is valid, the two sets of results should be similar. If, on the other hand, the zero correlation assumption is invalid, then the random effects results (as well as least squares results) will be biased while the fixed effects results are still unbiased, so that the two sets of results should diverge.[133] In this case, the fixed effects model would yield more accurate results.

10. Special Topics with Time Series Data

In a time series context, model specification and estimation proceeds differently depending on whether the time series in question are *stationary* or *nonstationary*.[134] In effect, a nonstationary time series shows a significant trend upward or downward.[135] Such trends can create problems. For example, suppose a dependent variable is increasing at the same time an explanatory variable is increasing. A simple regression then could find a positive and significant relationship between the explanatory variable and the dependent variable, because both variables are independently trending upward at the same time, even though the two variables are not related. Moreover, the R^2 from an OLS regression has no meaning when data are not stationary. Many economic time series are thought to be nonstationary, such as a country's Gross Domestic Product or the price of a stock.[136]

The properties of many econometric estimators, such as least squares regression, are derived under the assumption that the data series defining the dependent and explanatory variables are stationary. When the series

---

132. *Id.*
133. *Id.*
134. *See* KENNEDY, *supra* note 29, ch. 19. These issues should be considered if one has either a single time series, or panel data with reasonably long time series for each unit of observation. *Id.* at 287.
135. Technically, a stationary series has a finite variance, which means that in the future the series will not stray too far from its current values. *See* GREENE, *supra* note 16, at 752. A nonstationary series, however, exhibits infinite variance, which means that in the future the series will get infinitely far from its current value with positive probability. *Id.* at 776-81.
136. *Id.*

are nonstationary, a different set of econometric techniques should be employed.[137]

a. Tests for Stationarity

A number of different methods for testing for stationarity in a time series have been proposed.[138] These methods generally test for what econometricians call "unit roots."[139] These unit root tests are not perfect, and one should view the results of the tests within the context of economic theory and the dynamics of the market. For example, suppose the time series of interest are price and marginal cost. In general, it would be economically implausible that one of these series is nonstationary while the other is stationary, because price usually is significantly influenced by marginal cost.[140] Now, suppose the results strongly reject the null hypothesis that marginal cost is nonstationary, but do not reject the null hypothesis that price is stationary. It would be reasonable to conclude from the results of the two tests, plus the implication of economic theory (suggesting both series must be either stationary or nonstationary), that both price and marginal cost are stationary in this case.[141] If so, then standard regression analysis can be used.

137. *Id.*

138. For a brief introduction to stationarity tests, see MADDALA, *supra* note 14, ch. 14; *see also* GREENE, *supra* note 16, at 781-85; D. Kwiatkowski et al., *Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root*, 54 J. ECONOMETRICS 159 (1992).

139. *See* GREENE, *supra* note 16, at 780.

140. *See id.* at 790-91. There are some situations in which price might be stationary while cost was nonstationary, or vice versa. For example, cost and a third explanatory variable that affects price could both be nonstationary, but cointegrated (discussed further below). In that situation, a more complex econometric procedure must be used.

141. Similar considerations apply when one has panel data with long time series for each unit of observation. In many circumstances, it would be implausible for the price of one customer to be nonstationary while the price of another customer is stationary. *See* J. Hlouskova & M. Wagner, *The Performance of Panel Unit Root and Stationarity Tests: Results From a Large Scale Simulation Study*, 25 ECONOMETRIC REVIEWS 85 (2006).

b. Modelling Market Dynamics with Stationary Time Series

Although the usual econometric techniques can be applied to stationary time series, particular issues may arise when attempting to model the dynamics of such series. The starting point for modelling the dynamics is studying the economics of the industry and examining why dynamic effects may arise in the first place. That is, why do the economic outcomes not adjust instantaneously to changes in economic conditions, which would make modelling only contemporaneous influences on the dependant variable inappropriate? There may be long-term contracts or other institutional features that slow the speed of adjustment, so understanding the economic context can be important in accurately modelling the dynamics.

Three types of dynamics may need to be modelled. First, dynamic effects may arise if the dependent variable is influenced by earlier values. For example, where the dependent variable is price, the price in the current period can be affected by prices in prior periods if there are costs to firms changing their prices. These costs would discourage frequent price changes, so price changes would not occur as often and would likely be delayed until there was an accumulation of reasons to change price. Modelling this type of dynamics is often done by including lags (i.e., prior observations) of the dependent variable as additional explanatory variables.[142]

Second, dynamic effects may arise if the impact of a change in an explanatory variable plays out over two or more time periods. This can occur, for example, if long-term contracts allow for adjustments to price due to changes in costs, but only after a lag. Modelling these dynamics is often done by including not only the contemporaneous value of an explanatory variable, but also values of the variable from earlier time periods as additional explanatory variables.[143]

Finally, dynamic effects may arise if the error term includes unobserved factors that are themselves correlated over time. This induces *serial correlation* in the error term. The dynamics in this case may be limited to the error term, and can be addressed by modelling the form of the serial correlation.[144]

The economics of the industry can suggest that dynamics of one type or another are present, and the modelling should focus on that form of

142. *See* GREENE, *supra* note 16, ch. 17.
143. *Id.*
144. *Id.* ch. 13.

dynamics. As discussed above, if there is evidence that sales are made under contracts that adjust to cost changes with a lag, then lagged costs likely should be included as an explanatory variable for estimating the impact of cost on prices. In the case where there are reasons to believe dynamics of all three types may be present, it is important that the modelling of all three types be done jointly, rather than sequentially. There may be important interactions between the three that would be missed if the modelling was done sequentially.[145]

There is also an important question about the number of lags that should be included in the model. For example, if the analysis involves monthly prices and costs, should costs from just the previous month be included as an additional explanatory variable, or should the regression include costs from several previous months? The choice of lag length can be based on econometric tests, as we describe below. For example, a starting point for the modelling exercise might be the following model:

$$P_t = \sum_{j=1}^{P} \alpha_j P_{t-j} + \sum_{j=0}^{Q} X_{t-j} \beta_{t-j} + \varepsilon_t$$

$$\varepsilon_t = \sum_{j=1}^{R} \rho_j \varepsilon_{t-j} + u_t$$

where P, Q, and R represent initial choices of lag length for the lagged dependent variable, the lagged exogenous variables, and the serial correlation in the error term.[146] P, Q, and R should be chosen to be outer limits on the lag lengths. The optimal lag lengths then can be chosen by estimating the model using different combinations of the three lag lengths, and using model selection criteria such as the Schwartz criteria to choose between these models.[147]

It is important to recognize that the specification above, which includes lags of the dependent variable as additional explanatory variables, potentially complicates estimation of the model. In particular, it creates a correlation between the explanatory variables and the error term if the error term itself exhibits serial correlation.[148] Under this

145. *Id.* ch. 17.
146. The error term could also be specified to have *moving average* components. For simplicity of presentation, we focus on autoregressive components only.
147. *See* GREENE, *supra* note 16, at 717.
148. *Id.* at 534-35.

condition, least squares estimation of the model can lead to biased and inconsistent estimates. If it appears that least squares estimation in this type of specification could lead to biased and inconsistent estimates, other econometric methods should be used (as discussed below).

It is also important to recognize that, as long as the rest of the model is properly specified, the appropriate estimation method will generally be unbiased and consistent even if the serial correlation in the error term is left unmodeled.[149] As discussed above, however, it will still be necessary to account for the serial correlation when estimating the standard errors to ensure correct statistical inference, and this can be done using the techniques described above.

c. Modelling Dynamics in Nonstationary Time Series

If the unit root stationary tests indicate that the series are nonstationary, a different set of econometric techniques in general should be used to estimate the relationships between the series.[150] The simplest approach is to take the observed value of a variable in each time period observation "t" and subtract from it the value of the same variable in the previous time period "t-1." Doing this calculation for each observation creates a new time series data set of the changes in the variable from one period to the next, and this transformation is typically done for all of the variables involved in the regression.[151] This *first differences* approach frequently removes the nonstationarity from the data, although the transformed data should be tested to ensure that is the case.[152] The approach, however, will also often remove some information about the relationship between the dependent variable of interest and the exogenous variables. Thus, goodness of fit measures such as R^2 and the measured significance of the statistical tests usually will be lower than regressions based on the untransformed data.[153]

149. *Id.* at 533.
150. *See* MADDALA, *supra* note 14, § 6.10 & ch. 14.
151. *See* GREENE, *supra* note 16, at 790. Depending on the economics of the market and the nature of the alleged anticompetitive behavior, a dummy variable may not be transformed this way.
152. *Id.*
153. *See* GEORGE C.S. WANG & CHAMAN L. JAIN, REGRESSION ANALYSIS: MODELING & FORECASTING 88 (2003).

More sophisticated econometric techniques can be applied if the time series in question are nonstationary, but are *cointegrated*.[154] Two (or more) nonstationary series are said to be cointegrated if a linear combination of the different series is itself stationary.[155] For example, consider two time series. If one series plus a constant multiplied by the second series is stationary, the two series are cointegrated. Econometric methods exist for testing whether two or more series are cointegrated.[156]

Cointegration can arise if economic forces prevent the nonstationary series from getting too far from each other, if they have a long-run relationship. For example, again consider a price series and a marginal cost series. While price and cost may each be nonstationary and trend upward, the economic forces at play would tend to keep the gap between price and marginal cost (i.e., the margin) from becoming too large or too small. If the margin grew too large, entry would occur or customers would cut back their purchases and drive the margin back down. If the margin grew too small, exit would occur or customers would increase their purchases and the margin would increase again. Thus, the margin can be stationary even though price and cost are themselves nonstationary.

If the tests for cointegration indicate that the series are cointegrated, then the relationship between the series can be specified and estimated as a *vector error correction model* (VECM).[157] A VECM models both the cointegrating equations (the long-run relationships between the nonstationary variables) and also the short-run dynamics of the residuals of the cointegrating equations.[158]

---

154. If two series are nonstationary, but not cointegrated, regression techniques can still be applied. However, the estimated coefficients will have nonstandard sampling distributions. *See* Peter C.B. Phillips & Steven N. Durlauf, *Multiple Time Series Regression With Integrated Processes*, 53 REV. ECON. STUD. 473 (1986).
155. *See* GREENE, *supra* note 16, at 790.
156. *Id.* at 794-96. One of the most widely used of these tests is presented in Soren Johansen, *Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregression Models*, 59 ECONOMETRICA 1551 (1991).
157. *See* GREENE, *supra* note 16, at 793-94.
158. *Id.*

d. Considerations in Building a Forecasting Model

Time series models may be used to make forecasts. For example, as discussed below, a time series model may be estimated using data from a competitive period, and then used to forecast but-for prices during a conspiracy period.[159] Such a forecasting model obviously will be useful only to the extent that it provides reliable forecasts. For example, the start and end of a conspiracy period may be characterized by sharp "turning points" in a price series. Accordingly, it is important that the forecasting model be able to predict turning points that occur for nonconspiracy reasons. Otherwise, turning points due to nonconspiracy reasons might be incorrectly attributed to the conspiracy.[160]

How does one tell if the forecasting model is reliable? Looking at R^2 is generally not a good way to assess a forecasting model. Models with lagged dependent variables or serially correlated errors often appear to track the historical data well and achieve a very high R^2. However, as discussed previously, such models may suffer from a form of "over-fitting."[161] That is, while they fit the historical data well, they fail to capture the true economic structure of the dynamic pricing process. The good fit may come from the lagged dependent variables, not from the other explanatory variables that cause changes in the direction of the dependent variable. Thus, when such models are used to forecast prices during an alleged conspiracy period, they do a poor job forecasting turning points in the dependent variable.

One way to test a forecasting model is to see how well it predicts for some time period before the alleged anticompetitive act, assuming adequate data are available.[162] To implement such a test, one would divide the data from before the alleged anticompetitive act into two parts: one part is used to estimate the forecasting model and the other part is used to assess the accuracy of the forecasts of the model. It is particularly useful to exclude from the part of the data used to estimate the model observations that contain a turning point (e.g., a substantial change in prices). If the forecasting model is reliable, then it should produce a reasonably accurate forecast for the part of the data not used in estimation. In particular, it should be able to predict any turning points. If the model does a poor job forecasting in this context, then it is unlikely

159. *See* part E.1. of this chapter.
160. *See* Baker & Rubinfeld, *supra* note 2, at 396.
161. *See* part D.6. of this chapter.
162. *See* KENNEDY, *supra* note 29, at 102.

to produce a reliable estimate of but-for prices during the period of the anticompetitive act.

E. Before-During Approach

The before-during approach identifies the effect of the alleged conduct by using data from a period before the alleged conduct in combination with data from the period when the alleged conduct occurred.[163] Comparing the values of the dependent variable in the before period to the values it took on in the during period may serve to identify the effect of the alleged conduct.[164] In making this comparison, however, it is critical to account for any significant differences in other economic factors between the before and during periods.[165]

1. "Dummy Variable" Models Versus "Prediction" Models

For simplicity, assume that the best explanatory variable available for capturing the potentially anticompetitive effects is a dummy variable representing the time period when the alleged anticompetitive acts took place.[166] This variable will be 1 during that period and 0 at other times.

---

163. *See* van Dijk & Verboven, *supra* note 1, at 2335-36.
164. *Id.* The before period may be considered the control period and the during period the treatment period. See the discussion in part A of this chapter.
165. *See* van Dijk & Verboven, *supra* note 1, at 2335-36.
166. Depending on the nature of the antitrust allegations, the theory of causation, and specific facts of the case, there can be variables other than a dummy that may better measure the anticompetitive behavior in estimating damages. *Id.* Assume, for example, prices can only increase when existing customer contracts expire over time. The impact of the alleged conspiracy would not increase prices to all customers immediately, so average price increases related to the alleged conspiracy would be expected to trend upward over time. There could be several ways to model this effect other than using a simple dummy variable for the alleged conspiracy period, depending on the available data. *Id.* Under certain circumstances, one simple method to create a variable to approximate the conspiracy might be a "time trend" beginning at the period of the alleged conspiracy, with the value "1" in the first observation (e.g., a month) of the conspiracy period, "2" in the second, "3" in the third, etc. If average prices were the appropriate dependent variable, then the coefficient on this trend variable would be interpreted as the average monthly price increase due to the conspiracy. The

Specifically focusing on a reduced form model, one can account for the differences, if any, between the way price was determined in the during period (for example, with alleged price fixing) and the way price was determined in the before period (without price fixing). Suppose that the reduced form equation for price in the before period is

$$P_t^B = Z_t^B \Pi^B + \nu_t^B$$

where Z_t^B is a full set of exogenous explanatory variables (identified by any relevant economic analysis), Π^B are the estimated coefficients capturing the partial impact of these variables on the price P_t^B, and ν_t^B is the error term. Suppose that the price-fixing conspirators set price by adding a fixed amount "Δ" to the price that would have prevailed otherwise. The "overcharge" resulting from the price-fixing conspiracy would then simply be Δ. In that case, the reduced form equation for price in the after period would be

$$P_t^D = Z_t^D \Pi^B + \Delta + \nu_t^D$$

The coefficients on the exogenous variables in the during period, Π^B, are the same as their coefficients on the exogenous variables in the before period. Holding constant the values of the exogenous variables and the error term, prices in the during period are simply increased by the amount Δ.

The equations for the before period and the during period can be combined to obtain a single equation that applies to both periods. This equation is

$$P_t = Z_t \Pi^B + \Delta A_t + \nu_t$$

---

overcharge in any month during the conspiracy would be the corresponding time trend value multiplied by that coefficient. However, care must be used with this and any approach for creating a variable that measures the impact of an alleged anticompetitive act. For example, at some point all contracts will have been renegotiated, so it would be inappropriate to continue to predict price increases beyond a certain number of observations.

where A_t is a dummy or indicator variable that takes the value of 1 if the time period t is in the during period and the value of 0 if the time period t is in the before period. Because of the crucial role played by the dummy variable, this type of model is often called a *dummy variable model*.[167]

This model assumes that the coefficients on the predetermined variables are the same in the before and during periods. This assumption may not always be valid. For example, price-fixing conspirators may react differently to changes in supply and demand factors than they would have had they not been conspiring.

Thus, it may be important to test whether prices relate differently to the supply and demand factors in the during period than they did in the before period. In more technical terms, it can be important to test whether there has been a significant shift in the relationship between the exogenous variables and the dependent variable of interest.[168] Statistical tests of whether the additional restrictions inherent in the dummy variable model are appropriate can be performed by including *interaction* terms in the equation.[169] Interaction terms are variables calculated by multiplying other exogenous variables by the dummy variable.[170] Specifically, the reduced form price equation with interaction terms can be written

$$P_t = Z_t\Pi^B + \Delta A_t + A_t Z_t(\Pi^D - \Pi^B) + \nu_t$$

where A_t again is the dummy variable and $A_t Z_t$ is an interaction term. This equation will be equivalent to the reduced form equation of the dummy variable model if $\Pi^D = \Pi^B$, that is if the coefficients on the interaction terms will be zero. Thus, the appropriateness of the dummy variable model can be econometrically tested by testing the hypothesis $\Pi^D = \Pi^B$. If this hypothesis is rejected, this test suggests that the underlying assumption of the dummy variable model that the coefficients of the exogeneous variables in the before and during periods are the same is inconsistent with the data.

---

167. *See* Daniel L. Rubinfeld, *Econometrics in the Courtroom*, 85 COLUM. L. REV. 1048; James F. Nieberding, *Estimating Overcharges in Antitrust Cases Using a Reduced-Form Approach: Methods and Issues*, J. APPLIED ECON., Nov. 2006, at 361, 367-69.
168. *See* GREENE, *supra* note 16, at 287-93.
169. *Id.*
170. *Id.* at 326.

If the dummy variable model is found to be inappropriate, an alternative is a *prediction model*.[171] The first step in implementing a prediction model is to estimate the reduced form equation on the data from the before period. Then, this equation is used to predict the prices that would have prevailed in the during period, given the values of the exogenous variables in the during period. The "predicted" price for time t of the during period, $\hat{P}_t^D$, can therefore be written as

$$\hat{P}_t^D = Z_t^D \Pi^B$$

where Z_t^D are the values of the exogenous variables in the during period and Π^B are the coefficients for these variables estimated based only on data from the before period. As can be seen from this equation, the prediction model assumes that the coefficients on the exogenous variables would have been the same in the during period in the but-for world as they were in the before period in the actual world. But, under certain conditions, this assumption may be invalid. For example, assume there was a change in the relationship between the exogenous variables and the variable of interest (e.g., price) in the during period that had nothing to do with the alleged anticompetitive conduct (e.g., entry or exit of firms from the market) and the prediction model does not take that change into account. Under these conditions, estimates using data solely from the before period may not accurately predict pricing but for an alleged conspiracy or other anticompetitive act.[172] Moreover, as discussed above, a prediction model should be tested where possible to check whether it accurately predicts turning points in the data.[173] However, unlike the dummy variable model, the prediction model does not assume that the coefficients for the exogenous variables are the same across the before and during periods in the actual world.

171. *See* MADDALA, *supra* note 14, § 4.7; Nieberding, *supra* note 167, at 367-69.

172. In this way, the prediction model may suffer from some of the same limitations related to the dummy variable model discussed above.

173. A dummy variable model can suffer from a similar problem of failing to fit turning points well, and thus incorrectly attributing to the conduct turning points that may have nothing to do with the alleged conduct. *See* WOOLDRIDGE, *supra* note 3, at 61-62.

In addition, one also could fit the model to the during period and use the model to predict back ("backcast") into the before period. To the extent that the backcast is above the actual prices in the before period, that would suggest the price setting process in the during period produced higher prices than the price setting process in the before period. This finding would be consistent with the proposition that the alleged anticompetitive behavior in the during period increased prices, and the difference between the actual and predicted prices in the before period could form the basis for calculating an overcharge and damages.

2. *Before-During-After Models*

The previous section discussed before-during models that assume there is a single unaffected period (the before period). Sometimes, data might exist on several unaffected periods. For example, data may exist not only for the before period, but also for an after period. With more than one unaffected period, the question arises as to whether one should combine the two unaffected periods and, if not, then which of the two unaffected periods should be compared to the affected period.[174]

The before and after periods should be combined, if appropriate, since that will result in a larger sample and more precise estimation.[175] It is appropriate to combine them if the parameters of the reduced form equation are the same across the two periods.[176] Suppose the following are the reduced form equations for price, one for the before period and one for the after period:

$$P_t^B = X_t^B \Pi^B + u_t^B$$
$$P_t^A = X_t^A \Pi^A + u_t^A$$

If the relationship between price and the exogenous variables in the two periods is the same, then the coefficients are the same for each period ($\Pi^B = \Pi^A$). In that case, it would be appropriate to combine the two periods together and estimate a single model that applied to both.

174. *See* van Dijk & Verboven, *supra* note 1, at 2336. Of course, it is important to ensure that the supposedly unaffected periods really are unaffected.

175. *See* MADDALA, *supra* note 14, at 69, 270; A.H. STUDENMUND, USING ECONOMETRICS: A PRACTICAL GUIDE 68 (4th ed. 2001).

176. *See* TAKESHI AMEMIYA, ADVANCED ECONOMETRICS 399-402 (1985).

The hypothesis $\Pi^B = \Pi^A$ can be econometrically tested by estimating separate models for each period and then comparing the respective coefficients to see if they are the same, apart from statistical variation.[177]

The difficult decision arises when this test rejects combining the two periods into a single model and one must choose which unaffected period is the better basis of comparison for the affected period, assuming both are truly unaffected. Before addressing how to decide which unaffected period to use, first consider why two unaffected periods might exhibit different relationships between the dependent variable and the explanatory variables. There are economic models of competition that suggest that periods of more intense competition ("price wars") will alternate with periods of less intense competition.[178] These models suggest that the relationship between price and the explanatory variables may change over time with the intensity of competition. In this case, a before period that was dominated by more intense competition might look different than an after period that was dominated by less intense competition, even though both periods were unaffected by the anticompetitive conduct at issue.[179] Alternatively, there may have been substantial differences in market conditions between the two periods (e.g., there may have been substantial new entry into the market), which led to the differences in the model coefficients.

In the event that there are substantial differences between the before and after periods, one needs to determine how to select which period to use. One way to answer this question is to try to model explicitly the switches between competitive regimes, and then use this model to predict which competitive regimes would have prevailed during the affected period in the but-for world. There are econometric techniques called *switching models* that can be used to do this.[180] Although not commonly done, using a switching model of this sort may also provide a more realistic picture of competition in the during period. In contrast to the implicit assumption of the dummy variable model, a price-fixing conspiracy may be subject to breakdowns, where price would return to competitive levels (or even subcompetitive levels) for a period before the conspiracy price level was reestablished. Failure to model this possibility may lead to less precision in the estimate of the damages, or

177. *See* GREENE, *supra* note 16, at 287-93.
178. *See, e.g.*, Jerry Green & Robert Porter, *Noncooperative Collusion Under Imperfect Price Information*, 52 ECONOMETRICA 87 (1984).
179. *Id.*
180. *See* AMEMIYA, *supra* note 176, at 399-402.

possibly even to an unreliable estimate of damages if explanatory variables are correlated with the onset of price wars because of omitted variable bias.[181]

An alternative approach for choosing between the before and after periods can involve using general market information and nonquantitative evidence from the case. For example, suppose that the after and during periods were characterized by strong demand for the product, while the before period was characterized by weak demand. These factors would suggest that the after period would provide a better basis of comparison for the affected period than would the before period. It is also possible that the effects of the anticompetitive conduct at issue last beyond the time when the conduct actually ended. If so, then the after period is not truly "unaffected" and the before period may be the superior basis for comparison.

Another approach, more conservative from the point of view of plaintiffs, is to predict price (absent the allegedly anticompetitive acts) for the affected period alternatively using the before and after periods, and then take the maximum of these two predictions. Both of these predictions are consistent with competitive behavior. Using the higher of the predicted prices but for the alleged anticompetitive acts compared to the actual prices will yield a lower estimate of damages, since damages in this instance are based on the difference between the actual and but-for prices.

3. *Identifying Beginning and End Points of the Damages Period*

The beginning and the end of the damages period alleged in many cases may not accurately reflect the actual beginning or end of the alleged unlawful conduct. For example, in price-fixing class action cases, the plaintiff attorneys often choose the beginning and end dates for the "class period" before discovery is undertaken. Moreover, the beginning or end of the effects of the alleged unlawful conduct may not coincide with the beginning or end of the conduct itself. The effects might start later, end earlier, or last longer than the conduct. Accordingly, experts should rely on the evidence developed in discovery, market facts, and the analysis of liability experts when determining the relevant starting and ending dates for calculating damages.

Accordingly, damages experts should be flexible when determining the appropriate beginning and end dates for damages, and econometric

---

181. *See* part D.5.a. of this chapter.

analysis may be able to assist in this respect. Recently, econometric methods have been developed that can be used to choose beginning and end dates based on market data.[182] These methods involve estimating the point in time when an econometric model exhibits a *structural change*, or shift in the model parameters. In the case of alleged anticompetitive conduct that had an effect, a structural shift in the econometric model should have occurred when the effect of the conduct began and a second structural shift should have occurred when the effect of the conduct ended. Thus, for example, identifying when there were structural shifts in the regression model can provide useful evidence concerning the beginning and end of overcharges from a price-fixing conspiracy.[183]

F. Benchmark and Difference-in-Differences Approaches to Damages

The benchmark approach to damages is based on the same basic principles and econometric tools as the before-during-after approach. However, in a benchmark analysis, the variable of interest (e.g., price) in the affected market is compared to that variable in an unaffected benchmark market.[184] Instead of using past prices in the affected market as a control, the benchmark approach uses prices in the unaffected market as a control. The benchmark market (or markets) can be a different area or a different end use market. The key is that the benchmarks (the control group) were unaffected by the alleged anticompetitive conduct. This approach may be particularly useful if there are inadequate data for the before period, so a before-during approach cannot reliably be performed.

One potentially significant problem with the benchmark approach is that differences in economic conditions between benchmark and affected markets make it likely that prices in those markets would not be identical, even if there were no anticompetitive behavior.[185] Comparing variable profit margins between the market of interest and the benchmark

182. Jushan Bai & Pierre Perron, *Estimating and Testing Linear Models With Multiple Structural Changes*, 66 ECONOMETRICA 47 (1998). See Dennis Carlton & Gregory K. Leonard, *Correcting the Bias When Damage Periods Are Chosen to Coincide With Price Declines*, 2004 COLUM. BUS. L. REV. 304 (2004), for a discussion of the application of these methods to price-fixing cases.
183. *See* Carlton & Leonard, *supra* note 182.
184. *See* van Dijk & Verboven, *supra* note 1, at 2336.
185. *Id.*

can control for many cost- and demand-based differences in prices, although there are recognized problems in calculating and comparing economically meaningful profit margins.[186]

The difference-in-differences approach also can be used to account for at least some of the persistent unobserved differences between the benchmark market and the market of interest. This approach requires data for the market of interest and the benchmark that cover both a period of time that is unaffected by the alleged anticompetitive act and the period of time when the market of interest allegedly was affected. Thus, the difference-in-differences approach combines the before-during and benchmark approaches.[187]

The difference-in-differences approach can be implemented using a regression that takes into account the various factors that differ between markets and time periods. This approach in general would use the equation

$$P_{it} = \mu_i + \delta_t + \theta D_{it} + \varepsilon_{it},$$

where i indexes markets (affected or unaffected), t indexes time, μ_i is a market-specific unobservable effect that is constant over time, δ_t is a time-specific effect that is constant over markets, D_{it} is a dummy variable equal to one if the market is the affected market and the time period t is during the affected period, and θ is the partial effect of the alleged anticompetitive conduct.

The difference-in-differences method can be expanded by using the following regression equation to control for other economic factors that might have changed over time.

$$P_{it} = X_{it}\Pi + \mu_i + \delta_t + \theta D_{it} + \varepsilon_{it},$$

186. *See, e.g.*, Franklin Fisher, *On the Misuse of the Profits-Sales Ratio to Infer Monopoly Power*, 18 RAND J. ECON. 384-96 (1987). However, if the benchmark and the market of interest can be shown (1) to have similar supply and demand characteristics and (2) to have similar accounting methodologies, then one may be able to estimate the difference in those margins and from that calculate whether prices are higher in the market of interest—even if data are only available in the during period.

187. *See* WOOLDRIDGE, *supra* note 3, at 130.

where X_{it} are exogenous variables specific to market i in period t.[188]

G. Summary of Key Points in Creating and Evaluating Econometric Models of Damages

As can be surmised from the previous sections, a number of considerations should be taken into account when building or evaluating an econometric model used to estimate damages.[189] Model specification and estimation involve many choices: explanatory variables, functional form, dynamic specification in a time series, estimation technique, and which of the basic approaches to damages to use. In many cases, statistical tests can help guide these choices. However, the starting point for any econometric model is economic theory and knowledge of the industry, and being sure that there are sufficiently reliable data to perform an econometric analysis.

Economic theory provides guidance as to the general types of explanatory variables that should be included in the model. For example, for an econometric model of price, economic theory suggests that variables representing cost, demand, and competitive factors would be important types of explanatory variables.[190] Knowledge of the industry helps determine with more specificity which variables to include. For example, if the industry in question were plastics, the price of petroleum (an important input to plastics) likely would be a relevant cost variable and production levels in downstream industries that use plastics likely would be relevant demand variables.

One must be able to obtain enough sufficiently reliable data to implement any economically sound econometric model. Econometric analysis based on very few data points in general will be much less reliable than econometric analysis based on a larger data set. There may also be issues with how accurate the data are, and whether they represent an unbiased sample of what has occurred in the market. Accordingly, the data need to be checked for accuracy and any potential selection bias. Finally, in many instances the data will need to be "cleaned" to some extent, removing obviously incorrect observations. For example, many damages analyses use data from a company's sales database. Care must

---

188. *Id.*
189. Much of the discussion here and elsewhere in this chapter can be applied to any econometric analysis. *See, e.g.*, WOOLDRIDGE, *supra* note 3; GREENE, *supra* note 16.
190. *See* van Dijk & Verboven, *supra* note 1, at 2335.

be taken with such data to ensure that all significant discounts have been accounted for, and to weed out any obvious errors.[191] In some circumstances, limitations on the availability or quality of data may make it impossible to conduct a meaningful econometric analysis.

Often economic theory and industry knowledge might suggest a large number of potential explanatory variables. While there is a benefit to parsimony (namely potentially increased precision of the estimators), there is typically a greater concern with omitting a potentially important explanatory variable that can induce omitted variable bias. Thus, generally speaking, one should exercise great care if potentially important variables are to be excluded from a model.[192] In general, statistical tests or the AIC and SC criteria should be employed before dropping variables. It is rarely advisable to drop a theoretically important variable, especially if its coefficient is statistically significantly different from zero. Similarly, one generally should not drop a group of theoretically important variables when their coefficients are jointly statistically significantly different from zero (even if they are individually not statistically significantly different from zero).[193]

The relationship between the dependent variables and the explanatory variables can be expressed in various functional forms of the underlying variables, and this is another important aspect of model specification.[194] For example, the dependent variable might be the unadjusted price of the product at issue and an exogenous variable could be variable costs. This model would provide an estimate of how much price would go up for a dollar increase in cost. However, there may be good theoretical or industry specific reasons to believe that the relationship of price to cost is better measured in percentage change terms for each variable. If so, then the functional form of the estimating equation may be better specified using the logarithms of price and cost as variables. Statistical tests can help determine which functional form is most appropriate for the situation at hand.[195]

In a time series context, specification of the dynamics is important. Again, the particular specification might be guided by industry

191. For example, frequently these transactions data sets include credits for returned goods, etc. It is important to use net revenue and net sales to calculate a measure of the price actually paid.
192. *See* part D.5.a. of this chapter.
193. *See* KENNEDY, *supra* note 29, at 94.
194. *See* part D.1. of this chapter.
195. *Id.*

knowledge, and statistical tests, and the AIC and SC can help make choices about (for example) the lag lengths employed.[196]

Care must be taken to specify correctly the structure of the error term in the estimation of the econometric model and to use estimates of the standard errors that are robust to deviations from the assumption of independent, identically distributed error terms. Only with correct standard errors will the resulting statistical inference be valid. The correct method for calculating standard errors can be determined based on the outcome of statistical tests, and in general such tests should be performed.[197]

The decision of which of the basic damages models to use can be of great importance. A reduced form model typically will provide the most efficient and simplest model for damage estimation. However, the explanatory variables to be included in such a model need to reflect the important underlying supply and demand aspects of the market, and in some instances estimating the underlying structural model will be a superior approach.[198]

The decision regarding which damages approach to employ also will depend on the available data and the nature of the market. Are there sufficient data to implement a before-during or before-during-after model? Is there a good benchmark? Can a difference-in-differences approach be used? In general, if reliable data are available from outside the period of that alleged anticompetitive act, then techniques that take advantage of those data will likely be the most reliable.[199]

Deciding whether to use a dummy variable or prediction model can also be important. Is there evidence that the dummy variable technique is unreliable due to substantial changes in the model coefficients due to the alleged illegal act? Will the prediction model yield reliable results? Again, there are statistical tests that can help determine if one approach is clearly superior to another.[200] Since neither approach likely is perfect, in the absence of strong statistical test results favoring one or the other, it may make sense to use both approaches and see if they yield similar results.

Once the econometric model is specified and estimated, another series of statistical tests and checks can be run on the reliability and

---

196. *See* part D.10. of this chapter.
197. *See* part D.4. of this chapter.
198. *See* part D.8. of this chapter.
199. *See* parts E and F of this chapter.
200. *See* part E.1. of this chapter.

robustness of the estimates.[201] The starting place is typically examining the sign and statistical significance of the coefficients. If, for example, the coefficient on cost in a price regression is negative and statistically significantly different from zero, this would cast doubt on the reliability of the model, since we would normally expect cost to have a positive effect on price. When performing this sort of check, however, one must take care that there is a clear interpretation to the coefficients; there may not be one in a reduced form model, for instance.[202]

When there is reason to question a model's specification, a variety of specification tests can and should be employed. For example, as discussed above, a model may impose potentially questionable restrictions on the coefficients, such as that the coefficients are the same during two time periods. Specifically, an econometric model of pricing at the customer level might assume that all customers have the same coefficients on the cost and demand variables. However, there may be reason to think that customers have different responses to cost and demand conditions, so there could be different coefficients on the explanatory variables across customers, including the variable measuring damages. In this case, it may be necessary to estimate the model for different groups of customers, or otherwise to employ a more complex model that takes these differences into account.[203] Similarly, a time series model might assume that the coefficients are the same over the course of the entire time period, but there may be evidence that the structure of the model changed at some point in time. A Chow test (or a more sophisticated test for structural breaks) can be used to test whether such a change occurred.[204] If so, the appropriate model needs to adequately control for such a change.

There may be reasons to question the crucial assumption that the error term of the regression is uncorrelated with the explanatory variables. This assumption often can be tested using a form of the Hausman Specification Test. For example, if we are concerned that an explanatory variable might be endogenous (which would bias OLS regression results), then the Hausman Specification Test comparing the

---

201. *See* part D.7. of this chapter.
202. *See* Baker & Rubinfeld, *supra* note 2, at 392.
203. *See, e.g.*, Johnson & Leonard, *supra* note 52, at 351-52.
204. The Chow test is a statistical and econometric test of whether the coefficients in two linear regressions on different data sets are equal. *See* Gregory C. Chow, *Tests of Equality Between Sets of Coefficients in Two Linear Regressions*, 28 ECONOMETRICA 591 (1960).

OLS results to the IV results can be used to check for this problem.[205] If such a problem is found to exist, then more complex econometric techniques can be used to correct for the bias.[206]

Finally, virtually any regression model eventually will fail one or more tests if enough tests and specifications are run, even if nothing is wrong with the model. However, failure of a test should be taken seriously, and a model should be rejected when it fails a test of a critical assumption, or if it fails a large number of the specification tests to which it is subjected.[207]

H. Case Study—The Application of the Before-During and Benchmark Approaches

Econometric techniques can be applied in a variety of ways to estimate the impact of the alleged anticompetitive conduct on certain economic outcomes, such as estimating lost sales or price changes due to the alleged conduct. Some of the econometric techniques and tests discussed above and their applications are illustrated using the before-during and benchmark approaches to estimate the price effects of alleged collusive behavior by analyzing actual data relating to Southern Powder River Basin (SPRB) coal.

1. *Background*

Virtually all coal produced in the United States is used for electric power generation.[208] About one-third of the coal now comes from the SPRB, which contains vast amounts of low-sulfur coal available in low-cost surface mines. SPRB coal is relatively homogeneous, with two main types: 8400 Btu/lb. coal and 8800 Btu/lb. coal. SPRB coal mining began with the 8400 Btu coal, which was closer to the surface and easier to mine. During the 1970s, many coal-powered electric generation units were constructed to burn this coal. Virtually all of these units also can burn the higher-quality 8800 Btu coal. As sulfur emissions regulations became more stringent, older electric generation units were converted from burning 12,000 Btu bituminous coals to burning 8800 Btu SPRB

205. *See* Hausman, *supra* note 108.
206. *See* WOOLDRIDGE, *supra* note 3, ch. 5.
207. *See* KENNEDY, *supra* note 29, at 76-79.
208. U.S. ENERGY INFORMATION ADMINISTRATION, ANNUAL COAL REPORT 2007 9 (2009), *available at* http://www.eia.doe.gov/cneaf/coal/page/acr/acr.pdf.

coals. Because these plants already suffer capacity losses from burning lower Btu SPRB coals, they generally will not burn 8400 Btu coal.

During 1999 and 2000, the three leading SPRB coal producers conducted what economists may describe as "cheap talk." Cheap talk refers to communications between oligopolists about future intentions. Such talk can facilitate understanding of each other's goals and intentions, which can lead to collusive production and pricing.[209] The statements were of the nature that they would reduce production "until prices improved." The largest producer, Peabody Energy, closed a mine in 1999. The third largest producer—Arch Coal, Inc.—closed another mine in 2000, and several other planned expansions were cancelled. The result was a spot price increase in 2001. In the aftermath of the capacity reductions, prices settled about 30 percent above the level before the shut-downs. The DOJ opened an investigation in July 2001 but ultimately took no action against the SPRB producers.

2. *Economic Model, Data, and Choice of Explanatory Variables*

For the purpose of this case study, the simplest economic model hypothesizes that if the alleged collusion had an effect, then prices of SPRB coal during the damages period should be higher than before the damages period after controlling for other supply and demand influences. As will generally be the case, the most straightforward econometric models for estimating damages use a reduced form, rather than a structural form,[210] but the model still needs to include the major supply and demand influences on price that would be included in a structural model. Here, we assume a linear relationship between the relevant market forces and SPRB prices. A full analysis would investigate (and potentially test) whether the relationship of these factors to price should be specified differently, such as in percentage change terms.[211]

The price of SPRB coal used in the analysis is the spot price reported by the utilities that purchased the SPRB coal.[212] The utilities reported the

209. *See* MICHAEL D. WHINSTON, LECTURES ON ANTITRUST ECONOMICS 20 (2006). Whether such behavior constitutes an unlawful price-fixing "agreement" is a legal question.

210. See part D.8. of this chapter for a discussion of structural form models vs. reduced form models.

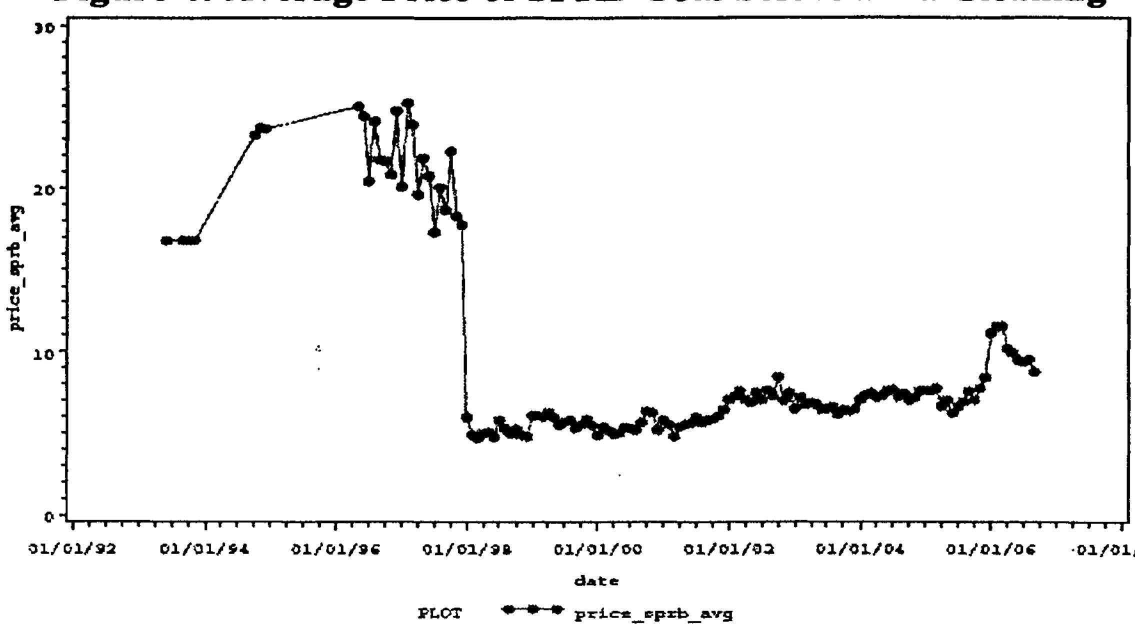
211. *See* WILLIAM H. GREENE, ECONOMETRIC ANALYSIS § 8.3 (2d ed. 1990).

212. The data source is FERC Form-423; Energy Velocity. Regulated utilities report the delivered price of their fuels in FERC Form-423. Energy Velocity estimates transportations costs and subtracts these from the

spot price at the time of delivery instead of at the time of actual sales. As a result, there is a lag between the reported spot price and the spot price at the time of actual sales that ranges from one quarter to one year. Therefore, although the alleged collusion occurred in the first quarter of 2001, the reported spot price did not increase until late 2001. This case study defines the before collusion period as before November 2001 and the during collusion period as between November 2001 and May 2005.

Before one can engage in econometric analysis of the data, it is important to review the data and check whether there are potential coding errors or other issues in the data that deserve further investigation. As discussed in section A, data do not have to be perfectly accurate to be useful in econometric analysis, but it is important to ensure that the data are reasonably reliable. To illustrate the importance of initial data review before running any regression analysis, the original data on the average SPRB coal price are plotted in Figure 4.

Figure 4: Average Price of SPRB Coal before Data Cleaning

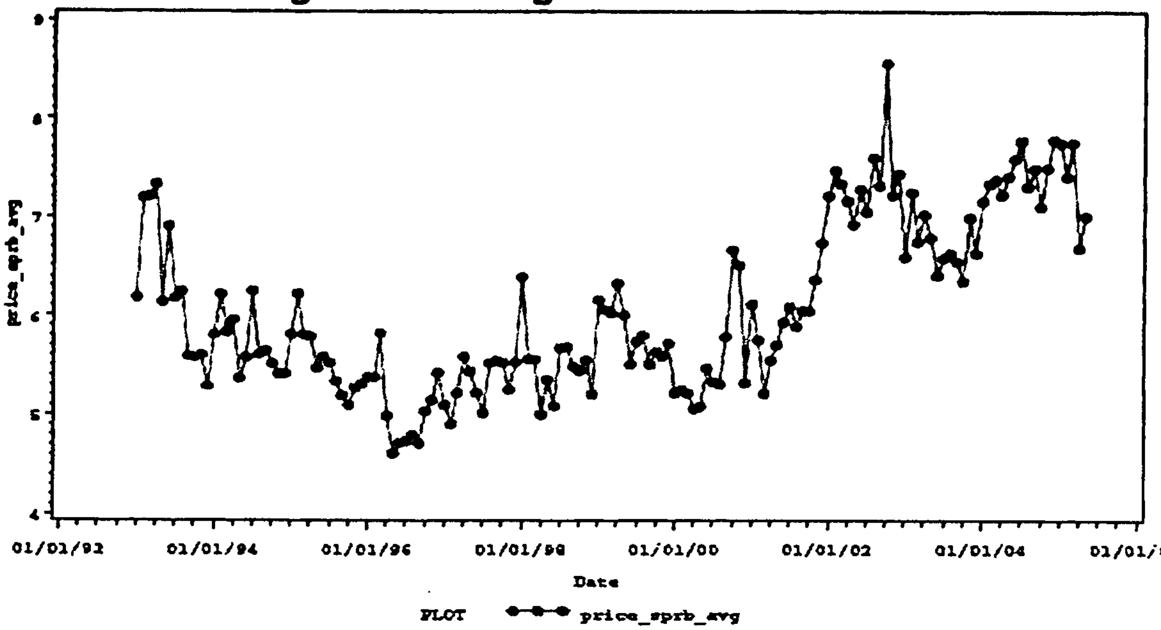


As can be seen clearly from Figure 4, something is not right with the SPRB coal price data before 1998. Further review of the data reveals

delivered costs to provide a comparable FOB mine price for each utility. The average of these monthly spot FOB prices provides data for our dependent variable.

that before 1998, the transportation cost from the coal mines to the utilities was in the column for the spot price and vice versa. Correcting these data and conducting some other cleaning of the data, such as estimating the spot prices that were missing, results in data on the corrected average spot price for SPRB coal that are presented in Figure 5.

Figure 5: Average Price of SPRB Coal



These prices represent the prices of SPRB coal delivered from January 1993 to May 2005. Through most of the 1990s, these prices display significant volatility, but they did not have any trend up or down. Beginning in November 2001, prices began increasing from average levels of around \$5.60/ton in the before period to over \$7.00/ton in the during period. The question is whether this increase in prices was caused by changes in demand, changes in costs, or changes in the behavior of mine operators.

To take into account the major demand and supply factors that potentially could affect the prices of the SPRB coal, data on these factors were obtained from public sources. Supply data include quality measures of SPRB coal (Btu content, ash percentage, and sulfur content); productivity of the mines that produce the SPRB coal as measured by the average tons of SPRB coal produced per person day; and costs of fuels that are used in the production of SPRB coal, including diesel fuel cost and electricity cost.[213] Demand was measured using data on total electric power production.[214]

3. *Assumption of Liability and Causation*

In performing a damage analysis, experts usually assume that the plaintiff has been found liable for the alleged anticompetitive conduct. Accordingly, this case study assumes that SPRB coal producers engaged in collusive behavior from November 2001 to May 2005. As discussed in section C, however, even assuming collusion did occur, one still needs to study whether the observed price increase was at least in part due to other demand or supply changes or could be attributed to the collusive behavior. To help answer this question and estimate the price increase that resulted from the assumed collusion, the dummy variable model and the prediction model under the before-during approach are used to isolate the price effects of the alleged collusive behavior.

4. *Stationarity of the Time Series*

As discussed in section D, when using time series data to conduct regression analysis, one should test whether the time series are stationary. If they are not stationary, different econometric techniques should be employed. To study whether nonstationarity is an issue in the time series data, Dickey-Fuller/Augmented Dickey-Fuller unit root tests were conducted[215] on the SPRB coal price data and certain independent variables.

---

213. Time series data on capacity for SPRB coal production were not available. If the data were available, one may want to consider capacity in the analysis as well since it may affect the supply of SPRB coal.

214. The spot prices of SPRB coal were reported by the utilities at the time of delivery instead of at the time of sales, so there is a lag between the time when the prices were reported and the time when the prices were determined. This delay in reporting is not an issue for other independent variables, such as diesel cost, electricity cost, electricity production and SPRB coal productivity measurement. Therefore, in theory, one should lag these independent variables for certain periods and include these lagged variables as independent variables in the regression as well. However, this undertaking is beyond the scope of this illustration.

215. *See* MADDALA, *supra* note 14, at 548-52. In the Dickey-Fuller/ Augmented Dickey-Fuller test, the null hypothesis is that the time series is nonstationary. *Id.* For a brief introduction to stationarity tests, see MADDALA, *supra* note 14, ch. 14.

The unit root test results indicate that the SPRB coal price data are stationary while some of the independent variables may be nonstationary, such as the diesel price series. However, as pointed out in section D, many unit root tests have a low ability to reject nonstationarity. Therefore, one should view the results of the tests within the context of the economic theory. In this case, if the dependent variable, SPRB coal price, is stationary, it is unlikely that the demand or supply factors that interact with each other to determine the SPRB coal price are nonstationary. If these demand or supply factors were nonstationary, they likely would drive the SPRB coal price to be nonstationary as well.

To further investigate the nonstationarity issue, a Dickey-Fuller/Augmented Dickey-Fuller test and a KPSS test were conducted on the residuals from the OLS regression where the SPRB coal price is the dependent variable and other demand and supply factors along with a collusion dummy variable are included as the explanatory variables.[216] These stationarity tests indicate that the residuals from the OLS regression are stationary. Since the SPRB coal price is stationary, this result again suggests that nonstationarity of other explanatory variables is not an issue in the case study. Thus, an OLS regression analysis was conducted to assess how the assumed collusion affected the SPRB coal price.

5. *The Before-During Approach with the Dummy Variable Model*

The dummy variable model under the before-during approach can be used to measure how assumed collusion changed the price of SPRB coal. The dummy variable is constructed by creating a variable equal to 0 before November 2001 (before period) and equal to 1 from November

216. The OLS regression model specification is discussed in more detail in part H.5. of this chapter. In conducting the Dickey-Fuller/Augmented Dickey-Fuller tests of the OLS regression residuals, one should use the Engle-Granger critical values instead of the standard Dickey-Fuller critical values because the Dickey-Fuller test of the residuals is a co-integration test. For more details on the Engle-Granger critical values, see WALTER ENDERS, APPLIED ECONOMETRIC TIME SERIES 335-42 (2d ed. 2003). However, the exact Engle-Granger critical values for these tests were not available because the number of independent variables in our regression exceeds those listed in the table of Engle-Granger critical values in Enders. Therefore, the KPSS test was also performed. In the KPSS test, the null hypothesis is that the time series is stationary. See MADDALA, *supra* note 14, ch. 14, for more details.

2001 onward (during period). Together with other explanatory variables, the model is specified as

$$SPRB\_price_t = \alpha + \beta_1 \cdot Btu_t + \beta_2 \cdot Ash_t + \beta_3 \cdot lbsSO2_t + \beta_4 \cdot ton\_prod_t$$
$$+ \beta_5 \cdot diesel\_price_t + \beta_6 \cdot electricity\_price_t + \beta_7 \cdot electricity\_production_t$$
$$+ \gamma \cdot dummy_t + \varepsilon_t$$

where Btu_t is the Btu content of the SPRB coal, Ash_t is the ash percentage of the SPRB coal, $lbsSO2_t$ is the pounds of sulfur per million Btu of energy in the coal, ton_prod_t is the average tons of SPRB coal produced per person day, $diesel\_price_t$ is the price of diesel, $electricity\_price_t$ is the price of electricity, $electricity\_production_t$ is the production of electricity, $dummy_t$ is the dummy variable, and ε_t is the error term.

If the assumed collusion led to an increase in the price of SPRB coal, the coefficient on the dummy variable should be positive and statistically significant after controlling for the demand and supply factors that affect the price of SPRB coal. As discussed in part D of this chapter, valid hypothesis testing requires consistent estimates of the standard error of the coefficient estimates. Tests for serial correlation indicate that the error terms in the regression are serially correlated. To estimate the standard errors consistently, the Newey-West procedure was used to obtain the heteroskedasticity and autocorrelation consistent Newey-West standard errors.[217] The result of the OLS regression with the Newey-West standard errors is presented in Table 1.[218]

217. *See* JEFFREY M. WOOLDRIDGE, INTRODUCTORY ECONOMETRICS—A MODERN APPROACH 410-13 (2d ed. 2003).

218. Obtaining the Newey-West standard error requires determining the maximum order of lag in advance. This illustration follows the practice of using the smallest integer greater than or equal to T1/4 where T is the sample size. *See* GREENE, *supra* note 16, at 465.

Table 1: Regression Results for Dummy Variable Method

Dependent Variable:	Average Price of SPRB Coal
R-squared:	0.7132
Adj R-squared:	0.6968
Number of Obs:	149

	Coefficient	Newey-West Standard Error	t-Statistic	P>|t|
Dummy variable for the alleged collusion period	1.2362	0.1798	6.8800	0.0000
Btu content of SPRB coal	-0.0013	0.0011	-1.1300	0.2590
Ash percentage of the SPRB coal	-1.0727	0.7633	-1.4100	0.1620
Pounds of sulfur per million Btu	-1.1665	1.5701	-0.7400	
Average tons of SPRB coal produced per person day	0.0015	0.0024	0.6200	0.5330
Price of diesel	0.0978	0.3283	0.3000	0.7660
Price of electricity	0.0497	0.0483	0.2400	0.8080
Production of electricity	4.33E-10	1.78E-09	0.2400	0.8080
Constant	20.7707	11.0578	1.8800	0.0620

As can be seen from Table 1, the variables that control for supply and demand changes are of the expected signs. Individually they are not statistically significant from zero at the 5 percent level, although they are statistically significant when taken together.[219] The coefficient on the dummy variable is positive and statistically significant. The regression result indicates that on average, SPRB coal prices were \$1.23/ton (or about 22%) higher in the during period compared to the before period, holding constant the exogenous variables.

As discussed in part E of this chapter, the dummy variable model assumes that the relationship between the dependent variable and the explanatory variables in the during period is the same as in the before period. To test this assumption, interaction terms involving the dummy variable and each of the other explanatory variables except the intercept term were added as explanatory variables in the model specification as

219. In a more detailed analysis, other measures of supply and demand might be considered or the estimating equation might be modified to better take into account more specific demand and supply considerations for each of the utilities.

well.[220] The null hypothesis that all of the coefficients on the interaction terms are jointly equal to zero was tested. If this hypothesis is not rejected, then the data are consistent with the view that the relationship between the price of SPRB coal and the demand and supply factors did not change from the before period to the during period. If the null hypothesis is rejected, it suggests that the relationship has changed, either due to the alleged behavior or because of other reasons.[221]

As shown in Table 2, the test rejected the null hypothesis that the relationship between the price of the SPRB coal and the demand and supply factors remained the same in the during period.[222] As discussed in part E of this chapter, even though the test suggests that the coefficients on the demand and supply factors in the during period have changed (contrary to the assumption of the dummy variable model), it does not necessarily mean that the coefficient estimate of the dummy variable will not provide any economically meaningful information. A possibility that should be investigated is that the restrictions on the coefficients of the demand and supply factors imposed by the dummy variable model may not have substantially affected the model's estimate of the effect of the alleged antitrust violation. Accordingly, the next section uses the prediction model to assess the impact of the assumed collusion on the price of SPRB coal.[223]

---

220. These interaction terms are equal to zero in the pre-collusion period and equal to the respective values of the original explanatory variables during the collusion period. *See* MADDALA, *supra* note 14, at 307-10.
221. *See* part E.1. of this chapter.
222. To implement the joint F test correctly, again, the Newey-West covariance matrix is used. *See* WOOLDRIDGE, *supra* note 217.
223. When one fully interacts the dummy variable with other explanatory variables, one in fact is assuming that the regression equations in the before period and in the during period are completely different. *See* MADDALA, *supra* note 14, at 309. For example, the regression equation in the during period can have smaller intercept terms, hence a negative coefficient on the dummy variable, but higher slope terms. As a result, one cannot interpret the coefficient on the dummy variable as the measurement of the average impact of the alleged collusion anymore.

Table 2: Regression Results—Testing for Relationship Change in the Dummy Variable Method

Dependent Variable:	**Average Price of SPRB Coal**
R-squared:	0.7417
Adj R-squared:	0.7126
Number of Obs:	149

	Coefficient	Newey-West Standard Error	t-Statistic	P>|t|
Dummy variable for the alleged collusion period	-57.0152	23.4560	-2.4300	0.0160
Btu content of SPRB coal	-0.0021	0.0016	-1.3000	0.1940
Ash percentage of the SPRB coal	-1.9244	0.8683	-2.2200	0.0280
Pounds of sulfur per million Btu	0.9237	1.7494	0.5300	0.5980
Average tons of SPRB coal produced per person day	0.0038	0.0025	1.5300	0.1280
Price of diesel	-0.0652	0.4501	-0.1400	0.8850
Price of electricity	0.0296	0.0860	0.3400	0.7310
Production of electricity	9.85E-10	2.32E-09	0.4300	0.6710
Interaction term for Btu content of SPRB coal	0.0056	0.0025	2.2600	0.0260
Interaction term for ash percentage of the SPRB coal	2.5203	1.2294	2.0500	0.0420
Interaction term for pounds of sulfur per million Btu	-5.2794	2.4693	-2.1400	0.0340
Interaction term for average tons of SPRB coal produced per person day	0.0038	0.0081	0.4700	0.6370
Interaction term for price of diesel	0.5961	0.5924	1.0100	0.3160
Interaction term for price of electricity	-0.0406	0.1036	-0.3900	0.6960
Interaction term for production of electricity	-6.07E-10	2.96E-09	-0.2100	0.8380
Constant	29.8307	14.1026	2.1200	0.0360
Joint F test for all interaction terms are equal to zero	$F(7, 133) = 3.09$		Prob>F = 0.0048	

6. *The Before-During Approach with the Prediction Model*

The prediction model assumes that the relationship between the price of SPRB coal and the demand and supply factors in the actual world before the alleged anticompetitive behavior would have remained the same but for the alleged behavior in the during period.[224] If the relationship change found in the last section is all due to the collusive behavior, then the prediction model still would be valid.

To use the prediction model to estimate the impact of the assumed collusion on the price of SPRB coal, first run an OLS regression similar to the dummy variable model, but using the data in the before collusion period only and excluding the dummy variable from the explanatory variables. The result of this regression is presented in Table 3.

Table 3: Regression Results for Prediction Model

Dependent Variable:	Average Price of SPRB Coal
R-squared:	0.2104
Adj R-squared:	0.1540
Number of Obs:	106

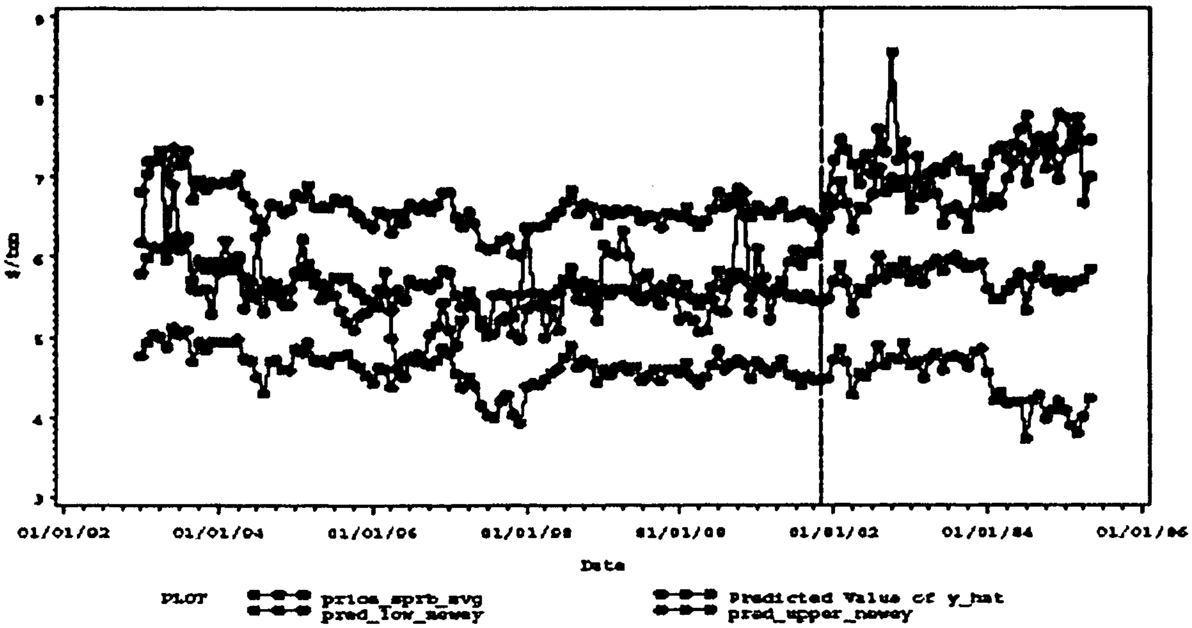
	Coefficient	Newey-West Standard Error	t-Statistic	P>|t|
Btu content of SPRB coal	-0.0021	0.0015	-1.3300	0.1870
Ash percentage of the SPRB coal	-1.9244	0.8532	-2.2600	0.0260
Pounds of sulfur per million Btu	0.9237	1.7189	0.5400	0.5920
Average tons of SPRB coal produced per person day	0.0038	0.0025	1.5600	0.1220
Price of diesel	-0.0652	0.4423	-0.1500	0.8830
Price of electricity	0.0296	0.0845	0.3500	0.7270
Production of electricity	9.85E-10	2.28E-09	0.4300	0.6660
Constant	29.8307	13.8571	2.1500	0.0340

Given the coefficients from the before period OLS regression, the next step is to predict the price of SPRB coal in the but-for world without collusion, using the data on the explanatory variables in the during period. Figure 6 shows the actual price of the SPRB coal, the predicted price of the SPRB coal based on relationships estimated in the before

224. See the discussion in part A of this chapter.

period, and the lower and upper bound of the 95 percent confidence interval for the predicted price of the SPRB coal.[225]

Figure 6: Average vs. Predicted Price of SPRB Coal—Newey Standard Error



As can be seen in Figure 6, in every month in the during period, the actual price of the SPRB coal is higher than its predicted price. Moreover, in 28 of the 43 months in the during period, the actual price of the SPRB coal is higher than the upper bound of the 95 percent confidence interval for the predicted price of the SPRB coal. The average predicted price of the SPRB coal for the during period is \$5.72/ton, while the average actual price of the SPRB coal is \$7.16/ton in the during period. Therefore, according to the prediction model, the

225. The 95% confidence interval for the predicted value of the SPRB coal price means that the there is a 95% "confidence" that the price of the SPRB coal in the but-for world will fall within this interval. See GREENE, *supra* note 16, § 6.6, for detailed discussion on how to construct a confidence interval for the predicted values. Constructing a confidence interval for the predicted value requires estimating the variances of the predicted value and the error term. The illustration uses the Newey-West standard errors in estimating the variance of the predicted value and uses the residuals from the OLS regression to estimate the variance of the error term.

assumed collusion resulted in a \$1.66/ton (or about a 30 percent) increase of the SPRB coal price. The prediction model results suggest a somewhat larger impact of the assumed collusion on the price of the SPRB coal compared to the \$1.23/ton (or 22 percent) increase of the SPRB coal price based on the dummy variable model. The results of both of these approaches are consistent with the alleged behavior causing higher prices. However, the prediction approach in this case would yield greater damages.

Given that the test shown in Table 2 above rejected the hypothesis that the relationship between the price of the SPRB coal and the demand and supply factors were the same in the before and during periods, the prediction approach would appear to be the more reliable quantification of the impact of the hypothesized conspiracy. However, to determine if the prediction approach is clearly superior, one could test the prediction approach for a portion of the period before the alleged anticompetitive behavior, as discussed above in part D. If it predicted prices before the alleged anticompetitive act reasonably accurately, then the prediction approach presumably would be more accurate in this example. If it does not predict prices before the alleged anticompetitive act that well, then it would be more accurate to say that the impact of the alleged anticompetitive behavior likely falls in the range indicated by the two methodologies.

7. *Benchmark Approach*

In a benchmark analysis, the first question is whether there are any products that might provide a reasonably good benchmark. As discussed in part F above, the benchmark should be (1) subject to similar supply and demand factors as SPRB coal and (2) unaffected by alleged anticompetitive acts. This illustration will use the price of bituminous coal (BIT).[226]

An econometric analysis using the benchmark approach can be implemented in several ways. To keep the illustration relatively simple,

226. Utilities also purchase this coal to generate electricity, and so BIT's prices likely reflect many of the same cost and demand factors. There is no allegation of the conspiracy extending to BIT. Moreover, most utilities cannot substitute SPRB and BIT, at least in the short run, so the alleged conspiracy should have no direct effect on the price of BIT. In *Federal Trade Commission v. Arch Coal, Inc.*, the court found "the relevant product market is no broader and no narrower than SPRB coal." FTC v. Arch Coal, Inc., 329 F. Supp. 2d 109, 123 (2004).

the impact of the alleged conspiracy is tested using the dummy variable model discussed above. Most of the supply and demand factors used in the before-during approaches are replaced by the benchmark BIT price, since it should be capturing those effects.[227] The model includes the same dummy variable, as well as three variables to control for the quality of SPRB coal. The estimating equation is:

$$SPRB\_price_t = \alpha + \beta_1 \cdot Btu_t + \beta_2 \cdot Ash_t + \beta_3 \cdot lbsSO2_t + \beta_4 \cdot BIT\_price_t + \gamma \cdot dummy_t + \varepsilon_t$$

where BIT_price_t is the price of BIT coal in period t. The results of this regression are shown in Table 4.

Table 4: Regression Results for Benchmark Approach

Dependent Variable:	Average Price of SPRB Coal
R-squared:	0.7191
Adj R-squared:	0.7093
Number of Obs:	149

	Coefficient	Newey-West Standard Error	t-Statistic	P>|t|
Dummy variable for the alleged collusion period	1.1527	0.1867	6.1700	0.0000
Btu content of SPRB coal	-0.0005	0.0010	-0.4800	0.6300
Ash percentage of the SPRB coal	-1.1329	0.6171	-1.8400	0.0680
Pounds of sulfur per million Btu	-1.1130	1.3201	-0.8400	0.4010
Average tons of SPRB coal produced per person day	0.0063	0.0028	2.2800	0.0240
Constant	15.7235	9.7266	1.6200	0.1080

As can be seen in Table 4, the BIT price is of the expected sign (positive) and is statistically significant. The results show a statistically significant impact on SPRB prices in the period during the alleged conspiracy, again consistent with the alleged conspiracy increasing price. This specification of the benchmark dummy variable model estimates a \$1.15/ton (or about 21 percent) increase in the SPRB coal price, which is somewhat smaller than the reduced form estimates in the before-during

227. The supply and demand factors could still be included in the model, and if they affect the benchmark differently than the price in question, including them will generally improve the model.

models using either the dummy variable or prediction approaches. Although BIT should not have been directly affected by the alleged anticompetitive act related to SPRB coal, it is possible that the customers who bought BIT faced substantially different supply and demand conditions than the plants purchasing the SPRB coal. In deciding whether the benchmark approach yields a more accurate quantification of the alleged anticompetitive behavior than the dummy variable or prediction approaches, one would need to analyze in more depth the comparability of the exogenous influences on BIT and SPRB to determine how good a benchmark BIT is for SPRB. If there are substantial questions about how comparable BIT is to SPRB, then the other approaches likely would yield more accurate estimates.

Exhibit 4

MOSTLY HARMLESS ECONOMETRICS

An Empiricist's Companion

Joshua D. Angrist and Jörn-Steffen Pischke

Published by Princeton University Press, 41 William Street,
Princeton, New Jersey 08540
In the United Kingdom: Princeton University Press,
6 Oxford Street, Woodstock, Oxfordshire OX20 1TW

Library of Congress Cataloging-in-Publication Data

Angrist, Joshua David.
Mostly harmless econometrics : an empiricist's companion /
Joshua D. Angrist, Jörn-Steffen Pischke.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-691-12034-8 (hardcover : alk. paper) —
ISBN 978-0-691-12035-5 (pbk. : alk. paper) 1. Econometrics.
2. Regression analysis. I. Pischke, Jörn-Steffen. II. Title.
HB139.A54 2008
330.01'5195—dc22 2008036265

British Library Cataloging-in-Publication Data is available

This book has been composed in Sabon
with Hel. Neue Cond. family display

Illustrations by Karen Norberg

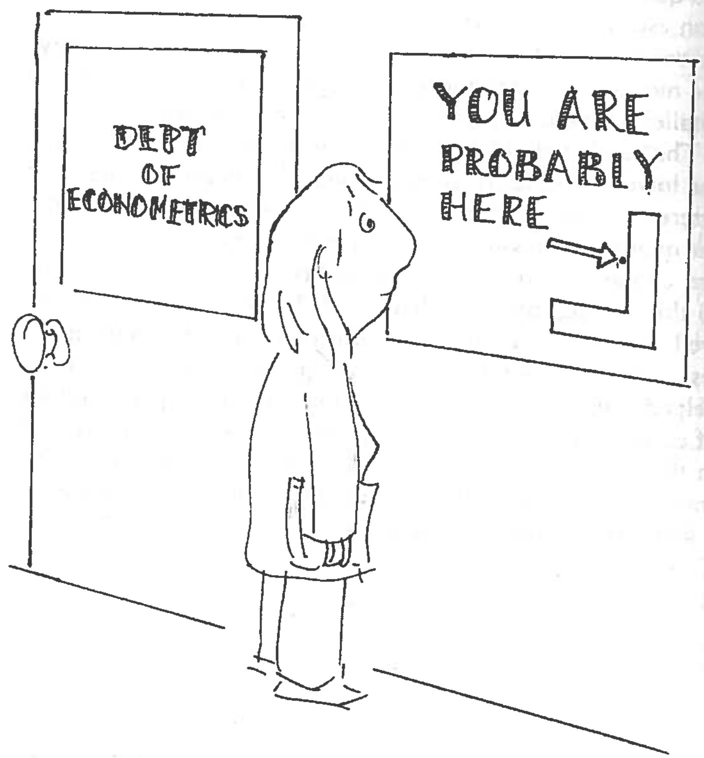
Printed on acid-free paper. ∞

press.princeton.edu

Printed in the United States of America

5 7 9 10 8 6

CONTENTS



Chapter 8

Nonstandard Standard Error Issues

> We have normality. I repeat, we have normality. Anything you still can't cope with is therefore your own problem.
>
> Douglas Adams, *The Hitchhiker's Guide to the Galaxy*

Today, software packages routinely compute asymptotic standard errors derived under weak assumptions about the sampling process or underlying model. For example, you get regression standard errors based on formula (3.1.7) using the Stata option `robust`. Robust standard errors improve on old-fashioned standard errors because the resulting inferences are asymptotically valid when the regression residuals are heteroskedastic, as they almost certainly are when regression approximates a nonlinear conditional expectation function (CEF). In contrast, old-fashioned standard errors are derived assuming homoskedasticity. The hangup here is that estimates of robust standard errors can be misleading when the asymptotic approximation that justifies these estimates is not very good. The first part of this chapter looks at the failure of asymptotic inference with robust standard error estimates and some simple palliatives.

A pillar of traditional cross section inference—and the discussion in section 3.1.3—is the assumption that the data are independent. Each observation is treated as a random draw from the same population, uncorrelated with the observation before or after. We understand today that this sampling model is unrealistic and potentially even foolhardy. Much as in the time series studies common in macroeconomics, cross section analysts must worry about correlation between observations. The most important form of dependence arises in data with a group structure—for example, the test scores of children observed within classes or schools. Children in the same school or class tend to have test scores that are correlated, since they are subject to some of the same environmental and family background influences. We call this correlation the clustering problem, or the Moulton problem, after Moulton (1986), who made it famous. A closely related problem is correlation over time in the data sets commonly used to implement differences-in-differences (DD) estimation strategies. For example, studies of state-level minimum wages must confront the fact that state average employment rates are correlated over time. We call this the serial correlation problem, to distinguish it from the Moulton problem.

Researchers plagued by clustering and serial correlation also have to confront the fact that the simplest fixups for these problems, like Stata's `cluster` option, may not be very good. The asymptotic approximation relevant for clustered or serially correlated data relies on a large number of clusters or time series observations. Alas, we are not always blessed with many clusters or long time series. The resulting inference problems are not always insurmountable, though often the best solution is to get more data. Econometric fixups for clustering and serial correlation are discussed in the second part of this chapter. Some of the material in this chapter is hard to work through without matrix algebra, so we take the plunge and switch to a mostly matrix motif.

8.1 The Bias of Robust Standard Error Estimates★

In matrix notation

$$\hat{\beta} = \left[\sum_i X_i X_i'\right]^{-1} \sum_i X_i Y_i = (X'X)^{-1}X'y,$$

where X is the $N \times K$ matrix with rows X_i' and y is the $N \times 1$ vector of Y_i's. We saw in section 3.1.3 that $\hat{\beta}$ has an

asymptotically normal distribution. We can write:

$$\sqrt{N}(\hat{\beta} - \beta) \sim N(0, \Omega)$$

where Ω is the asymptotic covariance matrix and $\beta = E[X_i X_i']^{-1}E[X_i Y_i]$. Repeating (3.1.7), the formula for Ω in this case is

$$\Omega_r = E[X_i X_i']^{-1}E[X_i X_i' e_i^2]E[X_i X_i']^{-1}, \qquad (8.1.1)$$

where $e_i = Y_i - X_i'\beta$. When residuals are homoskedastic, the covariance matrix simplifies to $\Omega_c = \sigma^2 E[X_i X_i']^{-1}$, where $\sigma^2 = E[e_i^2]$.

We are concerned here with the bias of robust standard error estimates in independent samples (i.e., no clustering or serial correlation). To simplify the derivation of bias, we assume that the regressor vector can be treated as fixed, as it would be if we sampled stratifying on X_i. Nonstochastic regressors gives a benchmark sampling model that is often used to look at finite-sample distributions. It turns out that we miss little of theoretical importance by making this assumption, while simplifying the derivations considerably.

With fixed regressors, we have

$$\Omega_r = \left(\frac{X'X}{N}\right)^{-1}\left(\frac{X'\Psi X}{N}\right)\left(\frac{X'X}{N}\right)^{-1}, \qquad (8.1.2)$$

where

$$\Psi = E[ee'] = diag(\psi_i)$$

is the covariance matrix of residuals. Under homoskedasticity, $\psi_i = \sigma^2$ for all i and we get

$$\Omega_c = \sigma^2\left(\frac{X'X}{N}\right)^{-1}.$$

Asymptotic standard errors are given by the square root of the diagonal elements of Ω_r and Ω_c, after removing the asymptotic normalization by dividing by N.

In practice, the pieces of the asymptotic covariance matrix are estimated using sample moments. An old-fashioned or conventional covariance matrix estimator is

$$\hat{\Omega}_c = (X'X)^{-1}\hat{\sigma}^2 = (X'X)^{-1}\left(\sum \frac{\hat{e}_i^2}{N}\right),$$

where $\hat{e}_i = Y_i - X_i'\hat{\beta}$ is the estimated regression residual and

$$\hat{\sigma}^2 = \sum \frac{\hat{e}_i^2}{N}$$

estimates the residual variance. The corresponding robust covariance matrix estimator is

$$\hat{\Omega}_r = N(X'X)^{-1}\left(\sum \frac{X_iX_i'\hat{e}_i^2}{N}\right)(X'X)^{-1}. \tag{8.1.3}$$

We can think of the middle term as an estimator of the form $\sum \frac{X_iX_i'\hat{\psi}_i}{N}$, where $\hat{\psi}_i = \hat{e}_i^2$ estimates ψ_i.

By the law of large numbers and Slutsky's theorem, $N\hat{\Omega}_c$ converges in probability to Ω_c, while $N\hat{\Omega}_r$ converges to Ω_r. But in finite samples, both variance estimators are biased. The bias in $\hat{\Omega}_c$ is well-known from classical least squares theory and easy to correct. Less appreciated is the fact that if the residuals are homoskedastic, the robust estimator is more biased than the conventional estimator, perhaps a lot more. From this we conclude that robust standard errors can be more misleading than conventional standard errors in situations where heteroskedasticity is modest. We also propose a rule of thumb that uses the maximum of old-fashioned and robust standard errors to avoid gross misjudgments of precision.

Our analysis begins with the bias of $\hat{\Omega}_c$. With nonstochastic regressors, we have

$$E[\hat{\Omega}_c] = (X'X)^{-1}\hat{\sigma}^2 = (X'X)^{-1}\left(\sum \frac{E(\hat{e}_i^2)}{N}\right).$$

To analyze $E[\hat{e}_i^2]$, start by expanding $\hat{e} = y - X\hat{\beta}$:

$$\hat{e} = y - X(X'X)^{-1}X'y = [I_N - X(X'X)^{-1}X'](X\beta + e) = Me,$$

where e is the vector of population residuals, $M = I_N - X(X'X)^{-1}X'$ is a nonstochastic residual-maker matrix with

ith row m_i', and I_N is the $N \times N$ identity matrix. Then $\hat{e}_i = m_i'e$, and

$$\begin{aligned} E(\hat{e}_i^2) &= E(m_i'ee'm_i) \\ &= m_i'\Psi m_i. \end{aligned}$$

To simplify further, write $m_i = \ell_i - h_i$, where ℓ_i is the ith column of I_N and $h_i = X(X'X)^{-1}X_i$, the ith column of the projection matrix $H = X(X'X)^{-1}X'$. Then

$$\begin{aligned} E(\hat{e}_i^2) &= (\ell_i - h_i)'\Psi(\ell_i - h_i) \\ &= \psi_i - 2\psi_i h_{ii} + h_i'\Psi h_i, \end{aligned} \tag{8.1.4}$$

where h_{ii}, the ith diagonal element of H, satisfies

$$h_{ii} = h_i'h_i = X_i'(X'X)^{-1}X_i. \tag{8.1.5}$$

Parenthetically, h_{ii} is called the *leverage* of the ith observation. Leverage tells us how much pull a particular value of X_i exerts on the regression line. Note that the ith fitted value (ith element of Hy) is

$$\hat{Y}_i = h_i'y = h_{ii}Y_i + \sum_{j\neq i} h_{ij}Y_j. \tag{8.1.6}$$

A large h_{ii} means that the ith observation has a large impact on the ith predicted value. In a bivariate regression with a single regressor, x_i,

$$h_{ii} = \frac{1}{N} + \frac{(x_i - \bar{x})^2}{\sum (x_j - \bar{x})^2}. \tag{8.1.7}$$

This shows that leverage increases when x_i is far the mean. In addition to (8.1.6), we know that h_{ii} is a number that lies in the interval $[0, 1]$ and that $\sum_{i=1}^{N} h_{ii} = K$, the number of regressors (see, e.g., Hoaglin and Welsch, 1978).[1]

[1] The property $\sum_{i=1}^{N} h_{ii} = K$ comes from the fact that H is idempotent, and so has trace equal to rank. We can also use (8.1.7) to verify that in a bivariate regression, $\sum_{i=1}^{N} h_{ii} = 2$.

Suppose residuals are homoskedastic, so that $\psi_i = \sigma^2$. Then (8.1.4) simplifies to

$$E(\hat{e}_i^2) = \sigma^2[1 - 2h_{ii} + h_i'h_i] = \sigma^2(1 - h_{ii}) < \sigma^2.$$

So $\hat{\Omega}_c$ tends to be too small. Using the properties of h_{ii}, we can go one step further:

$$\sum \frac{E(\hat{e}_i^2)}{N} = \sigma^2 \sum \frac{1 - h_{ii}}{N} = \sigma^2 \left(\frac{N - \kappa}{N}\right).$$

Thus, the bias in $\hat{\Omega}_c$ can be fixed by a simple degrees-of-freedom correction: divide by $N - \kappa$ instead of N in the formula for $\hat{\sigma}^2$. This correction is used by default in most regression software.

We now want to show that under homoskedasticity, the bias in $\hat{\Omega}_r$ is likely to be worse than the bias in $\hat{\Omega}_c$. The expected value of the robust covariance matrix estimator is

$$E[\hat{\Omega}_r] = N(X'X)^{-1}\left(\sum \frac{X_i X_i' E(\hat{e}_i^2)}{N}\right)(X'X)^{-1}, \qquad (8.1.8)$$

where $E(\hat{e}_i^2)$ is given by (8.1.4). Under homoskedasticity, $\psi_i = \sigma^2$ and we have $E(\hat{e}_i^2) = \sigma^2(1 - h_{ii})$ as in $\hat{\Omega}_c$. It's clear, therefore, that the bias in $\hat{e}_i^2$ tends to pull robust standard errors down. The general expression, (8.1.8), is hard to evaluate, however. Chesher and Jewitt (1987) show that as long as there is not "too much" heteroskedasticity, robust standard errors based on $\hat{\Omega}_r$ are indeed biased downward.[2]

How do we know that $\hat{\Omega}_r$ is likely to be *more* biased than $\hat{\Omega}_c$? Partly this comes from Monte Carlo evidence (e.g., MacKinnon and White, 1985, and our own small study, discussed below). We also prove this here for a bivariate example, where the single regressor, $\tilde{x}_i$, is assumed to be in deviations-from-means form, so there is a single coefficient. In this case, the estimator of interest is $\hat{\beta}_1 = \frac{\sum \tilde{x}_i y_i}{\sum \tilde{x}_i^2}$ and the leverage is

[2]In particular, as long as the ratio of the largest ψ_i to the smallest ψ_i is less than 2, robust standard errors are biased downward.

$h_{ii} = \frac{\tilde{x}_i^2}{\sum \tilde{x}_i^2}$ (we lose the $\frac{1}{N}$ term in (8.1.7) by partialing out the constant). Let $s_x^2 = \frac{\sum \tilde{x}_i^2}{N}$. For the conventional covariance estimator, we have

$$E[\hat{\Omega}_c] = \frac{\sigma^2}{Ns_x^2}\left[\frac{\sum(1 - h_{ii})}{N}\right] = \frac{\sigma^2}{Ns_x^2}\left[1 - \frac{1}{N}\right],$$

so the bias here is small. A simple calculation using (8.1.8) shows that under homoskedasticity, the robust estimator has expectation:

$$\begin{aligned} E[\hat{\Omega}_r] &= \frac{\sigma^2}{Ns_x^2} \sum \frac{(1 - h_{ii})}{N}\left(\frac{\tilde{x}_i^2}{s_x^2}\right) \\ &= \frac{\sigma^2}{Ns_x^2} \sum (1 - h_{ii})h_{ii} = \frac{\sigma^2}{Ns_x^2}\left[1 - \sum h_{ii}^2\right]. \end{aligned}$$

The bias of $\hat{\Omega}_r$ is therefore worse than the bias of $\hat{\Omega}_c$ if $\sum h_{ii}^2 > \frac{1}{N}$, as it is by Jensen's inequality unless the regressor has constant leverage, in which case $h_{ii} = \frac{1}{N}$ for all i.[3]

We can reduce the bias in $\hat{\Omega}_r$ by trying to get a better estimator of ψ_i, say $\hat{\psi}_i$. The estimator $\hat{\Omega}_r$ sets $\hat{\psi}_i = \hat{e}_i^2$, as proposed by White (1980a) and our starting point in this section. The residual variance estimators discussed in MacKinnon and White (1985) include this and three others:

$$HC_0 : \hat{\psi}_i = \hat{e}_i^2$$

$$HC_1 : \hat{\psi}_i = \frac{N}{N - \kappa}\hat{e}_i^2$$

[3]Think of h_{ii} as a random variable with a uniform distribution in the sample. Then

$$E[h_{ii}] = \frac{\sum h_{ii}}{N} = \frac{1}{N},$$

and

$$E[h_{ii}^2] = \frac{\sum h_{ii}^2}{N} > (E[h_{ii}])^2 = \left(\frac{1}{N}\right)^2$$

by Jensen's inequality unless h_{ii} is constant. Therefore $\sum h_{ii}^2 > \frac{1}{N}$. The constant leverage case occurs when $(\tilde{x}_i)^2$ is constant.

$$HC_2 : \hat{\psi}_i = \frac{1}{1-h_{ii}}\hat{e}_i^2$$
$$HC_3 : \hat{\psi}_i = \frac{1}{(1-h_{ii})^2}\hat{e}_i^2.$$

HC_1 is a simple degrees of freedom correction as is used for $\hat{\Omega}_c$. HC_2 uses the leverage to give an unbiased estimate of the variance of the ith residual when the residuals are homoskedastic, while HC_3 approximates a jackknife estimator.[4] In the applications we've seen, the estimated standard errors tend to get larger as we go down the list from HC_0 to HC_3, but this is not a theorem.

Time-Out for the Bootstrap

Bootstrapping is a resampling scheme that offers an alternative to inference based on asymptotic formulas. A bootstrap sample is a sample drawn from our own data. In other words, if we have a sample of size N, we treat this sample as if it were the population and draw repeatedly from it (with replacement). The bootstrap sampling distribution is the distribution of an estimator across many draws of this sort. Intuitively, we expect the sampling distribution constructed by sampling from our own data to provide a good approximation to the sampling distribution we are after.

There are many ways to bootstrap regression estimates. The simplest is to draw pairs of $\{Y_i, X_i\}$ values, sometimes called the "pairs bootstrap" or a "nonparametric bootstrap." Alternatively, we can keep the X_i values fixed, draw from the distribution of residuals ($\hat{e}_i$), and create a new estimate of the dependent variable based on the predicted value and the residual draw for each observation. This procedure, which is a type of parametric bootstrap, mimics a sample drawn with nonstochastic regressors and ensures that X_i and the regression residuals are independent. On the other hand, we don't want independence if we're interested in standard errors under heteroskedasticity. An alternative residual bootstrap, called the wild bootstrap, draws $X_i'\hat{\beta} + \hat{e}_i$ (which, of course, is just the original Y_i) with probability 0.5, and $X_i'\hat{\beta} - \hat{e}_i$ otherwise (see, e.g., Mammen, 1993, and Horowitz, 1997). This preserves the relationship between residual variances and X_i observed in the original sample, while imposing mean-independence of residuals and regressors, a restriction that improves bootstrap inference when true.

[4] A jackknife variance estimator estimates sampling variance from the empirical distribution generated by omitting one observation at a time. Stata computes HC_1, HC_2, and HC_3. You can also use a trick suggested by Messer and White (1984): divide Y_i and X_i by $\sqrt{\hat{\psi}_i}$ and instrument the transformed model by $X_i/\sqrt{\hat{\psi}_i}$ for your preferred choice of $\hat{\psi}_i$.

Bootstrapping is useful as a computer-intensive but otherwise straightforward calculator for asymptotic standard errors. The bootstrap calculator is especially useful when the asymptotic distribution of an estimator is hard to compute or involves a number of steps (e.g., the asymptotic distributions of the quantile regression and quantile treatment effects estimates discussed in chapter 7 require the estimation of densities). Typically, however, we have no problem deriving or evaluating asymptotic formulas for the standard errors of OLS estimates.

More relevant in this context is the use of the bootstrap to improve inference. Improvements in inference potentially come in two forms: (1) a reduction in finite-sample bias in estimators that are consistent (for example, the bias in estimates of robust standard errors) and (2) inference procedures which make use of the fact that the bootstrap sampling distribution of test statistics may be closer to the finite-sample distribution of interest than the relevant asymptotic approximation. These two properties are called asymptotic refinements (see, e.g., Horowitz, 2001).

Here we are mostly interested in use of the bootstrap for asymptotic refinement. The asymptotic distribution of regression estimates is easy enough to compute, but we worry that the traditional robust covariance estimator (HC_0) is biased. The bootstrap can be used to estimate this bias, and then, by a simple transformation, to construct standard error estimates that are less biased. However, for now at least, bootstrap bias correction of regression standard errors is not often used in empirical practice, perhaps because the bias calculation is not automated and perhaps because bootstrap bias corrections introduce extra variability. Also, for simple estimators like regression coefficients, analytic bias corrections such as HC_2 and HC_3 are readily available (e.g., in Stata).

An asymptotic refinement can also be obtained for hypothesis tests (and confidence intervals) based on statistics that are *asymptotically pivotal*. These are statistics that have asymptotic distributions that do not depend on any unknown parameters. An example is a t-statistic: this is asymptotically standard normal. Regression coefficients are not asymptotically pivotal; they have an asymptotic distribution that depends on the unknown residual variance. To refine inference for regression coefficients, you calculate the t-statistic in each bootstrap sample and compare the analogous t-statistic from your original sample to this bootstrap "t-distribution." A hypothesis is rejected if the absolute value of the original t-statistic is above, say, the 95th percentile of the absolute values from the bootstrap distribution.

Theoretical appeal notwithstanding, as applied researchers, we don't like the idea of bootstrapping pivotal statics very much. This is partly because we're not only (or even primarily) interested in formal hypothesis testing: we like to see the standard errors in parentheses under our regression coefficients. These provide a summary measure of precision that can be used to construct confidence intervals, compare estimators, and test any hypothesis that strikes us, now or later. In our view, therefore, practitioners worried about the finite-sample behavior of robust standard errors should focus on bias corrections like HC_2 and HC_3. As we show below, for moderate heteroskedasticity at least, an inference strategy that uses the larger of conventional and bias-corrected standard errors often seems to give us the best of both worlds: reduced bias with a minimal loss of precision.

An Example

For further insight into the differences between robust covariance estimators, we analyze a simple but important example that has featured in earlier chapters in this book. Suppose you

are interested in an estimate of β_1 in the model

$$\text{Y}_i = \beta_0 + \beta_1 \text{D}_i + \varepsilon_i, \tag{8.1.9}$$

where D_i is a dummy variable. The OLS estimate of β_1 is the difference in means between those with D_i switched on and off. Denoting these subsamples by the subscripts 1 and 0, we have

$$\hat{\beta}_1 = \bar{\text{Y}}_1 - \bar{\text{Y}}_0.$$

For the purposes of this derivation we think of D_i as nonrandom, so that $\sum \text{D}_i = N_1$ and $\sum (1 - \text{D}_i) = N_0$ are fixed. Let $r = N_1/N$.

We know something about the finite-sample behavior of $\hat{\beta}_1$ from statistical theory. If Y_i is normal with equal but unknown variance in both the $\text{D}_i = 1$ and $\text{D}_i = 0$ populations, then the conventional t-statistic for $\hat{\beta}_1$ has a t-distribution. This is the classic two-sample t-test. Heteroskedasticity in this context means that the variances in the $\text{D}_i = 1$ and $\text{D}_i = 0$ populations are different. In this case, the testing problem in small samples becomes surprisingly difficult: the exact small-sample distribution for even this simple problem is unknown.[5] The robust variance estimators HC_0–HC_3 give asymptotic approximations to the unknown finite-sample distribution for the case of unequal variances.

The differences between HC_0, HC_1, HC_2, and HC_3 are differences in how the sample variances in the two groups defined by D_i are processed. Define $S_j^2 = \sum_{\text{D}_i = j} (\text{Y}_i - \bar{\text{Y}}_j)^2$ for $j = 0, 1$. The leverage in this example is

$$h_{ii} = \begin{cases} \frac{1}{N_0} & \text{if } \text{D}_i = 0 \\ \frac{1}{N_1} & \text{if } \text{D}_i = 1 \end{cases}.$$

Using this, it's straightforward to show that the five variance estimators we've been discussing are

$$\textit{Conventional}: \frac{N}{N_0 N_1}\left(\frac{S_0^2 + S_1^2}{N-2}\right) = \frac{1}{Nr(1-r)}\left(\frac{S_0^2 + S_1^2}{N-2}\right)$$

[5] This is called the Behrens-Fisher problem (see, e.g., DeGroot and Schervish, 2001, chap. 8).

$$HC_0: \frac{S_0^2}{N_0^2} + \frac{S_1^2}{N_1^2}$$

$$HC_1: \frac{N}{N-2}\left(\frac{S_0^2}{N_0^2} + \frac{S_1^2}{N_1^2}\right)$$

$$HC_2: \frac{S_0^2}{N_0(N_0-1)} + \frac{S_1^2}{N_1(N_1-1)}$$

$$HC_3: \frac{S_0^2}{(N_0-1)^2} + \frac{S_1^2}{(N_1-1)^2}.$$

The conventional estimator pools subsamples: this is efficient when the two variances are the same. The White (1980a) estimator, HC_0, adds separate estimates of the sampling variances of the means, using the consistent (but biased) variance estimators, $\frac{S_j^2}{N_j}$. The HC_2 estimator uses unbiased estimators of the sample variance for each group, since it makes the correct degrees-of-freedom correction. HC_1 makes a degrees-of-freedom correction outside the sum, which will help but is generally not quite correct. Since we know HC_2 to be the unbiased estimate of the sampling variance under homoskedasticity, HC_3 must be too big.[6] Note that with $r = 0.5$, a case where the regression design is said to be balanced, the conventional estimator equals HC_1 and all five estimators differ little.

A small Monte Carlo study based on (8.1.9) illustrates the pluses and minuses of alternative estimators and the extent to which a simple rule of thumb goes a long way toward ameliorating the bias of the HC class. We choose $N = 30$ to highlight small sample issues, and $r = 0.10$ (10 percent treated), which implies $h_{ii} = \frac{1}{3}$ if $D_i = 1$ and $h_{ii} = \frac{1}{27}$ if $D_i = 0$. This is a highly unbalanced design. We draw residuals from the distributions:

$$\varepsilon_i \sim \begin{cases} N(0, \sigma^2) & \text{if } D_i = 0 \\ N(0, 1) & \text{if } D_i = 1 \end{cases}$$

and report results for three cases. The first has lots of heteroskedasticity, with $\sigma = 0.5$, while the second has relatively

[6]In this simple example, HC_2 is unbiased whether or not residuals are homoskedastic.

little heteroskedasticity, with $\sigma = 0.85$. No heteroskedasticity is the benchmark case.

Table 8.1.1 displays the results. Columns 1 and 2 report means and standard deviations of the various standard error estimates across 25,000 replications of the sampling experiment. The standard deviation of $\hat{\beta}_1$ is the sampling variance we are trying to measure. With lots of heteroskedasticity, as in the upper panel of the table, conventional standard errors are badly biased and, on average, only about half the size of the Monte Carlo sampling variance that constitutes our target. On the other hand, while the robust standard errors perform better, except for HC_3, they are still too small.[7]

The standard errors are themselves estimates and have considerable sampling variability. Especially noteworthy is the fact that the robust standard errors have much higher sampling variability than the conventional standard errors, as can be seen in column 2.[8] The sampling variability of estimated standard errors further increases when we attempt to reduce bias by dividing the residuals by $1 - h_{ii}$ (HC_2) or $(1 - h_{ii})^2$ (HC_3). The worst case is HC_3, with a standard deviation about 50 percent above the standard deviation of the White (1980a) standard error, HC_0.

The last two columns in the table show empirical rejection rates in a nominal 5 percent test for the hypothesis $\beta_1 = 0$, the population parameter in this case. The test statistics are compared with a normal distribution and to a t-distribution with $N - 2$ degrees of freedom. Rejection rates are far too high for all tests, even with HC_3. Using a t-distribution rather than a normal distribution helps only marginally.

[7]Although HC_2 is an unbiased estimator of the sampling variance, the mean of the HC_2 standard errors across sampling experiments (0.52) is still below the standard deviation of $\hat{\beta}_1$ (0.59). This comes from the fact that the standard error is the square root of the sampling variance, the sampling variance is itself estimated and hence has sampling variability, and the square root is a concave function.

[8]The large sampling variance of robust standard error estimators is noted by Chesher and Austin (1991). Kauermann and Carroll (2001) propose an adjustment to confidence intervals to correct for this.

Table 8.1.1
Monte Carlo results for robust standard error estimates

Parameter Estimate	Mean (1)	Standard Deviation (2)	Empirical 5% Rejection Rates: Normal (3)	Empirical 5% Rejection Rates: t (4)
A. Lots of heteroskedasticity				
$\hat{\beta}_1$	−.001	.586		
Standard Errors				
Conventional	.331	.052	.278	.257
HC_0	.417	.203	.247	.231
HC_1	.447	.218	.223	.208
HC_2	.523	.260	.177	.164
HC_3	.636	.321	.130	.120
max(HC_0, Conventional)	.448	.172	.188	.171
max(HC_1, Conventional)	.473	.190	.173	.157
max(HC_2, Conventional)	.542	.238	.141	.128
max(HC_3, Conventional)	.649	.305	.107	.097
B. Little heteroskedasticity				
$\hat{\beta}_1$	.004	.600		
Standard Errors				
Conventional	.520	.070	.098	.084
HC_0	.441	.193	.217	.202
HC_1	.473	.207	.194	.179
HC_2	.546	.250	.156	.143
HC_3	.657	.312	.114	.104
max(HC_0, Conventional)	.562	.121	.083	.070
max(HC_1, Conventional)	.578	.138	.078	.067
max(HC_2, Conventional)	.627	.186	.067	.057
max(HC_3, Conventional)	.713	.259	.053	.045
C. No heteroskedasticity				
$\hat{\beta}_1$	−.003	.611		
Standard Errors				
Conventional	.604	.081	.061	.050
HC_0	.453	.190	.209	.193
HC_1	.486	.203	.185	.171
HC_2	.557	.247	.150	.136
HC_3	.667	.309	.110	.100
max(HC_0, Conventional)	.629	.109	.055	.045
max(HC_1, Conventional)	.640	.122	.053	.044
max(HC_2, Conventional)	.679	.166	.047	.039
max(HC_3, Conventional)	.754	.237	.039	.031

Notes: The table reports results from a sampling experiment with 25,000 replications. Columns 1 and 2 shows the mean and standard deviation of estimated *standard errors*, except for the first row in each panel which shows the mean and standard deviation of $\hat{\beta}_1$. The model is as described by (8.1.9), with $\beta_1 = 0$, $r = .1$, $N = 30$, and heteroskedasticity as indicated in the panel headings.

The results with little heteroskedasticity, reported in the second panel, show that conventional standard errors are still too low; this bias is now on the order of 15 precent. HC_0 and HC_1 are also too small, about as before in absolute terms, though they now look worse relative to the conventional standard errors. The HC_2 and HC_3 standard errors are still larger than the conventional standard errors, on average, but empirical rejection rates are higher for these two than for conventional standard errors. This means the robust standard errors are sometimes too small "by accident," an event that happens often enough to inflate rejection rates so that they exceed the conventional rejection rates.

One lesson we can take away from this is that robust standard errors are no panacea. They can be smaller than conventional standard errors for two reasons: the small sample bias we have discussed and their higher sampling variance. We therefore take empirical results where the robust standard errors fall below the conventional standard errors as a red flag. This is very likely due to bias or a chance occurrence that is better discounted. In this spirit, the maximum of the conventional standard error and a robust standard error may be the best measure of precision. This rule of thumb helps on two counts: it truncates low values of the robust estimators, reducing bias, and it reduces variability. Table 8.1.1 shows the empirical rejection rates obtained using max(HC_j, *Conventional*). Rejection rates using this rule of thumb look pretty good in panel B and are considerably better than the rates using robust estimators alone, even with lots of heteroskedasticity, as shown in panel A.[9]

Since there is no gain without pain, there must be some cost to using max(HC_j, *Conventional*). The cost is that the best standard error when there is no heteroskedasticity is the conventional estimate. This is documented in the bottom panel of the table. Use of the maximum inflates standard errors unnecessarily under homoskedasticity, depressing rejection rates. Nevertheless, the table shows that even in this case, rejection

[9]Yang, Hsu, and Zhao (2005) formalize the notion of test procedures based on the maximum of a set of test statistics with differing efficiency and robustness properties.

rates don't go down all that much. We also view an underestimate of precision as being less costly than an overestimate. Underestimating precision, we come away thinking the data are not very informative and that we should try to collect more or improve the research design, while in the latter case we may mistakenly draw important substantive conclusions.

A final comment on this Monte Carlo investigation concerns the small sample size. Labor economists like us are used to working with tens of thousands of observations or more. But sometimes we don't. In a study of the effects of busing on public school students, Angrist and Lang (2004) worked with samples of about 3,000 students grouped in 56 schools. The regressor of interest in this study varied within grade only at the school level, so some of the analysis uses 56 school means. Not surprisingly, therefore, Angrist and Lang (2004) obtained HC_1 standard errors below conventional OLS standard errors when working with school-level data. As a rule, even if you start with the microdata on individuals, when the regressor of interest varies at a higher level of aggregation—a school, state, or some other group or cluster—effective sample sizes are much closer to the number of clusters than to the number of individuals. Inference procedures for clustered data are discussed in detail in the next section.

8.2 Clustering and Serial Correlation in Panels

8.2.1 *Clustering and the Moulton Factor*

Heteroskedasticity rarely leads to dramatic changes in inference. In large samples where bias is not likely to be a problem, we might see standard errors increase by about 25 percent when moving from the conventional to the HC_1 estimator. In contrast, clustering can make all the difference.

The clustering problem can be illustrated using a simple bivariate model estimated in data with a group structure. Suppose we're interested in the bivariate regression,

$$\text{Y}_{ig} = \beta_0 + \beta_1 x_g + e_{ig}, \qquad (8.2.1)$$

where Y_{ig} is the dependent variable for individual i in cluster or group g, with G groups. Importantly, the regressor of interest, x_g, varies only at the group level. For example, data from the STAR experiment analyzed by Krueger (1999) come in the form of Y_{ig}, the test score of student i in class g, and class size, x_g.

Although students were randomly assigned to classes in the STAR experiment, the STAR data are unlikely to be independent across observations. The test scores of students in the same class tend to be correlated because students in the same class share background characteristics and are exposed to the same teacher and classroom environment. It's therefore prudent to assume that, for students i and j in the same class, g,

$$E[e_{ig}e_{jg}] = \rho_e \sigma_e^2 > 0, \qquad (8.2.2)$$

where ρ_e is the residual intraclass correlation coefficient and σ_e^2 is the residual variance.

Correlation within groups is often modeled using an additive random effects model. Specifically, we assume that the residual, e_{ig}, has a group structure,

$$e_{ig} = v_g + \eta_{ig}, \qquad (8.2.3)$$

where v_g is a random component specific to class g and η_{ig} is a mean-zero student-level error component that's left over. We focus here on the correlation problem, so both of these error components are assumed to be homoskedastic. The group-level error component is assumed to capture all within-group correlation, so the η_{ig} are uncorrelated.[10]

When the regressor of interest varies only at the group level, an error structure like (8.2.3) can increase standard errors sharply. This unfortunate fact is not news—Kloek (1981) and

[10] This sort of residual correlation structure is also a consequence of stratified sampling (see, e.g., Wooldridge, 2003). Most of the samples that we work with are close enough to random that we typically worry more about the dependence due to a group structure than clustering due to stratification. Note that there is no GLS estimator for equation 8.2.1 with error structure 8.2.3 because the regressor is fixed within groups. In any case, here as elsewhere we prefer a "fix-the-standard-errors" approach to GLS.

Moulton (1986) both made the point—but it seems fair to say that clustering didn't really become part of the applied econometrics zeitgeist until about 15 years ago.

Given the error structure, (8.2.3), the intraclass correlation coefficient becomes

$$\rho_e = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_\eta^2},$$

where σ_v^2 is the variance of v_g and σ_η^2 is the variance of η_{ig}. A word on terminology: ρ_e is called the *intraclass correlation coefficient* even when the groups of interest are not classrooms.

Let $V_c(\hat{\beta}_1)$ be the conventional OLS variance formula for the regression slope (a diagonal element of Ω_c in the previous section), while $V(\hat{\beta}_1)$ denotes the correct sampling variance given the error structure, (8.2.3). With nonstochastic regressors fixed at the group level and groups of equal size, n, we have

$$\frac{V(\hat{\beta}_1)}{V_c(\hat{\beta}_1)} = 1 + (n-1)\rho_e, \qquad (8.2.4)$$

a formula derived in the appendix to this chapter. We call the square root of this ratio the Moulton factor, after Moulton's (1986) influential study. Equation (8.2.4) tells us how much we overestimate precision by ignoring intraclass correlation. Conventional standard errors become increasingly misleading as n and ρ_e increase. Suppose, for example, that $\rho_e = 1$. In this case, all the errors within a group are the same, so the Y_{ig} values are the same as well. Making a data set larger by copying a smaller one n times generates no new information. The variance $V_c(\hat{\beta}_1)$ should therefore be scaled up from $V_c(\hat{\beta}_1)$ by a factor of n. The Moulton factor increases with group size because with a fixed overall sample size, larger groups mean fewer clusters, in which case there is less independent information in the sample (because the data are independent across clusters but not within).[11]

[11] With nonstochastic regressors and homoscedastic residuals, the Moulton factor is a finite-sample result. Survey statisticians call the Moulton factor the *design effect* because it tells us how much to adjust standard errors in stratified samples for deviations from simple random sampling (Kish, 1965).

Even small intraclass correlation coefficients can generate a big Moulton factor. In Angrist and Lavy (2008), for example, 4,000 students are grouped in 40 schools, so the average n is 100. The regressor of interest is school-level treatment status: all students in treated schools were eligible to receive cash awards for passing their matriculation exams. The intraclass correlation in this study fluctuates around .1. Applying formula (8.2.4), the Moulton factor is over 3, so the standard errors reported by default are only one-third what they should be.

Equation (8.2.4) covers an important special case where the regressors are fixed within groups and group size is constant. The general formula allows the regressor, x_{ig}, to vary at the individual level and for different group sizes, n_g. In this case, the Moulton factor is the square root of

$$\frac{V(\hat{\beta}_1)}{V_c(\hat{\beta}_1)} = 1 + \left[\frac{V(n_g)}{\bar{n}} + \bar{n} - 1\right]\rho_x\rho_e, \qquad (8.2.5)$$

where $\bar{n}$ is the average group size, and ρ_x is the intraclass correlation of x_{ig}:

$$\rho_x = \frac{\sum_g \sum_j \sum_{i \neq j} (x_{ig} - \bar{x})(x_{jg} - \bar{x})}{V(x_{ig}) \sum_g n_g(n_g - 1)}.$$

Note that ρ_x does not impose a variance components structure like (8.2.3); here, ρ_x is a generic measure of the correlation of regressors within groups. The general Moulton formula tells us that clustering has a bigger impact on standard errors with variable group sizes and when ρ_x is large. The impact vanishes when $\rho_x = 0$. In other words, if the x_{ig} values are uncorrelated within groups, the grouped error structure does not matter for standard errors. That's why we worry most about clustering when the regressor of interest is fixed within groups.

We illustrate formula (8.2.5) using the Tennessee STAR example. A regression of kindergartners' percentile score on class size yields an estimate of $-.62$ with a robust (HC_1) standard error of .09. In this case, $\rho_x = 1$ because class size is fixed within classes, while $V(n_g)$ is positive because classes vary in size (in this case, $V(n_g) = 17.1$). The intraclass correlation coefficient for residuals is .31 and the average class size is 19.4. Plugging these numbers into (8.2.5) gives a value of about 7 for $\frac{V(\hat{\beta}_1)}{V_c(\hat{\beta}_1)}$, so that conventional standard errors should be multiplied by a factor of $2.65 = \sqrt{7}$. The corrected standard error is therefore about 0.24.

The Moulton factor works similarly with 2SLS estimates. In particular, we can use (8.2.5), replacing ρ_x with $\rho_{\hat{x}}$, where $\rho_{\hat{x}}$ is the intraclass correlation coefficient of the first-stage fitted values and ρ_e is the intraclass correlation of the second-stage residuals (Shore-Sheppard, 1996). To understand why this works, recall that conventional standard errors for 2SLS are derived from the residual variance of the second-stage equation divided by the variance of the first-stage fitted values. This is the same asymptotic variance formula as for OLS, with first-stage fitted values playing the role of the regressor.

To conclude, we list and compare solutions to the Moulton problem, starting with the parametric approach described above.

1. Parametric: Fix conventional standard errors using (8.2.5). The intraclass correlations ρ_e and ρ_x are easy to compute and supplied as descriptive statistics in some software packages.[12]
2. Cluster standard errors: Liang and Zeger (1986) generalize the White (1980a) robust covariance matrix to allow for clustering as well as heteroskedasticity. The clustered covariance matrix is

$$\hat{\Omega}_{cl} = (X'X)^{-1}\left(\sum_g X_g \hat{\Psi}_g X_g\right)(X'X)^{-1}, \text{ where} \tag{8.2.6}$$

[12]Use Stata's `loneway` command, for example.

$$\hat{\Psi}_g = a\hat{e}_g\hat{e}_g'$$
$$= a\begin{bmatrix} \hat{e}_{1g}^2 & \hat{e}_{1g}\hat{e}_{2g} & \cdots & \hat{e}_{1g}\hat{e}_{n_g g} \\ \hat{e}_{1g}\hat{e}_{2g} & \hat{e}_{2g}^2 & \cdots & \vdots \\ \vdots & \vdots & & \hat{e}_{n_g-1,g}\hat{e}_{n_g g} \\ \hat{e}_{1g}\hat{e}_{n_g g} & \cdots & \hat{e}_{n_g-1,g}\hat{e}_{n_g g} & \hat{e}_{n_g g}^2 \end{bmatrix}.$$

Here, X_g is the matrix of regressors for group g and a is a degrees of freedom adjustment factor similar to that which appears in HC_1. The clustered estimator is consistent as the number of groups gets large given any within-group correlation structure and not just the parametric model in (8.2.3). $\hat{\Omega}_{cl}$ is not consistent with a fixed number of groups, however, even when the group size tends to infinity. Consistency is determined by the law of large numbers, which says that we can rely on sample moments to converge to population moments (section 3.1.3). But here the sums are at the group level and not over individuals. Clustered standard errors are therefore unlikely to be reliable with few clusters, a point we return to below.

3. Use group averages instead of microdata: let $\bar{Y}_g$ be the mean of Y_{ig} in group g. Estimate

$$\bar{Y}_g = \beta_0 + \beta_1 x_g + \bar{e}_g$$

by WLS using the group size as weights. This is equivalent to OLS using micro data but the grouped-equation standard errors reflect the group structure, (8.2.3).[13] Again, the asymptotics here are based on the number of groups and not the group size. Importantly, however, because the group means are close to normally distributed with modest group sizes, we can expect the good finite-sample properties of regression with normal errors to kick in. The standard errors that come out of grouped estimation are therefore likely to be more reliable than clustered standard errors in samples with few clusters.

[13]The grouped residuals are heteroskedastic unless group sizes are equal but this is less important than the fact that the error has a group structure in the microdata.

Grouped-data estimation can be generalized to models with microcovariates using a two-step procedure. Suppose the equation of interest is

$$Y_{ig} = \beta_0 + \beta_1 x_g + \beta_2 w_{ig} + e_{ig}, \qquad (8.2.7)$$

where w_{ig} is a covariate that varies within groups. In step 1, construct the covariate-adjusted group effects, μ_g, by estimating

$$Y_{ig} = \mu_g + \beta_2 w_{ig} + \eta_{ig}.$$

The μ_g, called group effects, are coefficients on a full set of group dummies. The estimated $\hat{\mu}_g$ are group means adjusted for differences in the individual level variable, w_{ig}. Note that, by virtue of (8.2.7) and (8.2.3), $\mu_g = \beta_0 + \beta_1 x_g + \nu_g$. In step 2, therefore, we regress the estimated group effects on group-level variables:

$$\hat{\mu}_g = \beta_0 + \beta_1 x_g + \{\nu_g + (\hat{\mu}_g - \mu_g)\}. \qquad (8.2.8)$$

The efficient GLS estimator for (8.2.8) is WLS, using the reciprocal of the estimated variance of the group-level residual, $\{\nu_g + (\hat{\mu}_g - \mu_g)\}$, as weights. This can be a problem, since the variance of ν_g is not estimated very well with few groups. We might therefore weight by the reciprocal of the variance of the estimated group effects, the group size, or use no weights at all.[14] In an effort to better approximate the relevant finite-sample distribution, Donald and Lang (2007) suggest that inference for grouped equations like (8.2.8) be based on a t-distribution with $G - K$ degrees of freedom.

Note that the grouping approach does not work when x_{ig} varies within groups. Averaging x_{ig} to $\bar{x}_g$ is a version of IV, as we saw in chapter 4. So with micro-variation in the regressor of interest, grouped estimation identifies parameters that differ from the target parameters in a model like (8.2.7).

[14]See Angrist and Lavy (2008) for an example of the latter two weighting schemes.

4. Block bootstrap: In general, bootstrap inference uses the empirical distribution of the data by resampling. But simple random resampling won't do in this case. The trick with clustered data is to preserve the dependence structure in the target population. We can do this by block bootstrapping, that is, drawing blocks of data defined by the groups g. In the Tennessee STAR data, for example, we'd block bootstrap by resampling entire classes instead of individual students.
5. In some cases, you may be able to estimate a GLS or maximum likelihood model based on a version of (8.2.1) combined with a model for the error structure like (8.2.3). This fixes the clustering problem but also changes the estimand unless the CEF is linear, as detailed in section 3.4.1 for LDV models. We therefore prefer other approaches.

Table 8.2.1 compares standard-error fixups in the STAR example. The table reports six estimates: conventional robust standard errors (using HC_1); two versions of corrected standard errors using the Moulton formula (8.2.5), the first using the formula for the intraclass correlation given by Moulton and the second using Stata's estimator from the `loneway` command; clustered standard errors; block-bootstrapped standard errors; and standard errors from weighted estimation at the group level. The coefficient estimate is −.62. In this case, all cluster adjustments deliver similar results, a standard error of about .23. This happy outcome is due in large part to the fact that with 318 classrooms, we have enough clusters for group-level asymptotics to work well. With few clusters, however, things are much dicier, a point we return to at the end of the chapter.

8.2.2 *Serial Correlation in Panels and Difference-in-Difference Models*

Serial correlation—the tendency for one observation to be correlated with those that have gone before—used to be Somebody Else's Problem, specifically, the unfortunate souls who make their living out of time series data (macroeconomists, for

TABLE 8.2.1
Standard errors for class size effects in the STAR data (318 clusters)

Variance Estimator	Std. Err.
Robust (HC_1)	.090
Parametric Moulton correction (using Moulton intraclass correlation)	.222
Parametric Moulton correction (using Stata intraclass correlation)	.230
Clustered	.232
Block bootstrap	.231
Estimation using group means (weighted by class size)	.226

Notes: The table reports standard errors for the estimates from a regression of kindergartners' average percentile scores on class size using the public use data set from Project STAR. The coefficient on class size is −.62. The group level for clustering is the classroom. The number of observations is 5,743. The bootstrap estimate uses 1,000 replications.

example). Applied microeconometricians have therefore long ignored it.[15] But our data often have a time dimension, too, especially in DD models. This fact combined with clustering can have a major impact on statistical inference.

Suppose, as in section 5.2, that we are interested in the effects of a state minimum wage. In this context, the regression version of DD includes additive state and time effects. We therefore get an equation like (5.2.2), repeated below:

$$\text{Y}_{ist} = \gamma_s + \lambda_t + \delta \text{D}_{st} + \varepsilon_{ist}, \tag{8.2.9}$$

[15]The Somebody Else's Problem (SEP) field, first identified as a natural phenomenon in Adams's *Life, the Universe, and Everything*, is, according to Wikipedia, "a generated energy field that affects perception. . . . Entities within the field will be perceived by an outside observer as 'Somebody Else's Problem,' and will therefore be effectively invisible unless the observer is specifically looking for the entity."

As before, Y_{ist} is the outcome for individual i in state s in year t and D_{st} is a dummy variable that indicates treatment states in posttreatment periods.

The error term in (8.2.9) reflects the idiosyncratic variation in potential outcomes across people, states, and time. Some of this variation is likely to be common to individuals in the same state and year, for example, a regional business cycle. We can model this common component by thinking of ε_{ist} as the sum of a state-year shock, v_{st}, and an idiosyncratic individual component, η_{ist}. So we have:

$$\text{Y}_{ist} = \gamma_s + \lambda_t + \delta \text{D}_{st} + v_{st} + \eta_{ist}. \tag{8.2.10}$$

We assume that in repeated draws across states and over time, $E[v_{st}] = 0$, while $E[\eta_{ist}|s,t] = 0$ by definition.

State-year shocks are bad news for DD models. As with the Moulton problem, state- and time-specific random effects generate a clustering problem that affects statistical inference. But that might be the least of our problems in this case. To see why, suppose we have only two periods and two states, as in the Card and Krueger (1994) New Jersey-Pennsylvania study. The empirical DD estimator is

$$\hat{\delta}_{CK} = (\overline{\text{Y}}_{s=NJ,t=Nov} - \overline{\text{Y}}_{s=NJ,t=Feb}) - (\overline{\text{Y}}_{s=PA,t=Nov} - \overline{\text{Y}}_{s=PA,t=Feb}).$$

This estimator is unbiased, since $E[v_{st}] = E[\eta_{ist}] = 0$. On the other hand, assuming we think of probability limits as increasing group size while keeping the choice of states and periods fixed, state-year shocks render $\hat{\delta}_{CK}$ inconsistent:

$$\begin{aligned} &plim\ \hat{\delta}_{CK} \\ &\quad = \delta + \{(v_{s=NJ,t=Nov} - v_{s=NJ,t=Feb}) - (v_{s=PA,t=Nov} - v_{s=PA,t=Feb})\}. \end{aligned}$$

Averaging larger and larger samples within New Jersey and Pennsylvania in a pair of periods does nothing to eliminate the regional shocks specific to a given location and period. With only two states and years, we have no way to distinguish the differences-in-differences generated by a policy change from the difference-in-dfferences due to the fact that, say, the New Jersey economy was holding steady in 1992 while Pennsylvania was experiencing a cyclical downturn. The presence of v_{st} amounts to a failure of the common trends assumption discussed in section 5.2.

The solution to the inconsistency induced by random shocks in differences in differences models is to analyze samples including multiple time periods or many states (or both). For example, Card (1992) uses 51 states to study minimum wage changes, while Card and Krueger (2000) take another look at the New Jersey-Pennsylvania experiment with a longer monthly time series of payroll data. With multiple states or periods, we can hope that the v_{st} average out to zero. As in the first part of this chapter on the Moulton problem, the inference framework in this context relies on asymptotic distribution theory with many groups and not on group size (or, at least, not on group size alone). The most important inference issue then becomes the behavior of v_{st}. In particular, if we are prepared to assume that shocks are independent across states and over time—that is, that they are serially uncorrelated—we are back to the plain vanilla Moulton problem in section 8.2.1, in which case clustering standard errors by state × year should generate valid inferences. But in most cases, the assumption that v_{st} is serially uncorrelated is hard to defend. Almost certainly, for example, regional shocks are highly serially correlated: if things are bad in Pennsylvania in one month, they are likely to be about as bad in the next.

The consequences of serial correlation for clustered panels are highlighted by Bertrand, Duflo, and Mullainathan (2004) and Kézdi (2004). Any research design with a group structure where the group means are correlated can be said to have the serial correlation problem. The upshot of recent research on serial correlation in data with a group structure is that, just as we must adjust our standard errors for the correlation within groups induced by the presence of v_{st}, we must further adjust for serial correlation in the v_{st} themselves. There are a number of ways to do this, not all equally effective in all situations. It seems fair to say that the question of how best to approach the serial correlation problem is currently under study, and a consensus has not yet emerged.

The simplest and most widely applied approach is to pass the clustering buck one level higher. In the state-year example, we can report Liang and Zeger (1986) standard errors clustered by state instead of by state and year (e.g., using Stata `cluster`). This might seem odd at first blush, since the model controls for state effects. The state effect, γ_s, in (8.2.10) removes the state mean of v_{st}, which we denote by $\bar{v}_s$. Nevertheless, $v_{st} - \bar{v}_s$ is probably still serially correlated. Clustering standard errors at the state level takes account of this, since the one-level-up clustered covariance estimator allows for unrestricted residual correlation within clusters, including the time series correlation in $v_{st} - \bar{v}_s$. This is a quick and easy fix.[16] The problem here is that passing the buck up one level reduces the number of clusters. And asymptotic inference supposes we have a large number of clusters because we need many states or periods to estimate the correlation between $v_{st} - \bar{v}_s$ and $v_{st-1} - \bar{v}_s$ reasonably well. A paucity of clusters can lead to biased standard errors and misleading inferences.

8.2.3 *Fewer than 42 Clusters*

Bias from few clusters is a risk in both the Moulton and the serial correlation contexts because in both cases, inference is cluster-based. With few clusters, we tend to underestimate either the serial correlation in a random shock like v_{st} or the intraclass correlation, ρ_e, in the Moulton problem. The relevant dimension for counting clusters in the Moulton problem is the number of groups, G. In a DD scenario where you'd like to cluster on state or some other cross-sectional dimension, the relevant dimension for counting clusters is the number of states or cross-sectional groups. Therefore, following Douglas Adams's dictum that the ultimate answer to life, the universe, and everything is 42, we believe the question is: How many clusters are enough for reliable inference using the standard cluster adjustment derived from (8.2.6)?

If 42 is enough for the standard cluster adjustment to be reliable, and if less is too few, then what should you do when

[16] Arellano (1987) appears to have been the first to suggest higher-level clustering for models with a panel structure.

the cluster count is low? First-best is to get more clusters by collecting more data. But sometimes we're too lazy for that, or the number of groups is naturally fixed, so other ideas are detailed below. It's worth noting at the outset that not all of these ideas are equally well-suited for the Moulton and serial correlation problems.

1. Bias correction of clustered standard errors: Clustered standard errors are biased in small samples because $E(\hat{e}_g\hat{e}_g') \neq E(e_g e_g') = \Psi_g$, just as with the residual covariance matrix in section 8.1. Usually, $E(\hat{e}_g\hat{e}_g')$ is too small. One solution is to inflate residuals in the hopes of reducing bias. Bell and McCaffrey (2002) suggest a procedure (called bias-reduced linearization, or BRL) that adjusts residuals by

$$\hat{\Psi}_g = a\tilde{e}_g\tilde{e}_g'$$
$$\tilde{e}_g = A_g\hat{e}_g$$

where A_g solves

$$A_g'A_g = (I - H_g)^{-1},$$
$$H_g = X_g(X'X)^{-1}X_g',$$

and a is a degrees-of-freedom correction.

This is a version of HC_2 for the clustered case. BRL works for the straight-up Moulton problem with few clusters but for technical reasons cannot be used for the typical DD serial correlation problem.[17]

[17]The matrix A_g is not unique; there are many such decompositions. Bell and McCaffrey (2002) use the symmetric square root of $(I - H_g)^{-1}$, or

$$A_g = R\Lambda^{1/2},$$

where R is the matrix of eigenvectors of $(I - H_g)^{-1}$ and $\Lambda^{1/2}$ is the diagonal matrix of the square roots of the eigenvalues. One problem with the Bell and McCaffrey adjustment is that $(I - H_g)$ may not be of full rank, and hence the inverse may not exist for all designs. This happens, for example, when one of the regressors is a dummy variable that is one for exactly one of the clusters, and zero otherwise. This scenario occurs in the panel DD model discussed by Bertrand et al. (2004), which includes a full set of state dummies and clusters by state.

2. Recognizing that the fundamental unit of observation is a cluster and not an individual unit within clusters, Bell and McCaffrey (2002) and Donald and Lang (2007) suggest that inference be based on a t-distribution with $G - K$ degrees of freedom rather than on the standard normal distribution. For small G, this makes a difference: confidence intervals will be wider, thereby avoiding some mistakes. Cameron, Gelbach, and Miller (2008) report Monte Carlo examples where the combination of a BRL adjustment and the use of t-tables works well.
3. Donald and Lang (2007) argue that estimation using group means works well with small G in the Moulton problem, and even better when inference is based on a t-distribution with $G - K$ degrees of freedom. But, as we discussed in section 8.2.1, for grouped estimation the regressor should be fixed within groups. The level of aggregation is the level at which you'd like to cluster, such as schools in Angrist and Lavy (2008). For serial correlation, this is the state, but state averages cannot be used to estimate a model with a full set of state effects. Also, since treatment status varies within states, averaging up to the state level averages the regressor of interest as well, changing the rules of the game in a way we may not like (the estimator becomes IV using group dummies as instruments). The group means approach is therefore out of bounds for the serial correlation problem. Note also that if the grouped residuals are heteroskedastic, and you therefore use robust standard errors, you may have to worry about bias of the form discussed in section 8.1. In some cases, heteroskedasticity in the grouped residuals can be fixed by weighting by the group size. But weighting changes the estimand when the CEF is nonlinear, so the case for weighting is not open and shut (Angrist and Lavy, 1999, chose not to weight school-level averages because the variation in their study comes mostly from small schools). Weighted or not, a conservative approach when working with group-level averages is to use our rule of thumb from section 8.1: take the larger of robust and conventional standard errors as your measure of precision.

4. Cameron, Gelbach, and Miller (2008) report that some forms of a block bootstrap work well with small numbers of groups, and that the block bootstrap typically outperforms Stata-clustered standard errors. This appears to be true both for the Moulton and serial correlation problems. But Cameron, Gelbach, and Miller (2008) focus on rejection rates using (pivotal) test statistics, while we like to see standard errors.
5. Parametric corrections: For the Moulton problem, this amounts to use of the Moulton factor. With serial correlation, this means correcting your standard errors for first-order serial correlation at the group level. Based on our sampling experiments with the Moulton problem and a reading of the literature, parametric approaches may work well, and better than the nonparametric cluster estimator (8.2.6), especially if the parametric model is not too far off (see, e.g., Hansen, 2007a, which also proposes a bias correction for estimates of serial correlation parameters). Unfortunately, however, beyond the greenhouse world of controlled Monte Carlo studies, we're unlikely to know whether parametric assumptions are a good fit.

Alas, the bottom line here is not entirely clear, nor is the more basic question of when few clusters are fatal for inference. The severity of the resulting bias seems to depend on the nature of your problem, in particular whether you confront straight-up Moulton or serial correlation issues. Aggregation to the group level as in Donald and Lang (2007) seems to work well in the Moulton case as long as the regressor of interest is fixed within groups and there is not too much underlying heteroskedasticity. At a minimum, you'd like to show that your conclusions are consistent with the inferences that arise from an analysis of group averages, since this is a conservative and transparent approach. Angrist and Lavy (2008) use BRL standard errors to adjust for clustering at the school level but validate this approach by showing that key results come out the same using covariate-adjusted group averages.

As far as serial correlation goes, most of the evidence suggests that when you are lucky enough to do research on U.S. states, giving 51 clusters, you are on reasonably safe ground with a naive application of Stata's `cluster` command at the state level. But you might have to study Canada, which offers only 10 clusters in the form of provinces, well below 42. Hansen (2007b) finds that Liang and Zeger (1986) (Stata-clustered) standard errors are reasonably good at correcting for serial correlation in panels, even in the Canadian scenario. Hansen also recommends use of a t-distribution with $G-K$ degrees of freedom for critical values.

Clustering problems have forced applied microeconometricians to eat a little humble pie. Proud of working with large microdata sets, we like to sneer at macroeconomists toying with small time series samples. But he who laughs last laughs best: if the regressor of interest varies only at a coarse group level, such as over time or across states or countries, then it's the macroeconomists who have had the most realistic mode of inference all along.

8.3 Appendix: Derivation of the Simple Moulton Factor

Write

$$y_g = \begin{bmatrix} Y_{1g} \\ Y_{2g} \\ \vdots \\ Y_{n_g g} \end{bmatrix} \qquad e_g = \begin{bmatrix} e_{1g} \\ e_{2g} \\ \vdots \\ e_{n_g g} \end{bmatrix}$$

and

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_G \end{bmatrix} \qquad x = \begin{bmatrix} \iota_1 x_1 \\ \iota_2 x_2 \\ \vdots \\ \iota_G x_G \end{bmatrix} \qquad e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_G \end{bmatrix},$$

where ι_g is a column vector of n_g ones and G is the number of groups. Note that

$$E(ee') = \Psi = \begin{bmatrix} \Psi_1 & 0 & \cdots & 0 \\ 0 & \Psi_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \Psi_G \end{bmatrix}$$

$$\Psi_g = \sigma_e^2 \begin{bmatrix} 1 & \rho_e & \cdots & \rho_e \\ \rho_e & 1 & & \vdots \\ \vdots & & \ddots & \rho_e \\ \rho & \cdots & \rho_e & 1 \end{bmatrix} = \sigma_e^2 \left[(1-\rho_e)I + \rho_e \iota_g \iota_g' \right],$$

where $\rho_e = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_\eta^2}$.

Now

$$X'X = \sum_g n_g x_g x_g'$$
$$X'\Psi X = \sum_g x_g \iota_g' \Psi_g \iota_g x_g'.$$

But

$$x_g \iota_g' \Psi_g \iota_g x_g' = \sigma_e^2 x_g \iota_g' \begin{bmatrix} 1+(n_g-1)\rho_e \\ 1+(n_g-1)\rho_e \\ \cdots \\ 1+(n_g-1)\rho_e \end{bmatrix} x_g'$$
$$= \sigma_e^2 n_g \left[1+(n_g-1)\rho_e\right] x_g x_g'.$$

Let $\tau_g = 1+(n_g-1)\rho_e$, so we get

$$x_g \iota_g' \Psi_g \iota_g x_g' = \sigma_e^2 n_g \tau_g x_g x_g'$$
$$X'\Psi X = \sigma_e^2 \sum_g n_g \tau_g x_g x_g'.$$

With this in hand, we can write

$$V(\hat\beta) = (X'X)^{-1} X'\Psi X (X'X)^{-1}$$
$$= \sigma_e^2 \Big(\sum_g n_g x_g x_g'\Big)^{-1} \sum_g n_g \tau_g x_g x_g' \Big(\sum_g n_g x_g x_g'\Big)^{-1}.$$

We want to compare this with the standard OLS covariance estimator

$$V_c(\hat\beta) = \sigma_e^2 \Big(\sum_g n_g x_g x_g'\Big)^{-1}.$$

If the group sizes are equal, $n_g = n$ and $\tau_g = \tau = 1+(n-1)\rho_e$, so that

$$V(\hat\beta) = \sigma_e^2 \tau \Big(\sum_g n x_g x_g'\Big)^{-1} \sum_g n x_g x_g' \Big(\sum_g n x_g x_g'\Big)^{-1}$$
$$= \sigma_e^2 \tau \Big(\sum_g n x_g x_g'\Big)^{-1}$$
$$= \tau V_c(\hat\beta),$$

which implies (8.2.4).