Causal Relations via Econometrics

Asad Zaman®

International Islamic University of Islamabad

ABSTRACT

Applied econometric work takes a superficial approach to causality. Understanding economic affairs, making good policy decisions, and progress in the economic discipline depend on our ability to infer causal relations from data. We review the dominant approaches to causality in econometrics, and suggest why they fail to give good results. We feel the problem cannot be solved by traditional tools, and requires some out-of-the-box thinking. Potentially promising approaches to solutions are discussed.

Key words: *Causality, Regression, Exogeneity, Hendry Methodology, Natural Experiments*
JEL Classifications: C, C5, C59

1. INTRODUCTION

The Cowles Commission had a clear approach to causality in econometric models. Causal chains were to be specified in advance of modeling, on the basis of theoretical considerations. The econometrician specified the exogenous and endogenous variables, and put in zero restrictions when theory indicated no role for a particular variable in a structural equation. This approach did not succeed for several reasons. General equilibrium suggests that everything causes everything else, and so theory does not provide an adequate guide to model specification. Thus, zero restrictions and exogeneity assumptions were made on pragmatic grounds. However, substantial conflicts and differences of opinions arose on these issues, which could not be resolved either empirically or theoretically. Large forecast errors in econometric models following the oil crisis in the 70's also cast a cloud of suspicion on these conventional methodologies. Keuzenkamp (2000) provides a history, further references, and a critical evaluation of a number of new methodologies, which have since been developed. Our focus will be on the use of regression models to establish causal relationships.

Most current econometric texts either make no mention of causality, or else contain a brief and superficial discussion. Establishing causality is often a central concern in many papers in applied econometrics. Differentiating between causes and effects of growth, poverty

---

® Asad Zaman, International Institute of Islamic Economic, International Islamic University of Islamabad, (email: asadzaman@alum.mit.edu), Tel. +9251 9257939, Fax: + 9251 9258019.
I would like to dedicate this paper to the memory of David Freedman, who made detailed comments on several rounds of earlier drafts of this paper, resulting in substantial improvements in clarity and conceptualization. The impress of his ideas on this paper is unmistakable, and I would have listed him as the co-author or the principal author but for the fear that the paper may not meet his standards of excellence. We never actually met, but had extensive correspondence via email over the course of several years, which was tremendously beneficial to me in improving my understanding of several controversial issues in statistics. Occasional personal notes revealed his deep concern for the larger problems facing human beings in these 'dark' times. I deeply regret his loss, both on a personal and on a professional level.
An earlier draft of this paper was presented at the South Asian and Far Eastern Meetings of the Econometrics Society (FEMES), July 2008, Singapore. I am grateful to participants for their comments.

reduction, inflation, etc. is of crucial importance to crafting suitable policy and developing an understanding of the world we live in. Due to lack of appropriate training, many published articles display very poor understanding of the evidence required to support causal claims. Freedman (2005) discusses numerous articles which use regression analysis, and shows how causality claims central to these articles reduce to claims based on observed correlations in non-experimental data. Even though the ideas that ‘correlation does not prove causation’ and that ‘Granger-Causality is not equivalent to causality’ are well-known and oft-repeated, authors nonetheless continue to rely on these tools to establish and validate causality claims. Freedman (1991) has argued that discovering causal relations requires more hard work than mere statistical analysis. We will argue that these problems are compounded in econometric analysis. It is essential to learn about causal chains. Economic theory gives us some, imperfect, guidance on this matter. Where the causal chains are clear, regressions are useful in assessing the quantitative strength of the causal effect. A discovery of surprising or unexpectedly strong correlations can lead us to interesting hypotheses about causal mechanisms. While regressions can be useful tools as part of exploratory causal analysis, they are not adequate for confirmation of causal mechanisms, for reasons to be discussed. Establishing causality will usually requiring going outside the range of conventional econometric techniques. Regression analysis may point the way, and may serve as part of the evidence for a causal mechanism, but establishing causality will require more broadly based evidence from different types of sources, and more attention to the structure of the real world mechanisms, which generate the data.

2. CARICATURE OF A TYPICAL ECONOMETRIC ARGUMENT

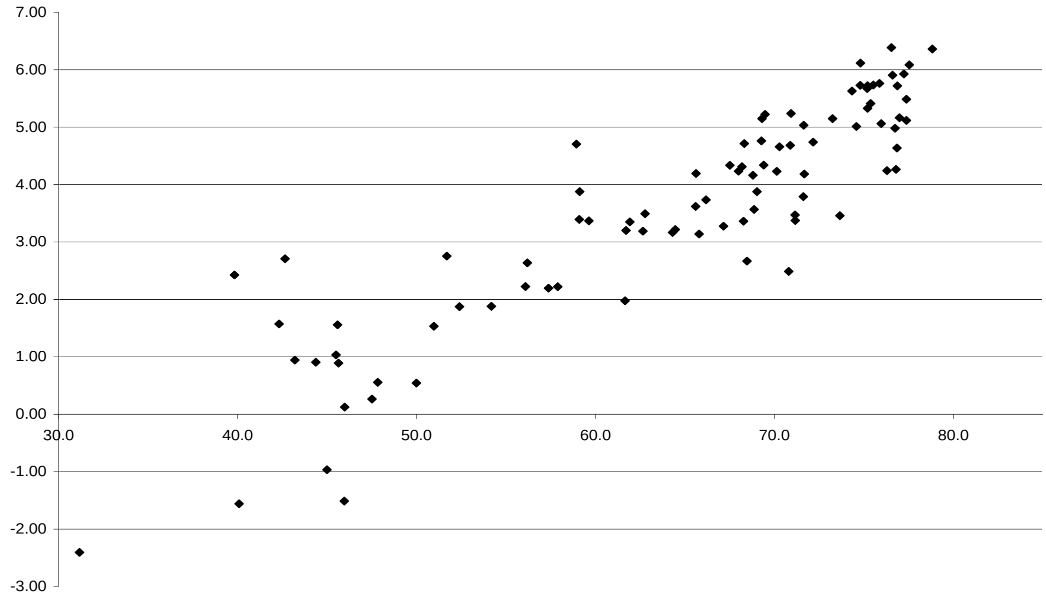
Figure 2.1 below plots the variables for all countries for which World Bank provides information on both variables in 1990. The graph shows the strong positive relation between Life Expectancy (*LE*) and log of number of newspapers published per 1000 people (*LN*). The regression of *LE* on *LN* has $R^2 = 0.81$, and *SER*=5.2 (standard errors are given in parentheses)

$$\underset{}{LE} = \underset{(1.2)}{45.0} + \underset{(0.3)}{5.48}\, LN + \varepsilon$$

LN is highly significant. The picture itself shows a clear and strong relationship, and the formal statistics does not really add much to the information displayed in the plot. The issue is: how to interpret this relationship.

Suppose someone were to argue that the significant *t*-statistic on *LN* proves that reading newspapers leads to longer lives, and argues that developing countries should focus on publishing more newspapers as a means for improving life expectancies. Surely, such an argument would be greeted with laughter. If this argument was published in a journal, we would be concerned about the sloppy standards of refereeing and editorship. Nonetheless, arguments equivalent in logical structure to this one are routinely published in respectable journals. For example, on the basis of essentially equivalent data (actually, weaker data) Lynn and VanHanen (2002) argue at book length that IQ causes GNP growth. Volken (2003) has written a rebuttal using careful and detailed statistical analysis – the seriousness with which such arguments are taken is another indication of fundamental problems with our understanding of causation. In “Why we learn nothing from regressing economic growth on government policies,” Rodrik (2005) has pointed out why causal claims regarding effects of policy on the basis of cross country regressions are not supported by the correlations, and cited several papers making such invalid claims. Rodrik and Rodrigues (2001) analyze three influential and highly cited papers, which argue that openness in trade leads to higher growth rates, and point out errors in the analysis, which invalidate the causal conclusion. Similarly, Friedman (2005) has also cited elementary mistakes about causation in influential papers published in leading journals in social sciences. These sources provide substantial evidence that cavalier treatment of causality leads to serious errors in numerous articles, even by the best authors publishing in the leading journals.

Figure 2.1 Log of number of newspapers published (*LN*) versus Life Expectancy (*LE*)



Some argue that multiple regression can solve the problem of significance of *LN* (log of Newspapers per 1000) in the regression above. The problem arises only because *LN* proxies for other important variables. This means that including the relevant variables will eliminate the significance of others. We tried this by including all health relevant variables for which data was available in sufficient quantities in the World Bank data base. This led to the following regression, with $R^2 = 0.74$, and $SER = 3.8$:

$$\underset{(8.9)}{LE = 57.2} + \underset{(2.7)}{6.0\,LN} - \underset{(2.8)}{0.7\,LHB} + \underset{(0.06)}{0.08\,ImpSan} - \underset{(0.16)}{0.04\,ImpWat} + \underset{(2.6)}{4.8\,LPhys} + \varepsilon$$

Here, *LHB* and *LPhys* are logs of Hospital Beds and Physicians per 1000 population respectively, while *ImpSan* and *ImpWat* measure improvements in sanitation and water supply. *LN* remains the only variable, which is significant at 95% level. Interestingly, its coefficient also remains stable despite the addition of several variables. All variables other than LN are insignificant. It appears likely that there is a complex set of factors related to the process of development, which all affect Life Expectancy. Newspapers per 1000 provides a better measure of this complex than other single or groups of variables and hence appears significant. No amount of regression tricks will enable us to disentangle the causal complex, without studying the actual structure of real world variables, which directly impact on life expectancy – this is the process of expending "shoe leather" which Friedman (1991) has described as essential to the discovery of causality. We will later provide additional arguments why the strategy of variable addition is likely to fail as a tool for discovering causal relations.

Typical authors and textbooks remain confused about these issues despite the fact that several authors have clearly differentiated between correlation and causation. Problems arising from confusing the two have been explicated in many papers ranging from the deep and sophisticated theoretical exposition of Engle et al. (1984) to the lucid textbook clarity of Freedman et al. (2007), and the practical and empirical Rodrik (2005). Freedman (2005) discusses the very first paper on multiple regression by Yule, published in 1899. Yule regressed percent change in Pauperism on Percent change in Outrelief ratio (outside vs inside the poorhouse–the policy debate was on the “new poor law” which made in-relief mandatory). The unit of analysis was the “union,” a clump of about 25 parishes. These were the bodies that administered relief. Multiple regression was used to control for several possible confounders. Yule came to the conclusion that increases in in-relief provisions lead to increases in the number who declare themselves poor. However, in a footnote, Yule writes that “strictly speaking, for ‘due to’ read associated with.” If this footnote is taken seriously, then the paper says nothing of relevance to the real world, and the policy conclusion drawn is meaningless. Countless instances of similar strategies can be pointed out in currently published papers. The text of the analysis makes the usual academic disclaimers and points out difficulties with data, and non-equivalence of Granger causality and our common understanding of causality. All these caveats are forgotten in the conclusions section which makes causal claims on the basis of regressions and policy recommendations on the basis of these causality claims.

The point of this discussion is that there are certain aspects of observational data, which simply cannot be assessed or analyzed by any econometric technique, whether crude or sophisticated. This is even at the simple level of assessing whether a correlation is genuine or spurious. Causal mechanisms require an even deeper knowledge of structure than simple correlations. Recently, Banerjee and Duflo (2002, 2004, 2007) and others have systematically resorted to large scale randomized experiments for a number of microeconomic propositions, with good results. Similarly, behavioral and experimental economists have started generating and gathering data on actual behavior in controlled situations. These are promising developments since both utilize experimental data, which is vastly superior to observational data in helping to reveal causal mechanisms.

3. EMPIRICAL FAILURE OF REGRESSION MODELS

Starting from the inauspicious beginning by Yule[1], more than a hundred years of regressions have failed to yield a single demonstrable and solid successful discovery of a causal relationship. Magnus (1999) writes about econometric tests of economic theories that “we invited readers to name a published paper that … significantly changed the way economists think about some economic proposition.” Total lack of response to this challenge suggests that it is time to step back and rethink strategy. In a talk on the 100th anniversary of the first published regression by Yule, Freedman (1997:113) writes: “For nearly a century, investigators in the social sciences have used regression models to deduce cause-and-effect relationships from patterns of association. … . In my view, this enterprise has not been successful.” For nearly every posited causal mechanism in economic theory, there are econometric papers on both sides of the issue. A widely believed causal claim is that growth of the money stock causes inflation. However, in a careful study based on the most recent methodological advances, Hendry and Ericsson (1991) dispute this claim of Friedman and argue that the causality runs in the other direction. The core of the rejoinder by Friedman is simply that complex econometric analyses often fail in the real world; in support of this, he

[1] It seems likely that Yule reached the wrong conclusion. Migrations of the poor from nearby parishes to those providing better relief would account for the same effect. Given the deliberately harsh and degrading conditions of the poorhouses, it seems unlikely that significant numbers chose to declare themselves as paupers if they could afford not to do so.

cites personal experience rather than analytical or theoretical arguments. The consumption function introduced by Keynes is at the heart of macroeconomics, and has been intensively studied. Despite its central importance and extensive research, there is a bewildering variety of variables with claims to be causes of consumption, all supported by econometric analysis. Furthermore, the best available models routinely fail; Thomas (1993:284) writes that “Perhaps the most worrying aspect of empirical work on aggregate consumption is the regularity with which apparently established equations break down when faced with new data. This has happened repeatedly in the UK since the 1970s. … the reader may be forgiven for wondering whether econometricians will ever reach the stage where even in the short run their consumption equations survive confrontation with new data.”

Why is there lack of clarity about an issue of such fundamental importance, which is central to our tasks as econometricians and economists? Hoover (2006) discusses the history of a tension and paradox originating with Hume, which continues to this day in economic practice. One the one hand, Hume recognizes the central importance of causality to the conduct of economic policy. On the other hand, Hume notes that only timing and correlations can be observed, and genuine causal relations are unobservable. This tension is reflected in econometrics in the frequency with which causation is discussed in the policy recommendations and implications section, and correlations and timings equated to causality in the econometric analysis.

An additional factor in the cavalier treatment of causality in econometrics is identified by Blaug (1998): “Economics as taught in graduate schools has become increasingly preoccupied with formal technique to the exclusion of studying real-world problems and issues.” Inertia and momentum of existing methodologies, which have acquired respectability and mechanisms in place for promotions and publication, are certainly to blame. For example, Hey (1997) summed up his experience of ten years of editorship of the Economic Journal by noting the overwhelming predominance of formal mathematical models of economic problems: “Many of the submissions do not appear to be written in order to further economic knowledge. … few economists ask themselves what are the crucial economic problems facing society. If they did so, they might well produce more relevant material.” Articles focused on econometric techniques alone, with data and real-world applications serving merely as framework and window dressing for demonstration of mathematics and/or clever new statistical techniques are easier to write, more easily publishable, and more prestigious, then marshalling of evidence from a variety of clues pieced together via less formal arguments often required for genuine causal analysis. Discussing several cases of successful discovery of causal relations, Freedman (2008) has emphasized the role of qualitative and informal insights in the process.

Discovering causal laws is difficult, and involves substantial effort. It is nonetheless possible, and numerous successes are documented by Freedman et al. (2007), including the effectiveness of Salk’s polio vaccine. The relation between smoking and cancer is famous for having been established purely on a statistical basis. To illustrate the subtle and complex issues, which must be resolved to distinguish between correlation and causation, note that the suggestion that a genetic factor might dispose one towards both smoking and cancer was disproven by looking at cancer rates among identical twins with discordant smoking habits. Along with numerous successes, failures and errors in discovering causality are also common. Freedman et al. (2007) discuss many cases in which observational studies led to wrong conclusions, sometimes with disastrous consequences. In the next section, we argue that while statisticians can sometimes arrive at the truth, special features of econometric methodology make this outcome unlikely for econometricians.

4. STRATEGIES OF SURRENDER

Hume pointed out that only correlations and timing are observable, while causality is inherently unobservable. This has led some to argue that we can get by without assessing causal relations; worse, some have argued that, because it is fundamentally unobservable, it is a pre-scientific notion and should be discarded. Prominent among these what one of the founders of statistics, Pearson (1911) who writes that “Beyond such discarded fundamental as ‘matter’ and ‘force’ lies still another fetish … namely, the category of cause and effect.” Hoover (2004) writes that ‘causal language in economics virtually collapsed between 1950 and about 1990’. Lack of emphasis on causality is easily demonstrated by some quotes collected in Pearl (2000:341):

- The Encylopedia of Statistical Science devotes twelve pages to correlation, but only two to causation, and spends one of these pages demonstrated that the two are not equivalent.

- Philip Dawid, the editor of Biometrika, admits that “Causal inference is on of the most important, most subtle, and *most neglected* of all the problems of statistics.” [emphasis mine]

- Terry Speed, a former president of the Biometric Society, declares: “Considerations of causality should be treated as they always have been in statistics: preferably not at all but, if necessary, with very great care.

- Cox and Wermuth (1996) in their recent text *Multivariate Dependencies: Models, Analysis and Interpretation*, write that “We did not in this book use the words *causal* or *causality* …”

We illustrate why we cannot get by without an understanding of causality by a simple example. Barro (1997) discovered that education affects development with about a 10 year lag, while other variables he considered were not significant. This implies that the single most significant component of a development strategy is investment on education. While this position has substantial intuitive appeal, we wish to consider whether the statistical evidence supports a causal link between education and development. The evidence is logically equivalent to the evidence supporting the link between newspapers and longevity. To decide on whether or not we should invest heavily in education as a means to promote development, it is crucial to distinguish between correlation and causation. If we consider this to be mere correlation and fail to invest in education, we could be guilty of losing a tremendous opportunity for improving lives of large numbers of people. Mistaking a correlation for a causal relation would lead to the opposite mistake, akin to investing heavily in newspaper production as a means for prolonging life. We cannot afford to be agnostic about the difference between correlations and causation.

In wake of disenchantment with econometrics following widespread forecast failures in the 70’s, schools of thought which place even greater emphasis on theory and even less on data have emerged. The real business cycle (RBC) school of thought in macroeconomics uses data to “calibrate” theoretical models – that is numerical magnitudes not specified in the theory are measured by using elementary descriptive statistics of the data; see Faust and Whiteman (1997). The question of using data to assess theory and revise postulated causal mechanisms does not arise in this approach.

The Lucas critique suggested that failure of macroeconomic models results from insufficient attention to theory. To remedy this, one should use the data to estimate unobservables which economic theory posits (termed 'deep parameters'), instead of directly estimating relationships among observables. Like the Cowles Commission approach, all of these strategies give up on the possibility of learning about causal mechanisms from the data. To a lesser extent, they also give up on the possibility of learning about the structure, functional forms, and nature of stochastic relations among the variables from the data. To be fair, gross conflicts between observations and theoretical predictions can take place, and have led to revisions of theory, and even of causal chains. Similarly, modifications of functional forms and of specifications for errors can and do take place routinely in course of estimation of regression models. However, all of these activities fall outside the purview of econometric theory (and are done in the basement, out of sight of the high priests who condemn such activities, in the metaphor of Leamer, 1983).

With a few exceptions to be discussed later, econometric theory does not formally allow for learning from data about many crucial aspects of the real world. If we had excellent economic theory, which provided a reliable guide to causal and structural mechanisms, then we could live with this situation. Unfortunately, we have compelling reasons to believe otherwise. Current macroeconomic textbooks discuss the diverse opinions of numerous schools of thought; see for example Seven Schools of Macroeconomic Thought by Phelps (1991). Papers presented at the AEA Session in 1997 entitled "Is There a Core of Macroeconomics That We Should All Believe?" (see Bolch, 1998) highlighted the conflicts over nearly all fundamental macroeconomic propositions. In face of such severe conflicts among theorists, it is clear that theory does not provide a reliable guide. In such a situation, it is necessary to devise some mechanisms to allow us to learn about the world from data.

The substantial and persistent conflicts among theorists have led another school of thought, headed by Sims (1980), to the opposite extreme. This school recommends fitting a general VAR model to time series data, which treats all variables together as nameless pieces of data, and gives no role to theory. In my view, differentiating causality from correlation requires some knowledge of structure, and hence the real world quantities measured by the variables under study. A purely statistical analysis cannot discover causal effects and hence amounts to a strategy of surrender. Nonetheless, such analysis can discover patterns of correlation among the data as well as timing. By far the currently most popular approach to causality in econometrics is Granger Causality, which substitutes timing and correlations for causality. Since such an approach has the potential for learning about causality from the data, we defer a discussion to the next section.

5. METHODS FOR DISCOVERING CAUSALITY

Theory does not provide adequate guidance, and we cannot afford to be ignorant, so we must use observations and data to learn about causal mechanisms. Aldrich (1989) has argued that progress of science is directly related to discovery of causal mechanisms. Hoover (2006) has reviewed all extant approaches to causality currently in use in economics and econometrics, and also provided a useful categorization and classification. Below we follow an alternative scheme, and classify approaches according to the evidence used to justify causal inference. This leads to four types of approaches, discussed separately below.

Conformity to Theoretical Specifications: We have discussed different types of approaches based on a priori specifications of causal schemes. Here a model, which fits the data and conforms to the a priori specification, is taken as evidence of support for the a priori causal specifications. Because of the diverse and flexible classes of models available to the econometrician, it is very easy to produce models, which fit the data, even for very bad theories. This is discussed in greater detail in the next section. The literature is full of conflicting theories all of which are able to provide evidence of conformity. This is why such methods have failed to uncover any causal mechanisms or resolve any theoretical controversies.

Stability and Robustness to Specifications: It is sometimes suggested that ‘robustness’ is the key to causal relationships (for example Leamer, 1983, as well as several other authors). If a variable stands out as significant in a lot of different specifications, then this signals a causal relationship. In small data sets, robustness may come about purely by accident. For example, an economic theory suggests that annual changes in consumption are purely random – the consumer plans his lifetime consumption based on his evaluation of his lifetime income. Random shocks to his income can cause random changes in his plans, but these are not predictable from past data; see Thomas (1991:274-278).

Letting *DCHN* and *DAUS* be the annual change in consumption in China and Australia respectively, a regression yields the following results; (R^2 = 0.30, *SER* = 6.5 E+10)

$$\underset{}{DCHN} = \underset{(2.65\text{E}+10)}{1.39\text{E}+10} + \underset{(1.97)}{7.11}\, DAUS + \varepsilon$$

	Coefficients	*Standard Error*	*t-stat*
Intercept	5.87E+10	6.60E+10	0.89
DZA	-2.90	2.81	-1.03
ARG	-0.43	1.61	-0.27
AUS	8.44	3.12	2.71
AUT	-0.59	10.13	-0.06
BGD	0.30	0.38	0.79
BEL	-7.42	9.96	-0.75
BEN	0.95	0.58	1.65
BOL	-21.50	54.31	-0.40
BFA	0.30	0.42	0.72
CMR	-0.01	0.06	-0.11
CAN	-1.23	2.29	-0.54
TCD	-0.46	0.36	-1.27
CHL	0.00	0.01	0.09
COL	-0.02	0.02	-1.17
ZAR	-1.82	1.34	-1.36
CIV	0.08	0.08	0.97

Table 5.1 Regression of Change in Annual Consumption in China (*DCHN*) on Annual Consumption for 16 Countries (WDI three letter country codes).

This regression uses WDI Online data on Final consumption expenditure (constant LCU) for China (CHN) and Australia (AUS) from 1970 to 2003, and suggests that changes in Australian consumption significantly affects the Chinese, rejecting the economic theory. Differencing eliminates stochastic trends, so this phenomenon is not related to the classic “spurious regressions” produced by use of integrated time series. Remarkably, the coefficient remains numerically stable and statistically significant even after we add many additional countries to this same regression. We regress *DCHN* on the consumption of the countries listed (three letter codes as per WDI tables). The estimated coefficients and associated *t*-statistics are listed in Table 5.1; only Australia (AUS) is significant, and the coefficient is stable (close to what it is in the above regression without the other countries).

Backed by some plausible sounding theory such as imitation of Australian consumer behavior by the Chinese, and impressive statistical names like “Extreme Bounds Analysis,” we should be able to convince the unsuspecting victim that the robust and reliable relationship cannot be due to chance. The ease with which this and similar examples can be produced show that we cannot trust purely statistical analyses on observational data sets. Hendry (2000) given many other examples based on the idea that ‘integrated’ series lead to spurious regressions. Our example above uses differenced series and shows that spurious regressions are not just confined to integrated series.

Timing of Correlations: Granger causality is the most popular explicit approach to eliciting causal mechanisms from the data. This relies on evidence about timing of correlations. Asghar (2007) has shown that Granger causality is very sensitive to minor and apparently insignificant details of the testing procedure. Causal chains can reverse directions for small changes in specifications, time period of data, variable transforms, tests used for model selection and lag length selection, etc. Since econometricians routinely experiment with such changes, this accounts for the large numbers of conflicting claims about Granger causality, which can be found in the literature. A simple reason why timing evidence cannot be trusted can be given as follows. Suppose that there is an underlying variable *M*, which measures structural changes, associated with modernization, which reflect in several dimensions in the economy. This variable is not directly observed nor easily measurable. Suppose that *M* leads to increases in *LN* (log of newspapers per 1000) in the short run and increases in *LE* (life expectancy) in the long run. Then the data will show a correlation between *LN* and *LE* at a later date. This correlation will also be robust to addition of other variables, because it is based on real structure. Nonetheless, it cannot be relied upon to reveal causation. Nor can we expect that multiple regression techniques, or sophisticated treatments via instrumental variables, will reveal the problem. Deeper and more sophisticated arguments by Hoover (2001) show why we cannot rely on Granger causality to learn about causal mechanisms in the context of macroeconomic models.

Patterns of Correlations: No causal relation between *X* and *Y* implies stochastic independence and zero correlation. Similarly, all causal patterns between a collection of variables have implications for patterns of correlations which can exist between these variables. Pearl (2000) and associates have worked out methods to infer causal chains from the observed patterns of correlation in the data. A handful of papers have applied such methods in econometrics. See Hoover (2006) for further details and references. So far there are no recorded successes for these methods in econometrics. Some reasons for pessimism are given in Freedman (2004). Freedman (2005, Chapter 5) also provides a critique of path models as a means for discovering causality.

Natural Experiments: Using knowledge of history outside the purely statistical, one can isolate episodes and events where changes are clearly not caused by variables under study. This ability to isolate an uncaused event leads to the possibility of studying causes by examining relations, which remain stable through periods of change. This possibility has been exploited by a few researchers (for example, Angrist and Krueger, 2001; Hoover, 2001 and Asghar, 2007) to asses causality in various econometric models. Hendry's methodology employs a variant of this technique, and is discussed in section 7 below. This approach appears promising as it inputs additional genuine causal information outside the purely statistical ones into the procedure of econometric inference. Statisticians have been successful in using natural experiments to isolate cause and effect from observational data. The temptation to overfit, using complex models and equally complex error processes, tends to make it difficult for econometricians to learn from the data. If this tendency could be regulated, this approach should prove fruitful.

6. THE UNDERSTUDIED PROBLEM OF OVERFITTING

The extensive variety of models available to the econometrician as vehicles to expressing a theory in concrete forms leads to overfitting data in many ways, and overfitting leads to lack of validity for regression models. It is well known that changing the initial regression model in response to any aspect of the fit leads to lack of validity of the diagnostic statistics for the second model fitted. For example, if we select the best among 20 regressions, one of these will be significant at 95% level when all of the regressions are invalid. Jensen (2000) writes that

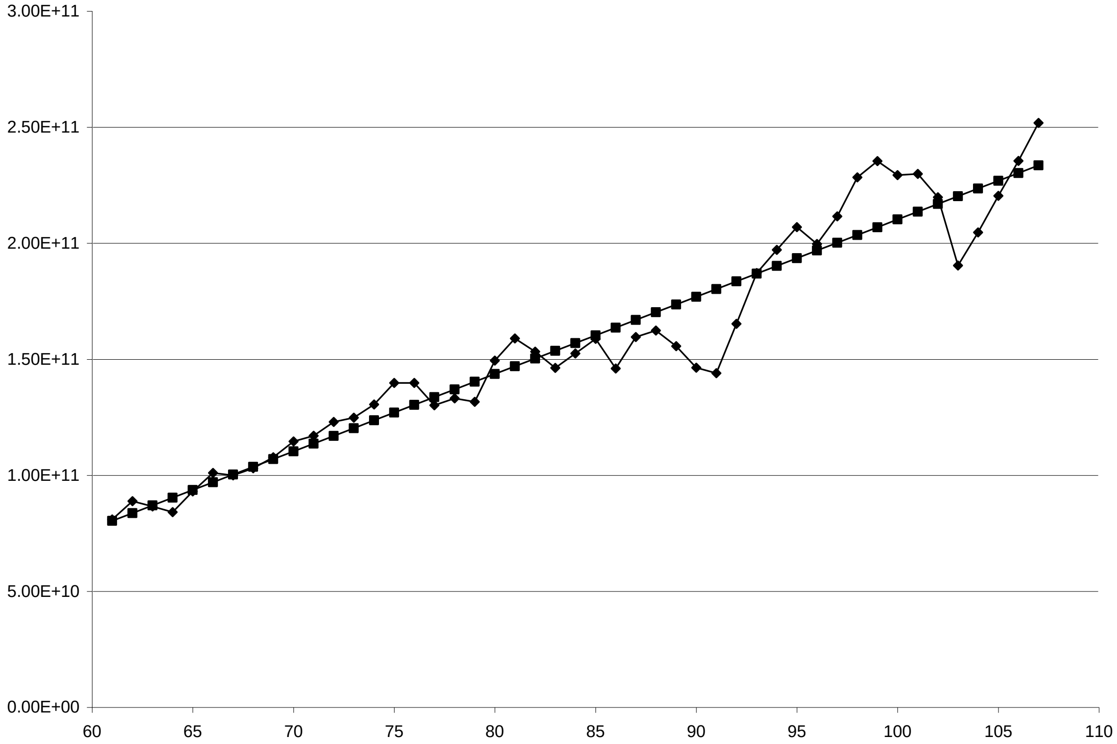
> "However, this "dark side" of data mining is still largely unknown to some practitioners, and problems such as overfitting and overestimation of accuracy still arise in knowledge discovery applications with surprising regularity. In addition, the statistical effects of search can be quite subtle, and they can trip up even experienced researchers and practitioners.

Despite the fact that model selection almost always takes place in practice and is known to affect the validity of regression statistics of the final model presented, it is nearly universally ignored in applications. Contributions of a few authors (e.g. Ashley, 1999) who have attempted to provide methods for adjusting statistics to account for the search process, following the seminal work of Leamer (1983), have been ignored in the literature.

If one has a sufficiently rich class of models to provide a perfect fit to any data set, then this class will almost certainly overfit the data, and therefore be almost useless for discovery of real world structures. Our contention is that the toolbox of the econometrician is too rich for comfort. The full data set for Final Consumption in constant LCU (local currency units) for Argentina, taken from WDI online, is pictured in Figure 6.2 below.

Suppose we wish to forecast the value of consumption for 2008. We could just use the last 5 years of data, which would give almost a perfect fit to a linear model. If we want an even more optimistic forecast, a quadratic function could be fitted to the last 10 points of data. If the quadratic appears strained, a second order ARMA model will provide similar results while hiding the 'cooking' of the data from a naïve audience. If the error process displays any irregularities, we could go to ARMA, ARCH, GARCH, or other types of complex error structures. If a pessimistic forecast is called for, restricting the data to start from 1991 will generate them. We can always find some event, such as change of monetary regimes, to justify introduction of a structural change, and hence discarding of previous data. This does not begin to exhaust the bag of tricks at our disposal. For example, we can take logs, differences, or make other suitable transforms of any or all variables before or during the modeling process. Faustman and White (1997) statement about the LSE approach that "one brings to the project a set of tools that virtually guarantees that one can find a model satisfying the test criteria on any dataset," is a valid critique of all econometric approaches to modeling. While we can make some intuitive judgment about the relative validity of the different approaches to forecasting described above, it is disturbing that we have no formal or theoretical criteria to guide us regarding this matter.

Figure 6.2 Consumption in Argentina



To put the problem in sharper focus, consider the issue of how we could prove correlation between two series, which are known a priori to be independent. Since (Granger) causality is just correlation with lags, these techniques can also be applied to prove causality between independent series. Here is a list of ways, all of which are used routinely by econometricians:

1. Select the series from among a large set. We found the strong correlation between Chinese and Australian consumption by looking at all the correlations in the data set.

2. Adjust the time period to suit. As in the data set above, by discarding different segments of the time series, we can change results to suit our tastes. Asghar (2007) gives several examples of cases where Granger causality shifts depending on the time period chosen. Since outliers have strong effects on Least Squares estimates, omitting or including a few observations can have a dramatic effect on standard estimates and results.

3. Use data transformations, such as logs, differences, percentage changes, Box-Cox transformations, or whatever fits best – Faust and Whiteman (1997) have aptly called some transformations "designer" variables.

4. Many concepts from economic theory can be creatively interpreted within a fixed data set. Wealth, income, consumption, permanent income, interest rate etc. are all flexible concepts and can be matched to different types of actual measures. We can find

dozens of measures of "openness" for economies, as well as evidence that authors searched for the definition which would suit their theses; see Rodriguez and Rodrik (2001).

5. It should be obvious that the use of suitable nonlinear functions would allow a perfect fit between any two variables. Nonparametric or semiparametric methods with a suitable collection of orthogonal basis functions systematize this procedure.

6. As Hendry (2000) has remarked, every model with a complex error process can be translated into a model with a simple error process and complex dynamics on observables. Thus apparently innocuous "generalization" to accommodate complex error processes can be used to fit arbitrarily complex dynamic models.

7. Even if all specification issues are settled correctly, there remains a vast choice of estimators and test procedures, which, like the Midas touch, can turn data into the gold of a significant result. Sufficient creativity in choosing instruments or appropriate moments to match for the generalized method of moments will always lead to the desired results. With vast numbers of tests, many with acknowledged low power, one can confirm the non-rejection of ones favorite hypothesis by selecting a suitable test.

For example, Baba et al. (1992) succeed in finding a stable money demand equation for M1 in USA up to 1980's. This time period contains several structural changes labeled "missing money" of the 70's and the "great velocity decline" of the 80's, which had frustrated previous attempts. Faust and Whiteman (1997) write that:

> Finding such a stable equation and a corresponding economic rationalization is a heroic achievement. The question is whether the achievement is testament to the ability of the method to uncover important economic regularities. The alternative, of course, is that the equation is testament to the ability of talented and imaginative practitioners to generate a relation that passes stability tests, regardless of the data.

They go on to discuss how subsequent failures of the model suggest that the second alternative holds and cast suspicion on the "designer" variables introduced to achieve good fit. The best practitioners can and do overfit, and our statistics fail to give us a clue on when this happens; this is a problem that needs serious consideration and further study.

7. LESSONS FROM HENDRY'S METHODOLOGY

Leamer (1983) remarked on the fact that the High Priests of econometrics who preach that models must be specified a priori undergo a remarkable transformation when they enter the computer lab in the basement. In less colorful language, it is virtually impossible to adhere to the maxim that regression models cannot be changed in response to the data. In economics, models are a dime a dozen while data is precious, so even if the first model fits well, exploration of fit of other models is carried out (and should be carried out) on the same data set. However, none of our statistics for validating and assessing regression models can cope with the issue of data-based model selection and change; see Jensen (2000) for a discussion and further references. Even if the selection is based on a glance at the data (as in the regression fits to Argentinian consumption data above), this disturbs the validity of the regression statistics which are based on the assumption that the model is "true" independent of the data, and before any observations are made. Leamer (1983) called this assumption the "Axiom of Correct Specification." He discussed how this is obviously false, since many econometric models simultaneously claim to be the unique true model for the data. At the same time, it is fundamental to econometric inference, since all conventional statistics are based directly on this assumption and invalid when it fails.

Hendry's (2000) methodology is the only one which explicitly takes the econometricians search for models into account. It has led to a number of deep insights into the process of systematic model search and its consequences, which have yet to be fully assimilated by the profession. The fact is the actual methodology utilized by econometricians (including Hendry and his followers; see Darnell and Evans, 1990) is much wilder and less subject to rules and regulation than the picture presented by Hendry. Even this tame picture overturns numerous suppositions built into the language used by econometricians to describe our work. We describe a few issues, which arise as an almost immediate consequence of viewing models as suitable reductions of the data generating process (DGP). For our present purposes, it suffices to consider the DGP as the unique true underlying stochastic process, which generates the observed data.

Suppose that the variables (X_t, Y_t, Z_t) have a trivariate normal distribution and are *i.i.d.* across time for t=1,2,…,T. Then the conditional distribution of X given (Y, Z) will be stable across time, and an econometrician who posits that X is determined by Y and Z will receive a perfect confirmation for his theory. However, the conditional distribution of Y given (X, Z) is also perfectly stable and the theory that X determines Y will also be confirmed by the data. In this situation, there are a large number of a priori theories, all of which will be correct, true, and perfectly compatible with the data. Any two of the variables may be taken as a cause for the third. Also, any variable can be taken as a cause for any other, with the third variable being a concomitant variable. The third variable Z can also be taken out of the picture by considering the marginal distribution of the first two (X, Y). This is also a perfectly valid reduction of the DGP, and yields a true bivariate model.

This existence of several true models is an embarrassment and does not correspond to the picture that econometricians have of the modeling process.

$$Y = a\,X + r \tag{7.1}$$

$$Y = b\,X + c\,Z + s \tag{7.2}$$

In the above framework, (7.1) and (7.2) are both valid reductions of the DGP. It does not make sense to think of the first model as misspecified (as per standard econometric terminology), and to think that estimates of "a" are biased because of the missing variable Z. This shows how difficult it would be to arrive at causal relations simply on the basis that the data conform to an a priori specification. Hendry proposes to solve the problem of multiple true models by using the concept of "variance encompassing" – the second model (7.2) is to be preferred because its residual will have smaller variance. In practical situations, the choice between (7.1) and (7.2) cannot be made on theoretical grounds. Missing observations, or noisy data for Z would lead to (7.1) being a preferred model. In forecasting situations, if Z cannot be accurately forecast, while X can be, (7.1) could be the preferred forecasting model.

Calculating the effects of interventions requires knowledge of causal mechanisms. In a smoothly functioning world with (X, Y, Z) as described above, causal mechanisms could not be discovered by observations. All possible causal relationships would consistent with the observations. Much as in the concept of natural experiments, Hendry proposes to use structural changes to find out about causal relations in this setup. If X is determined by Y, without reference to Z, then Y can be taken as the cause of X. If the marginal distribution of Z remains stable, the regression of Y and X and X on Y will both be stable relationships. If there is structural change and the distribution of Z shifts, then the distribution of X given Y will remains stable, while that of Y given X will shift and display instability. From this, we should be able to see that Y causes X and not the other way around. The fundamental idea appears sound – among a set of correlations, artificial ones will break down, and structural ones will persist through shakeups. Can the idea be implemented, and lead to revelations of structural relationships? Only a few researchers have used this approach, and we do not have enough evidence or experience to judge this as yet. For example, Hendry and Ericsson (1991) used this approach to argue that causation runs from prices to money, reversing the standard analysis. Subsequent failure of predictive validity of their 1991 model casts some doubt on their results, however.

Another major insight that results from taking model choice explicitly into account, and doing it in the systematic manner suggested by Hendry, is the realization that both models and residuals are simultaneously constructed by the econometrician. If valid models are required to have *i.i.d.* residuals, for example, then the modeling process will continue until residuals possessing such properties are obtained. The realization has led Hendry to change terminology for tests of residuals to "indices of conformity." Furthermore, the large collection of tools at our disposal virtually guarantees that we can find a model with satisfactory indices of conformity for any dataset, regardless of whether the underlying process satisfies the criteria (as noted by Hendry, 1993:24). This problem of overfitting plagues all areas of econometrics, and prevents the realization of the potential of a number of promising approaches, some of which have been mentioned earlier.

8. STRATEGIES FOR SUCCESS

In the process of translating a theory into implications about data, so many auxiliary assumptions are made that all contact with reality is lost. Any conflict can be resolved by adjusting the auxiliary assumptions. For example, suppose we want to learn if a production process satisfies diminishing, constant, or increasing returns to scale. The issue is of substantial significance from the point of view of theory. In carrying out the test, we assume a particular form of a production function, a particular way in which stochastic errors enter, and particular ways to aggregate and measure factors of production. The result of the test has no credibility, because we do not know what we are rejecting or accepting: the theory, or the auxiliary assumptions, or the ingenuity of the econometrician. This is why Keuzenkamps (2000) quotes Spanos' that 'no economic theory was ever abandoned because it was rejected by some empirical econometric test.'

Leamer (1978) has described the practice of econometricians as being a "specification search" – we look for a regression model, which conforms to the theory and also to the data. He thinks that this may be a fruitful task as an exploratory device, but taking this explicitly into account would substantially improve methodology. He has described the different types of devices, which could be used to produce models, which validate and confirm one's favorite econometric theory. First, one has to choose a set of observations, which measures a theoretical concept or category. For example, Rodriguez and Rodrik (2001) discuss how there is substantial flexibility in the choice of an index to measure the loose concept of "openness to trade." The choice of authors they cite appears to be guided by the need to produce a significant variable, which supports the theory that free trade promotes growth. Very often the chosen measure can be given several different interpretations, many of which would contradict the theory under study. In addition, one can choose the data set, by specifying the time period, the countries to be studied, etc. These choices may be justified by asserting that the theory is more likely to hold over this period and those countries, but is more often done in an opportunistic manner. Examples where slight changes in time period, choice of data set, functional forms, choices of lag length, modified measures for key variables, etc. significantly affect the final result of the analysis abound in the literature – indeed, nearly every analysis is of this type. Exploratory analysis of this analysis may have merit in producing interesting hypotheses, as Leamer argues, but cannot be used to provide evidence for causality.

As discussed earlier, I am not aware of a single instance where a causal mechanism was discovered by regression. In contrast, there are several stories of successful discovery of causal mechanisms in other areas; see Freedman (2008). Based on these success stories, I would like to offer some suggestions, which may help, improve the situation in the future.

8.1. Seek Confrontations with the Data

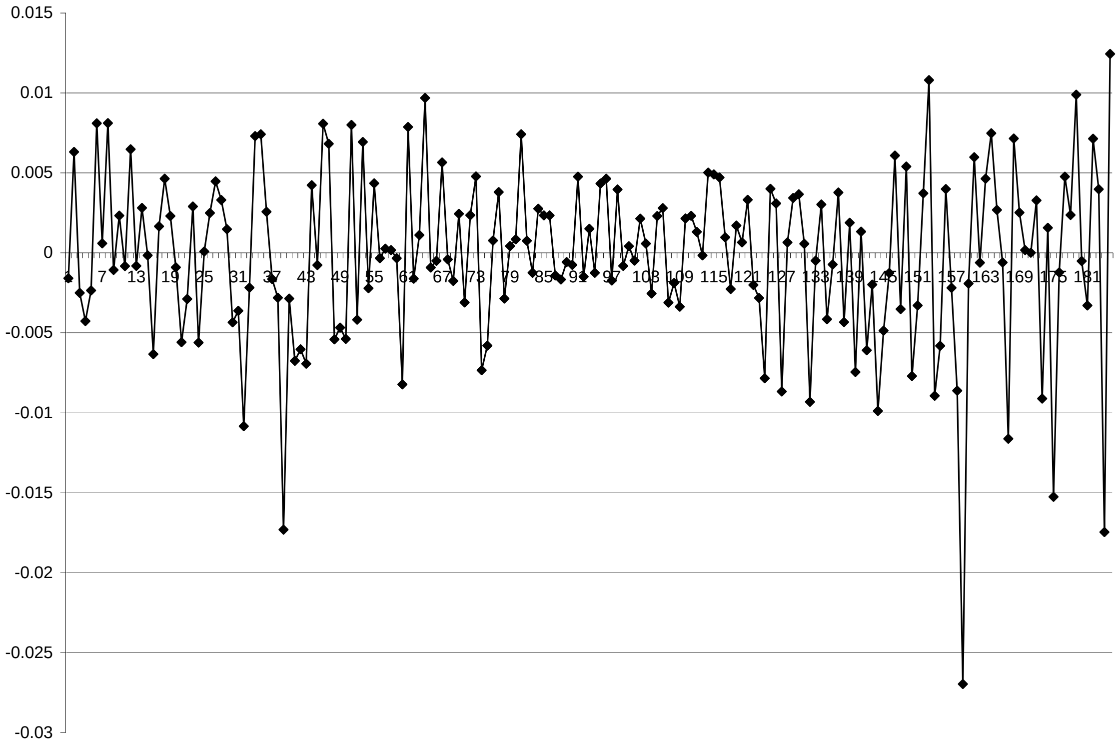
Current methodology seeks accommodation of the data, with calibration being an extreme example. Instead, we should follow positivist prescriptions: the power of a theory should be judged in terms of what observable phenomena it rules out. Theories should be valued for sharp predictions about observations, which can be refuted. Good theories will be those, which survive confrontations with data. Economics is capable of generating powerful theories with strong predictions about what may or may not be observed in the real world. Consider for example, the "efficient markets" hypothesis, which says that all profit opportunities are exploited. This means that studying past data on Yen-Dollar exchange rates cannot lead one to find a pattern useful for future predictions of this rate. In particular, the correlations between the percentage changes should be zero. In Figure 8.3, the graph of daily percentage change shows the (lack) of pattern predicted by economic theory. Regressions of the percentage change on lagged values confirm that there is no information in the past values on current exchange rate. The relation between theory and data is not mediated by a model of uncertain pedigree.

The idea that stock prices reflect the present discounted values of future returns can similarly be directly tested – and rejected. The stock prices are too volatile relative to the present discounted values of returns. This anomaly has led to deep examination of other possible explanations for determination of stock prices, and many advances in theories.

A clear rejection of static Keynesian consumption functions has led to substantial improvements in our understanding of the macro consumption function. Similarly, if consumers make consumption plans on the basis of their projections of lifetime income, then consumption paths should smooth out income fluctuations. Examination of relative volatility of consumption and income streams has led to greater understanding of the nature of the consumption function. In general, conflicts between theory and data have led to more sophisticated theory and data analysis in econometrics.

Anomalies, which could not be explained by dominant theories, have often played a key role in successful discoveries of causal structure, as documented in Freedman (2008). Instead of looking for auxiliary assumptions, functional forms, etc. to fit theories to data, we should actively seek ways of formulating theories so that they can directly confront data, and be capable of refutation by suitable observations. Such an empiricist methodology has a much better chance of discovering real causal chains.

Figure 8.3 Monthly Percent Change in Yen-Dollar Exchange Rates



8.2. Look for Alternative Explanations

A satisfactory fit between the data and any given theory is often the goal of econometric analysis. Examination of success stories for discovery of causal relations shows how important it is to go beyond this step. Having found a good fit, it is necessary to look for alternative explanations for the same data. If satisfactory alternative explanations cannot be found, this strengthens the case for our current theory, which does fit well. When there are other alternative explanations, which fit equally well, then we must do "differential diagnosis": look for ways to differentiate between competing explanations. There are numerous cases of how this was done cited in the literature. For example, the idea that there is a genetic predisposition to cancer, which also induces smoking, was disproven by the study of identical twins with different smoking habits. Active search for competing explanations and ways to differentiate between them has been an important element of successful discovery of causal mechanisms. The idea of "encompassing" which is one of the cornerstones of the Hendry methodology does capture an aspect of this, but this needs to be done more regularly and more widely than is current econometric practice.

8.3. Qualitative and Informal Supporting Evidence

This is one of the areas of greatest weakness, where more out of the box thinking is required. The role of hunches, informal reasoning, anomalies, and qualitative evidence in discovery of causal mechanisms is well documented – see Freedman (2008). More effort to find qualitative implications of hypotheses under consideration, and informal evaluations of these implications should prove rewarding. An interesting illustration of this is provided in Andrabi et al. (2007), who study education in Pakistan. Initial findings that parents send fewer girls to schools suggest discriminatory behavior. However, closer investigation shows that the “distance to school” is a key factor for the schooling decision for girls (and not so much for boys). Other data shows that parents invest heavily in schooling for children whom they consider bright, without discriminating between girls and boys. These hypotheses emerge only after expending “shoe leather,” – the authors went to the villages, and walked with the children to their schools. Mistakes regarding causal factor can and often do lead to wrong and ineffective policy prescriptions. On the ground investigations often reveal a reality at variance with general purpose a priori theories. Another important instance of this is furnished by Banerjee and Duflo (2004) who discuss how to reconcile empirical evidence with an aggregate production function and find it cannot be done.

In addition to using more shoe leather, investigating implications of hypotheses in different domains both qualitatively and quantitatively, econometricians need to do more informal and exploratory data analysis. History and inertia plays a large role in determining the shape of current practice. Econometric and statistical methods were developed in a pre-computer era, where many types of analyses were not possible. In particular graphical and exploratory data analysis of the type currently routine is not part of the standard training of econometricians. Making boxplots, graphs of different sorts, and in general doing informal data analyses would add a lot of value, and provide more convincing demonstrations than sophisticated econometric analyses in many situations.

8.4. Randomized and Behavioral Experiments

In general, randomization has been a powerful tool in disentangling causal chains. For obvious reasons, it has not been used much in economics. Nonetheless, it has been used successfully on a number of occasions in the past, and is becoming increasingly important. Burtless (1995), Greenberg and Shroder (2004), and many others provide strong evidence on the value of randomized experiments in providing evidence on genuine causal mechanisms. The Abdul Latif Jameel Poverty Action Lab, started in 2003 by Professors Banerjee, Duflo, and Mullainathan of MIT uses randomized experiments as basis for evaluating the effectiveness of poverty programs. Heckman, Lalonde and Smith (1999) discuss econometric evaluations of experiments in labor markets, which provide important evidence regarding the effects of policies.

Direct experimentation and data gathering unconstrained by theory and complex models has often led to substantial gains in understanding. Experiments by behavioral economists have led to many new insights and theoretical developments; see for example, Kahnemann (2003), and Barberis and Thaler (2003). Experimental evidence suggests that people increase efforts in response to wage increases, in conflict with neoclassical economic theory; see Falk and Fehr (2003). Andrabi et al. (2007) provide observational evidence, which shows that fourfold increase in teachers’ wages does not lead to improved educational outcomes in rural schools in Pakistan. The point I am trying to make here is that because causality is not directly observable, establishing causality requires piecing together evidence from different sources, and out of the box reasoning. Experimental and qualitative evidence provides strong supporting evidence, but no single piece of evidence may be conclusive.

9. CONCLUSIONS

Although there are many causes for the current confusions regarding causality in econometrics, my feeling is the over-specialization is at the heart of the matter. Julie Reuben (1996:176) writes that:

> Academic scientists coming of age in the early twentieth century rejected the utopian vision of science and the ideal of the unity of truth that had been so important to their predecessors. They embraced specialization and rejected efforts to synthesize all knowledge. They began to see the interests of their disciplines in a model of science that … associated objectivity with the rejection of moral values. … (They defined their role as) producing research and providing specialized training. This more limited role gave scientists more autonomy and freedom from administrative supervision.

Increasing specialization has led to the fragmentation of knowledge and a loss of vision regarding the grand enterprise of accumulation of knowledge to serve the humankind as a whole. Vartan Gregorian (1993), the president of Brown University, discusses many of the problems created by this fragmentation:

> Specialization, instead of uniting human beings into a general community of values and discourse, has by necessity divided them into small and exclusive coteries, narrow in outlook and interest. It isolates and alienates human beings. Social relations, as a result, cease to be the expression of common perceptions and common beliefs; they are reduced to political relations, to the interplay of competitive, and often antagonistic, groups. Specialized education makes our students into instruments to serve the specialized needs of a society of specialists.

In particular, the distinction between applied and theoretical work and specialization in this way has been very damaging. Theoreticians work on problems chosen for their mathematical structure, with no knowledge or awareness of, or even interest in, potential for applications. As a result, a massive amount of brainpower is brought to bear on problems, which are completely trivial when viewed in the perspective of problems, which currently threaten the future of mankind. At the same time, very little thought and effort goes into solving complex multidisciplinary problems of genuine importance. More efforts towards a holistic approach will prove rewarding. Encouraging authors to assess the relevance of their research towards solving genuine problems, and engaging more with the real world would create the motivation necessary for the discovery of genuine causal mechanisms.

REFERENCES

Aldrich, J. (1989). Autonomy. *Oxford Economic Papers New Series*, 41(1), 15-34.

Andrabi, T., D. Jishnu, A.I. Khwaja, T. Vishwanath and T. Zajonc (2007). PAKISTAN Learning and Educational Achievements in Punjab Schools (LEAPS): Insights to Inform the education policy debate. World Bank. http://go.worldbank.org/YUFOT05SA0 (accessed February 28, 2010).

Angrist, J.D. and B. Krueger Alan (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives*, 15(4), 69-85.

Ashley, R. (1998). A New Technique for Postsample Model Selection and Validation. *Journal of Economic Dynamics and Control*, 22, 647-65.

Baba, Y., D.F. Hendry and R.M. Starr (1992). The Demand for M1 in the U.S.A., 1960-1988. *Review of Economic Studies*, 59, 25-61.

Banerjee, A.V. and E. Duflo (2003). Inequality and Growth: What Can the Data Say? *Journal of Economic Growth*, 8(3), 267-299.

Banerjee, A.V. and E. Duflo (2004). Growth Theory through the Lens of Development Economics. *MIT Department of Economics Working Paper* No. 05-01.

Banerjee, A.V. and E. Duflo (2007). The Economic Lives of the Poor. *Journal of Economic Perspectives*, 21(1), 141-167.

Barro, R.J. (1997). *Determinants of Economic Growth: A Cross-Country Empirical Study*. MIT Press, Cambridge, MA.

Barberis, N. and R.H. Thaler (2002). A Survey of Behavioral Finance. *NBER Working Paper* No. W9222. http://ssrn.com/abstract=332266 (accessed February 28, 2010).

Blaug, M. (1998). Disturbing Currents in Modern Economics. *Challenge*. May-June 1998.

Bolch, B. (1998). Is Macroeconomics Believable? *The Independent Review*, 2(4) Spring.

Burtless, G. (1995), The Case for Randomized Field Trials in Economic and Policy Research. *The Journal of Economic Perspectives*, 9(2), 63-84.

Darnell, A.C. and J.L. Evans (1991). *The Limits of Econometrics*. Edward Elgar Publishers, Cheltenham, U.K.

Engle, R.F., D.F. Hendry and J.F. Richard (1983). Exogeneity. *Econometrica*, 51(2), 77-304.

Falk, A. and E. Fehr (2003). Why labour market experiments? *Labour Economics*, 10(4), 399-406.

Faust, J. and C.H. Whiteman, 1997, General-to-specific procedures fitting a data-admissible, theory-inspired, congruent, parsimonious, encompassing, weakly-exogenous, identified, structural model to the DGP: A translation and critique. In *Carnegie-Rochester Conference Series on Public Policy*, North-Holland, Amsterdam, 47, 121-161.

Freedman, D.A. (2008), On types of scientific enquiry: Nine success stories in medical research. In *The Oxford Handbook of Political Methodology*, ed. Janet M. Box-Steffensmeier, Henry E. Brady and David Collier. Oxford University Press, Oxford. http://www.stat.berkeley.edu/~census/anomaly.pdf (accessed February 28, 2010).

Freedman, D.A., R. Pisani and R. Purves (2007). *Statistics.* Norton, New York.

Freedman, D.A. (2005). *Statistical Models: Theory and Practice*. Cambridge University Press, Cambridge, MA.

Freedman, D. A. (2004). On specifying graphical models for causation and the identification problem. *Evaluation Review,* 26, 267-93. Reprinted in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. D.W.K. Andrews and J.H. Stock, Cambridge University Press (2005), 56-79. http://www.stat.berkeley.edu/~census/rothsag.pdf (accessed February 28, 2010).

Freedman, D.A. (1997). From Association to Causation via Regression. In *Causality in Crisis?* ed. V. McKim and S. Turner. University of Notre Dame Press, South Bend, (with discussion) 113-82. Reprinted in *Advances in Applied Mathematics*, 18, 59-110.

Freedman, D.A. (1991). Statistical Models and Shoe Leather. *Sociological Methodology*, 21, 291-313.

Gregorian, V. (1993). Technology, Scholarship, and the Humanities: The Implications of Electronic Information. http://www.cni.org/docs/tsh/Keynote.html (accessed February 28, 2010).

Heckman, J.J., R.J. LaLonde and J.A. Smith (1999). The Economics and Econometrics of Active Labor Market Programs. In *Handbook of Labor Economics*, ed. O. Ashenfelter, D. Card. Elsevier Science, New York, NY, 3, 1865-2097.

Hendry, D. F. (1995). *Dynamic Econometrics.* Oxford University Press, Oxford.

Hendry, D.F. (1993). The Roles of Economic Theory and Econometrics in Time-Series Economics. In *European Econometric Society*, Invited address, Stockholm.

Hendry, D.F. (2000). *Econometrics: Alchemy or Science? Essays in Econometric Methodology*. Oxford University Press, Oxford.

Hendry, D.F. and N.R. Ericsson (1991). An econometric analysis of UK money demand in `Monetary Trends in the United States and the United Kingdom by Milton Friedman and Anna J. Schwartz'. *American Economic Review*, 81, 8-38.

Hey, J. (1997). The Economic Journal: Report of the Managing Editor. *Royal Economic Society Newsletter*, January, 3-5.

Hoover, Kevin D. (2001). *Causality in Macroeconomics.* Cambridge University Press, Cambridge.

Hoover, Kevin D. (2004). Lost Causes. *Journal of the History of Economic Thought,* 26(2), 149-164.

Hoover, Kevin D. (2006). Causality in Economics and Econometrics. SSRN: http://ssrn.com/abstract=930739 (accessed April 17, 2010).

Jensen, D. (2000). Data Snooping, Dredging and Fishing: The Dark Side of Data Mining, A SIGKDD99 Panel Report. *SIGKDD Explorations*, 1(2), 52-54.

Kahneman, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. *American Economic Review*, 93(5), 1449-1475.

Keuzenkamp, H.A. (2000). *Probability, Econometrics and Truth: The Methodology of Econometrics*. Cambridge University Press, Cambridge.

Leamer, E. (1978). *Specification Searches: Ad Hoc Inference with Non Experimental Data*. John Wiley and Sons Inc., New York.

Leamer, E. (1983). Let's Take the Con Out of Econometrics. *American Economic Review*, 73, 31-43.

Lynn, R. and T. Vanhanen (2002). *IQ and the Wealth of Nations*. Praeger, Westport CT.

Magnus, J.R. (1999). The success of econometrics. *De Economist*, 147(1), 50-58.

Pearl, Judea (2000). *Causality*. Cambridge University Press, New York.

Phelps, E. (1999). *Seven Schools of Macroeconomics Thought*. Cambridge University Press, Cambridge.

Reuben, J. (1996). *The Making of the Modern University: Intellectual Transformation and the Marginalization of Morality*. University of Chicago Press, Chicago.

Rodríguez, F. and D. Rodrik (2001). Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence. In *Macroeconomics Annual 2000*, ed. Ben Bernanke and Kenneth S. Rogoff. MIT Press for NBER, Cambridge, MA.

Rodrik, D. (2005). Why We Learn Nothing from Regressing Economic Growth on Policies, *Harvard University Discussion Paper*.

Rossman, A.J. (1994). Televisions, Physicians, and Life Expectancy. *Journal of Statistics Education*, 2(2). http://www.amstat.org/publications/jse/v2n2/datasets.rossman.html (accessed February 28, 2010).

Sims, C.A. (1980). Macroeconomics and Reality. *Econometrica*, 48, 1-48.

Thomas, R.L. (1993). *Introductory Econometrics: Theory and Applications*. Second edition, Longman, New York.

Volken, T. (2003). IQ and the Wealth of Nations. A Critique of Richard Lynn and Tatu Vanhanen's Recent Book. *European Sociological Review*, 19(4), 411-412.

Zahid A. (2007). *Investigation of Granger Causality*, Ph.D. Thesis, Quaid-e-Azam University, Islamabad, Pakistan.