The Limits of Inference Without Theory

Kenneth I. Wolpin
University of Pennsylvania

Koopmans Memorial Lecture (2)
Cowles Foundation
Yale University
November 3, 2010

Introduction

"Fuller utilization of the concepts and hypotheses of economic theory … *as a part of the practices of observation and measurement* promises to be a shorter road, perhaps even the only road, to an understanding of cyclical fluctuations."

Tjalling C. Koopmans, "Measurement Without Theory," Cowles Commission Papers, No. 15, 1947

(italics in original)

Introduction

"Fuller utilization of the concepts and hypotheses of economic theory … *as a part of the practices of observation, measurement and* ***inference*** promises to be a shorter road, perhaps even the only road, to an understanding of **(fill in the blank)**."

Introduction

There appears to be a wide division within the micro-empirical community concerning what constitutes good inferentially-motivated empirical work.

Introduction

There appears to be a wide division within the micro-empirical community concerning what constitutes good inferentially-motivated empirical work.

That the disagreement entails a choice between “structural” and "reduced form" approaches is a *false* characterization.

Introduction

What distinguishes approaches to inferentially-motivated empirical research is the extent to which data and econometrics are related to economic models.

Introduction

What distinguishes approaches to inferentially-motivated empirical research is the extent to which data and econometrics are related to economic models.

Much empirical research eschews the use of theory as a way of justifying empirical specifications and interpreting results.

Introduction

What distinguishes approaches to inferentially-motivated empirical research is the extent to which data and econometrics are related to economic models.

Much empirical research eschews the use of theory as a way of justifying empirical specifications and interpreting results.

In that approach, statistical model parameters and auxiliary variables that serve as "controls" are not explicitly related to theory.

Introduction

This approach is predominant in many economics journals.

For example, in the maiden issue (January 2009) of the new AEA journal, *Applied Economics*, not a single paper includes an explicit economic model of the behavior that was being studied.

Introduction

Why does it matter?

Introduction

Why does it matter?

Let me give you three quite different examples where it does matter - the issue is, however, more general than the examples.

Introduction

The first example illustrates the importance of using theory in econometric specification and is relevant to dozens of papers spanning over 30 years of empirical research on unemployment duration.

This empirical literature was spawned by the development of search-theoretic models of unemployment - McCall (1970), Mortensen (1970).

Introduction

The second example illustrates the connection between theory and the recent econometrics literature on IV estimation with heterogeneous treatment response.

Heckman and Robb (1985), Bjorklund and Moffitt (1987), Imbens and Angrist (2005), Heckman (1997), Rosenzweig and Wolpin (2000), Heckman and Vytlacil (2005).

Introduction

The third example illustrates the importance of theory in addressing questions of interpretation and of external validity in an experimental setting.

Heckman and Smith (1995), Todd and Wolpin (2003), Deaton (2010), Heckman and Urzua (2010), Imbens (2009).

Introduction

There are many possible examples in the literature. I’ve chosen these for several reasons.

1. Their literatures are large.

2. The theory necessary for the illustrations is simple and easily described.

3. The examples are discussed in my papers.

Example 1 – Estimating the Effect of UI Benefits on Unemployment Duration

A great deal of effort has been expended on estimating the impact of the level of UI benefits on the duration of unemployment.

Usually, that empirical research appeals to the standard job search model.

Example 1 –UI Benefit Effect

Consider a standard infinite horizon search model. The reservation wage solves:

$$w^* = b + \frac{\lambda}{r}\int_{w^*}^{\infty}(x - w^*)dF(x)$$

$$= w^*(b, \frac{\lambda}{r}, F)$$

Thus, the reservation wage is a function of the *level of unemployment benefits*, the *ratio of the job offer arrival rate to the interest rate* and the *distribution of wage offers*.

This discussion is taken from Wolpin, *Empirical Methods for the Study of Labor Force Dynamics* (1995)

Example 1 –UI Benefit Effect

The hazard function of unemployment duration is given by

$$h(t_u) = \lambda(1 - F(w^*))$$

which implies that it is a function of

$$b, \lambda, r, F$$

Example 1 –UI Benefit Effect

A prototypical empirical paper of that genre specifies the hazard function as depending on observable variables that include something called the *replacement rate,* the ratio of the benefit amount to the wage on the job prior to the unemployment spell, that is,

$$b/w_{-1}$$

Example 1 –UI Benefit Effect

A prototypical empirical paper of that genre specifies the hazard function as depending on observable variables that include something called the *replacement rate,* the ratio of the benefit amount to the wage on the job prior to the unemployment spell, that is,

$$b/w_{-1}$$

It is in fact hard to find a paper that does not include either the replacement rate or the wage on the job prior to the unemployment spell.

Example 1 –UI Benefit Effect

However, neither the replacement rate nor w_{-1} itself appears in the hazard function derived from the search model.

Example 1 –UI Benefit Effect

So why do researchers include it?

Two possibilities:

Example 1 –UI Benefit Effect

So why do researchers include it?

Two possibilities:

1. The researcher has a different model in mind, one which includes w_{-1} in the structure in a way that introduces it into the hazard .

Example 1 –UI Benefit Effect

So why do researchers include it?

Two possibilities:

1. The researcher has a different model in mind, one which includes w_{-1} in the structure in a way that introduces it into the hazard.

In that case, the dynamic optimization problem, as well as the interpretation of the effect of UI benefits on search outcomes, will be considerably more complex. The agent must take into account the effect of accepting a wage in the given unemployment spell on the search problem in future unemployment spells (Ferrall, 1997).

Example 1 –UI Benefit Effect

2. The researcher believes that w_{-1} serves as a proxy for some omitted variable, for example, for some moment, such as the mean, of the wage offer distribution.

Example 1 –UI Benefit Effect

2. The researcher believes that w_{-1} serves as a proxy for some omitted variable, for example, for some moment, such as the mean, of the wage offer distribution.

The use of the proxy cannot be analyzed as a classical measurement error problem.

Example 1 –UI Benefit Effect

2. The researcher believes that w_{-1} serves as a proxy for some omitted variable, for example, for some moment, such as the mean, of the wage offer distribution.

The use of the proxy cannot be analyzed as a classical measurement error problem.

The reason is that the observed wage on the previous job is the outcome, that is, the accepted wage, of the prior search.

Example 1 –UI Benefit Effect

That accepted wage is therefore related to the same factors as the current search, including the UI benefit level.

Example 1 –UI Benefit Effect

That accepted wage is therefore related to the same factors as the current search, including the UI benefit level.

Varying the UI benefit level holding the previous accepted wage constant requires that the omitted factors, for which the previous accepted wage is serving as a proxy, also vary.

Example 1 –UI Benefit Effect

That accepted wage is therefore related to the same factors as the current search, including the UI benefit level.

Varying the UI benefit level holding the previous accepted wage constant requires that the omitted factors, for which the previous accepted wage is serving as a proxy, also vary.

The estimate of the UI effect in the presence of the proxy is therefore *biased* and the sign of the bias depends on which factors are unobserved.

Example 1 –UI Benefit Effect

Analyzing the effect of using proxy variables as "controls" relies on having an economic theory.

Example 1 –UI Benefit Effect

Analyzing the effect of using proxy variables as “controls” relies on having an economic theory.

A corollary is that “control” variables should be explicitly justified by the theory.

Example 2 – The Effect of Schooling on Earnings

Perhaps the single most frequently estimated parameter in economics is that of years of schooling in an earnings regression.

The main objective of that literature is to obtain an estimate that is free of ability bias.

Example 2 – The Effect of Schooling on Earnings

Among the most prominent papers in the literature, one that exemplifies the IV approach based on the use of a natural experiment, is that by Angrist and Kreuger (1991).

Example 2 – The Effect of Schooling on Earnings

Among the most prominent papers in the literature, one that exemplifies the IV approach based on the use of a natural experiment, is that by Angrist and Kreuger (1991).

AK exploit laws governing the ages at which children can enter and leave school that induce variation in completed schooling with respect to birth date.

Example 2 – The Effect of Schooling on Earnings

What is the interpretation we should give to the schooling effect estimated using the variation that AK exploit?

A reasonable strategy is to design a schooling decision model that captures that variation as closely as possible.

A Simple Model of Schooling Choice[1]

Assume:

1. that everyone works full-time for the same number of periods after leaving school so that actual work experience is the same as potential work experience.

1. This discussion is taken from Rosenzweig and Wolpin, “Natural ‘Natural Experiments’ in Economics,” (JEL, 2000).

A Simple Model of Schooling Choice

Assume:

1. that everyone works full-time for the same number of periods after leaving school so that actual work experience is the same as potential work experience.

2. Assume that there is only one decision period after reaching the compulsory schooling age, attend school or not attend school.

A Simple Model of Schooling Choice

Assume:

1. that everyone works full-time for the same number of periods after leaving school so that actual work experience is the same as potential work experience.

2. Assume that there is only one decision period after reaching the compulsory schooling age, attend school or not attend school.

3. Assume that there is a direct cost of attending school in that decision period.

A Simple Model of Schooling Choice

Define the attendance choice as

$$s_1 \in \{0, 1\}$$

A Simple Model of Schooling Choice

Define the attendance choice as

$$s_1 \in \{0, 1\}$$

and let S_0 be the number of years of schooling completed at the compulsory school leaving age. Then, completed schooling is

$$S_1 = S_0 + s_1$$

A Simple Model of Schooling Choice

Define the attendance choice as

$$s_1 \in \{0, 1\}$$

and let S_0 be the number of years of schooling completed at the compulsory school leaving age. Then, completed schooling is

$$S_1 = S_0 + s_1$$

Let the wage function be separable in schooling and other determinants of skill (work experience, x), but not in ability, μ :

$$\log y = f(S, \mu) + g(x, \mu)$$

A Simple Model of Schooling Choice

The present value of lifetime earnings for each schooling alternative is given by

$$V(s_1 = 1|S_0) = \exp[f(S_0 + 1, \mu)] \sum_{x=0}^{X} \beta^{x+1} \exp[g(x, \mu)] - c$$

$$V(s_1 = 0|S_0) = \exp[f(S_0, \mu)] \sum_{x=0}^{X} \beta^{x} \exp[g(x, \mu)]$$

A Simple Model of Schooling Choice

The present value of lifetime earnings for each schooling alternative is given by

$$V(s_1 = 1|S_0) = \exp[f(S_0 + 1, \mu)] \sum_{x=0}^{X} \beta^{x+1} \exp[g(x, \mu)] - c$$

$$V(s_1 = 0|S_0) = \exp[f(S_0, \mu)] \sum_{x=0}^{X} \beta^{x} \exp[g(x, \mu)]$$

The school attendance decision is:

$$s_1 = 1 \quad if \quad f(S_0 + 1, \mu) - f(S_0, \mu) \geq r + \log\left[\frac{c}{V(s_1=0|S_0)} + 1\right]$$

$$= 0 \quad \text{otherwise}$$

where $\beta = 1/(1 + r)$.

A Simple Model of Schooling Choice

Note that:

1. If $c = 0$, then we obtain the usual condition that attendance depends on whether the marginal return exceeds the interest rate: $\frac{\Delta \log y}{\Delta s} \geq r.$

A Simple Model of Schooling Choice

Note that:

1. If $c = 0$, then we obtain the usual condition attendance depends on whether the marginal return exceeds the interest rate: $\frac{\Delta \log y}{\Delta s} \geq r.$

2. If $\frac{\partial}{\partial \mu}[f(S_0+1,\mu) - f(S_0,\mu)] > 0$, then there exists a μ^* such that $s_1 = 1$ if $\mu \geq \mu^*$

A Simple Model of Schooling Choice

Given the condition in 2, those that attend school are of higher ability than those that do not attend.

A Simple Model of Schooling Choice

Given the condition in 2, those that attend school are of higher ability than those that do not attend.

Thus, $\frac{\Delta \log y}{\Delta s}$ as measured by the difference in earnings of the two schooling groups, will reflect ability differences.

A Simple Model of Schooling Choice

Given the condition in 2, those that attend school are of higher ability than those that do not attend.

Thus, $\frac{\Delta \log y}{\Delta s}$ as measured by the difference in earnings of the two schooling groups, will reflect ability differences.

The bias in the estimated schooling effect due to omitted (unobserved) ability is

$$E\left(\frac{\Delta \log y}{\Delta s}\right) = E_\mu\left[f(S_0+1,\mu)|\mu \geq \mu^*\right] - E_\mu\left[f(S_0,\mu)|\mu < \mu^*\right]$$

$$> E_\mu\left[f(S_0+1,\mu) - f(S_0,\mu)\right]$$

Example 2 – The Effect of Schooling on Earnings

Consider the interpretation of the Angrist and Krueger estimator within this schooling choice model.

Example 2 – The Effect of Schooling on Earnings

Consider the interpretation of the Angrist and Krueger estimator within this schooling choice model.

Assume that there are only two ability types:

$$\mu_1 \text{ and } \mu_2, \text{ with } \mu_1 > \mu_2.$$

Example 2 – The Effect of Schooling on Earnings

Consider the interpretation of the Angrist and Krueger estimator within this schooling choice model.

Assume that there are only two ability types:

$$\mu_1 \text{ and } \mu_2, \text{ with } \mu_1 > \mu_2.$$

The proportion of type 1 individuals is π_1 and the proportion of type 2's $1 - \pi_1$. Assume that the types are independent of date of birth (instrument validity).

Example 2 – The Effect of Schooling on Earnings

Consider the interpretation of the Angrist and Krueger estimator within this schooling choice model.

Assume that there are only two ability types:

$$\mu_1 \text{ and } \mu_2, \text{ with } \mu_1 > \mu_2.$$

The proportion of type 1 individuals is π_1 and the proportion of type 2's $1 - \pi_1.$ Assume that the types are independent of date of birth (instrument validity).

Suppose the optimal level of schooling for type 1's is $S_0 + 1$ and that of type 2's $S_0.$

Example 2 – The Effect of Schooling on Earnings

Compare two sets of children, those who just make the school entry date of birth deadline (older children) and those who just miss the deadline (younger children) – they differ in age, say, by only 1 day.

Example 2 – The Effect of Schooling on Earnings

Compare two sets of children, those who just make the school entry date of birth deadline (older children) and those who just miss the deadline (younger children) – they differ in age, say, by only 1 day.

The type 1 children complete $S_0 + 1$ years regardless of their date of birth because it's optimal for them.

Example 2 – The Effect of Schooling on Earnings

Compare two sets of children, those who just make the school entry date of birth deadline (older children) and those who just miss the deadline (younger children) – they differ in age, say, by only 1 day.

The type 1 children complete $S_0 + 1$ years regardless of their date of birth because it's optimal for them.

The older type 2 children complete S_0 years because that is optimal for them.

Example 2 – The Effect of Schooling on Earnings

Compare two sets of children, those who just make the school entry date of birth deadline (older children) and those who just miss the deadline (younger children) – they differ in age, say, by only 1day.

The type 1 children complete $S_0 + 1$ years regardless of their date of birth because it's optımal for them.

The older type 2 children complete S_0 years because that is optimal for them.

But, the younger type 2 children are forced to remain in school an extra year because they reach the school leaving age only after spending $S_0 + 1$ years in school.

Example 2 – The Effect of Schooling on Earnings

To get the Wald estimate of the schooling effect, note that:

Mean earnings for younger type 1's $= f(S_0 + 1, \mu_1)$

Mean earnings for younger type 2's $= f(S_0 + 1, \mu_2)$

Mean earnings for older type 1's $= f(S_0 + 1, \mu_1)$

Mean earnings for older type 2's $= f(S_0, \mu_2)$

Example 2 – The Effect of Schooling on Earnings

Mean earnings of younger children

$$= \pi_1 f(S_0 + 1, \mu_1) + (1 - \pi_1) f(S_0 + 1, \mu_2)$$

Mean earnings of older children

$$= \pi_1 f(S_0 + 1, \mu_1) + (1 - \pi_1) f(S_0, \mu_2)$$

Example 2 – The Effect of Schooling on Earnings

Thus, the difference in mean earnings between the younger and older children is

$$\pi_1 f(S_0 + 1, \mu_1) + (1 - \pi_1) f(S_0 + 1, \mu_2) \; -$$

$$\pi_1 f(S_0 + 1, \mu_1) + (1 - \pi_1) f(S_0, \mu_2)$$

Example 2 – The Effect of Schooling on Earnings

Thus, the difference in mean earnings between the younger and older children is

$$\pi_1 f(S_0+1,\mu_1) + (1-\pi_1) f(S_0+1,\mu_2) \;-$$

$$\pi_1 f(S_0+1,\mu_1) + (1-\pi_1) f(S_0,\mu_2)$$

$$= (1-\pi_1)[f(S_0+1,\mu_2) - f(S_0,\mu_2)]$$

Example 2 – The Effect of Schooling on Earnings

Thus, the difference in mean earnings between the younger and older children is

$$\pi_1 f(S_0 + 1, \mu_1) + (1 - \pi_1) f(S_0 + 1, \mu_2) \;\; -$$

$$\pi_1 f(S_0 + 1, \mu_1) + (1 - \pi_1) f(S_0, \mu_2)$$

$$= (1 - \pi_1)[f(S_0 + 1, \mu_2) - f(S_0, \mu_2)]$$

and the change in the population mean schooling is

$$\pi_1 0 + (1 - \pi_1) 1$$

Example 2 – The Effect of Schooling on Earnings

Thus, the Wald estimate is

$$\frac{\Delta E(Y)}{\Delta E(S)} = f(S_0 + 1, \mu_2) - f(S_0, \mu_2).$$

Example 2 – The Effect of Schooling on Earnings

Thus, the Wald estimate is

$$\frac{\Delta E(Y)}{\Delta E(S)} = f(S_0 + 1, \mu_2) - f(S_0, \mu_2).$$

This is the marginal effect of schooling on earnings for the *less able* only.

Example 2 – The Effect of Schooling on Earnings

When would the Wald estimate equal the marginal effect for the population (not just for the less able)?

Example 2 – The Effect of Schooling on Earnings

When would the Wald estimate equal the marginal effect for the population (not just for the less able)?

$$f(S, \mu) = f_1(S) + f_2(\mu)$$

i.e., if the marginal effect of schooling is independent of ability.

Example 2 – The Effect of Schooling on Earnings

An alternative instrument, based on sibling gender, is proposed by Butcher and Case (1994).

Example 2 – The Effect of Schooling on Earnings

An alternative instrument, based on sibling gender, is proposed by Butcher and Case (1994).

A reasonable interpretation is that the instrument creates variation in the cost of schooling.

Example 2 – The Effect of Schooling on Earnings

An alternative instrument, based on sibling gender, is proposed by Butcher and Case (1994).

A reasonable interpretation is that the instrument creates variation in the cost of schooling.

In that case, one can show that the Wald estimate recovers the marginal effect of schooling on earnings for the *more able* only.

Example 2 – The Effect of Schooling on Earnings

An alternative instrument, based on sibling gender, is proposed by Butcher and Case (1994).

A reasonable interpretation is that the instrument creates variation in the cost of schooling.

In that case, one can show that the Wald estimate recovers the marginal effect of schooling on earnings for the *more able* only.

The use of the simple schooling model provides a set of underlying assumptions under which the AK and BC interpretations of their IV estimators are valid.

Example 3 – The Effect of Class Size on Student Performance[1]

Assume that we have a randomized field experiment.

1. This discussion is taken from Todd and Wolpin, "On the Specification and Estimation the Production Function for Cognitive Achievement," (EJ, 2003)

Example 3 – The Effect of Class Size

Assume that we have a randomized field experiment.

To make it concrete, consider the Tennessee STAR experiment in which students and teachers were randomly assigned to classes of different sizes.

Example 3 – The Effect of Class Size

Assume that we have a randomized field experiment.

To make it concrete, consider the Tennessee STAR experiment in which students and teachers were randomly assigned to classes of different sizes.

The experiment is used to estimate the relationship between class size and cognitive achievement either in current or later grades.

Example 3 – The Effect of Class Size

Consider a regression of a measure of cognitive achievement in grade g, T_g, on class size, $C_{g'}$, where $g \geq g'$,

$$T_g = \alpha C_{g'} + u,$$

and where α is the effect of class size in grade g' on measured achievement in grade g.

Example 3 – The Effect of Class Size

How should we interpret this effect?

Example 3 – The Effect of Class Size

How should we interpret this effect?

That depends on which of the following is true:

$E(u|C_{g'}) = 0$ or $E(u|C_{g'}) \neq 0$

Example 3 – The Effect of Class Size

How should we interpret this effect?

That depends on which of the following is true:

$$E(u|C_{g'}) = 0 \quad \text{or} \quad E(u|C_{g'}) \neq 0$$

How do we assess which of these is true and why would we care?

Example 3 – The Effect of Class Size

First, in order to assess which is true, we need to know what's in u and to know what's in u, we need to know where the regression comes from, that is, what model would generate such a relationship.

Example 3 – The Effect of Class Size

First, in order to assess which is true, we need to know what's in u and to know what's in u, we need to know where the regression comes from, that is, what model would generate such a relationship.

One reasonable way to interpret the regression is as a production function in which the achievement measure is the output and class size is the input. Then, what's in u are all the other inputs that determine achievement (as well as invariant endowments, all possibly interacting with class size).

Example 3 – The Effect of Class Size

We know that from any underlying optimization problem in which parents are concerned about their children's achievement, an exogenous change in one input would normally imply an adjustment in other inputs.

Example 3 – The Effect of Class Size

We know that from any underlying optimization problem in which parents are concerned about their children's achievement, an exogenous change in one input would normally imply an adjustment in other inputs.

For example, if parents thought that a larger class size would adversely impact their child's achievement, they might work more themselves with the child or hire a private tutor. Or, teachers might use different teaching methods for larger class sizes.

Example 3 – The Effect of Class Size

We know that from any underlying optimization problem in which parents are concerned about their children's achievement, an exogenous change in one input would normally imply an adjustment in other inputs.

For example, if parents thought that a larger class size would adversely impact their child's achievement, they might work more themselves with the child or hire a private tutor. Or, teachers might use different teaching methods for larger class sizes.

The class size effect reflects all of these adjustments and cannot itself be interpreted as a production function parameter.

Example 3 – The Effect of Class Size

So, one point is simply that to interpret the effect requires a theory, that is, a statement as to where the relationship comes from and an answer to *what's in* u?"

Example 3 – The Effect of Class Size

So, one point is simply that to interpret the effect requires a theory, that is, a statement as to where the relationship comes from and an answer to *what's in* u?

But, suppose one argues that all we had wanted to identify was this effect of class size on achievement, that is, the *policy impact*.

Example 3 – The Effect of Class Size

Theory implies that the change in other inputs induced by differences in class size will depend on the circumstances of the families in the experiment.

Example 3 – The Effect of Class Size

Theory implies that the change in other inputs induced by differences in class size will depend on the circumstances of the families in the experiment.

Without understanding how the response of those inputs to the change in class size differs among families, we cannot generalize the estimated class size effect to other settings.

Example 3 – The Effect of Class Size

It is possible to make progress in this direction if one had collected additional information within the experiment about other inputs and their determinants (e.g., family income).

Example 3 – The Effect of Class Size

It is possible to make progress in this direction if one had collected additional information within the experiment about other inputs and their determinants (e.g., family income).

But, to know what information to collect or even to recognize the necessity of it, requires the researcher cum experimenter to have thought about the theory.

Example 3 – The Effect of Class Size

It is possible to make progress in this direction if one had collected additional information within the experiment about other inputs and their determinants (e.g., family income).

But, to know what information to collect or even to recognize the necessity of it, requires the researcher cum experimenter to have thought about the theory.

The best field experiments are those that have also collected auxiliary information that can be used to help extrapolate beyond the setting of the experiment (e.g., as in the Mexican Progresa program).

Conclusion

"But *the decision* not to use theories of man's economic behavior, even hypothetically, *limits the value to economic science and to the makers of policies*, of the results obtained or obtainable by the methods developed."

Tjalling C. Koopmans (Measurement Without Theory, Cowles Commission Papers, No. 15, 1947)

(italics added).