The Identification Zoo: Meanings of Identification in Econometrics

By Arthur Lewbel

Journal of Economic Literature 57.4 (2019): 835-903.

James J. Heckman

THE UNIVERSITY OF CHICAGO

- Over two dozen different terms for identification appear in the econometrics literature, including set identification, causal identification, local identification, generic identification, weak identification, identification at infinity, and many more.

- This survey:
  - **(i)** gives a new framework unifying existing definitions of point identification;
  - **(ii)** summarizes and compares the zooful of different terms associated with identification that appear in the literature; and
  - **(iii)** discusses concepts closely related to identification, such as normalizations and the differences in identification between structural models and causal, reduced form models. (JEL C01, C20, C50)

1. Introduction

- Econometric identification really means just one thing: model parameters or features being uniquely determined from the observable population that generates the data

- This survey then discusses the differences between identification in traditional structural models versus the so-called reduced form (or causal inference, or treatment effects, or program evaluation) literature

- Concepts that are closely related to identification, including normalizations, coherence, and completeness are also discussed

- The study of identification logically precedes estimation, inference, and testing

- The next section, **section 2**, begins by providing some historical background
- **Section 3** then provides examples of, and methods for obtaining, point identification
- Next is **section 4**, which defines and discusses the concepts of coherence and completeness of models
- This is followed by **section 5**, which is devoted to discussing identification concepts in what is variously known as the reduced form, or program evaluation, or treatment effects, or causal inference literature

- **Section 6** describes nonparametric identification, semiparametric identification, and set identification

- **Section 7** describes limited forms of identification, in particular, local identification and generic identification

- **Section 8** considers forms of identification that have implications for, or are related to, statistical inference

- **Section 9** then concludes, and an appendix provides some additional mathematical details

2. Historical Roots of Identification

- Before we can think about isolating, and thereby identifying, the effect of one variable on another, we need the notion of "ceteris paribus," that is, holding other things equal

- The textbook example of an identification problem in economics, that of separating supply and demand curves, appears to have been first recognized by Philip Wright (1915), who pointed out that what appeared to be an upward sloping demand curve for pig iron was actually a supply curve, traced out by a moving demand curve

- A standard identification problem in the statistics literature is that of recovering a treatment effect

- A different identification problem is that of identifying the true coefficient in a linear regression when regressors are measured with error

3. Point Identification

- In modern terminology, the standard notion of identification is formally called point identification

- Early formal definitions of (point) identification were provided by
  - Koopmans and Reiersøl (1950),
  - Hurwicz (1950),
  - Fisher (1966), and
  - Rothenberg (1971)

- In this survey I provide a new general definition of identification

- This generalization maintains the intuition of existing classical definitions while encompassing a larger class of models than previous definitions

3.1. Introduction to Point Identification

- Recall that θ is the parameter (which could include vectors and functions) that we want to identify and ultimately estimate

- Assume also that we have a model, which typically imposes some restrictions on the possible values ϕ could take on

- **Example 1:**

  Suppose for scalars Y, X, and θ, our model is that $Y = X\theta + e$ where $E(X^2) \neq 0$ and $E(eX) = 0$, and suppose that ϕ, what we can learn from data, includes the second moments of the vector (Y, X)

- **Example 2:**

  Let the model be that a binary treatment indicator X is assigned to individuals by a coin flip, and Y is each individual's outcome

- When discussing empirical work, a common question is, "what is the source of the identification?"
  That is, what feature of the data is providing the information needed to determine θ?
  This is essentially asking, what needs to be in ϕ?

- The definition of identification is somewhat circular or recursive

- We usually think of a model as a set of equations describing behavior

- A common starting assumption is that the DGP consists of *n* independently, identically distributed (IID) observations of a vector *W*,
  where the sample size *n* goes to infinity

- With more complicated DGPs (e.g., time series data, or cross section data containing social interactions or common shocks), part of the challenge in establishing identification is characterizing what information ϕ is knowable, and hence appropriate to use as the starting point for proving identification

- Even in the most seemingly straightforward situations, such as experimental design with completely random assignment into treatment and control groups, additional assumptions regarding the DGP (and hence regarding the model and ϕ) are required for identification of treatment effects

- In practice, it is often useful to distinguish between two types of DGP assumptions

3.2. Defining Point Identification

- Here we define point identification and some related terms, including structure and observational equivalence

- Define a model M to be a set of functions or constants that satisfy some given restrictions

- Examples of what might be included in a model are regression functions, error distribution functions, utility functions, game payoff matrices, and coefficient vectors

- Define a model value m to be one particular possible value of the functions or constants that comprise M

- Define ϕ to be a set of constants and/or functions about the DGP that we assume are known or knowable from data

- Define a set of *parameters* θ to be a set of unknown constants and/or functions that characterize or summarize relevant features of a model

- The set of parameters θ may also include *nuisance* parameters, which are defined as parameters that are not of direct economic interest, but may be required for identification and estimation of other objects that are of interest

- We assume that each particular value of *m* implies a particular value of ϕ and of θ

- Two parameter values, θ and $\tilde{\theta}$, are defined to be *observationally equivalent* if there exists a ϕ such that both $s(\phi, \theta)$ and $s(\phi, \tilde{\theta})$ are not empty

- Let Θ denote the set of all possible values that the model says θ could be

- We say that the *model is point identified* when no pairs of model values m and $\tilde{m}$ in M are observationally equivalent (treating m and $\tilde{m}$ as if they were parameters)

- The concepts of local and generic identification deal with cases where we can't establish point identification for all θ in Θ

- *Local identification* of θ_0 means that there exists a neighborhood of θ0 such that, for all values $\theta \neq \theta_0$ in this neighborhood, θ is not observationally equivalent to θ_0

- *Generic identification* roughly means that the set of values of θ in Θ that cannot be point identified is a very small subset of Θ

- *Parametric* identification is where θ is a finite set of constants and all the different possible values of ϕ also correspond to different values of a finite set of constants

- *Nonparametric identification* is where θ consists of functions or infinite sets

- Other cases are called *semiparametric identification*, which includes situations where, for example, θ includes both a vector of constants and nuisance parameters that are functions

3.3. Examples and Classes of Point Identification

- *Example 1: a median*

  Let the model M be the set of all possible distributions of a random variable W having a strictly monotonically increasing distribution function

- How does this example fit the general definition of identification?

- *Example 2: Linear regression*

  Consider a DGP consisting of observations of Y, X where Y is a scalar and X is a K-vector

- *Example 3: Treatment*

  Suppose the DGP consists of individuals who are assigned a treatment of $T = 0$ or $T = 1$, and each individual generates an observed outcome Y

- The key point for identification is not whether we can write a closed-form expression like $E(Y|T=1) - E(Y|T=0)$ for θ, but whether there exists a unique value of θ corresponding to every possible ϕ

- *Example 4: Linear Supply and Demand*

  Consider the textbook example of linear supply and demand curves

- *Example 5: Latent Error Distribution*

  Suppose the DGP is IID observations of scalar random variables Y, X, so ϕ is the joint distribution of Y, X

- In **examples 1, 2**, and **5** above, data are assumed to be IID observations of some vector we can call W, and therefore what we start by assuming is knowable, ϕ, is the distribution function of W

- Many identification arguments in econometrics begin with one of three cases:
  - either ϕ is a set of reduced-form regression coefficients, or
  - ϕ is a data distribution, or
  - ϕ is the maximizer of some function

- *Wright–Cowles Identification*

  The notion of identification most closely associated with the Cowles foundation concerns the simultaneity of linear systems of equations like supply and demand equations

- The model M is a set of linear structural equations

- A convenient feature of Wright–Cowles identification is that it can be applied to time series, panel, or other DGPs with dependence across observations, as long as the reduced form linear regression coefficients have some well-defined limiting value ϕ

- *Distribution-Based Identification*

  Distribution-based identification is equivalent to the general definition of identification given by Matzkin (2007, 2012)

- Here, θ could be parameters of a parameterized distribution function, or features of the distribution ϕ like moments or quantiles, including possibly functions like conditional moments

- Note a key difference between Wright–Cowles and distribution-based identification is that the latter assumes an entire distribution function is knowable, while the former is based on just having features of the first and second moments of data be knowable

- *Extremum-Based Identification*

- **Extremum estimators** are estimators that maximize an objective function, such as generalized method of moments (GMM) or least squares estimation

- To see the connection between extremum-based identification and estimation, consider the example of extremum estimators that maximize an average with IID data

- Suppose G is, as above, the probability limit of the objective function of a given extremum estimator

- Suppose we had considered extremum-based identification where the model consists of G functions defined by $G(\zeta) = -E[(W - |\zeta|)^2]$

3.4. Demonstrating Point Identification

- How can we show that a given set of parameters θ are point identified?
- Identification by construction means that we can write a closed-form expression for θ as a function of ϕ
- An important example of a direct consistency proof is the Glivenko-Cantelli theorem
- Another general method of showing that a parameter is identified is to prove that the true θ_0 is the unique solution to some maximization problem defined by the model
- Examples of identification proofs that apply many of the above techniques can be found in Matzkin (2005, 2007, 2012)
- An interesting property of point identification is that it can be applied without reference to any data at all

3.5. Common Reasons for Failure of Point Identification

- Parameters θ often fail to be point identified for one of six somewhat overlapping reasons:
  - Model incompleteness
  - Perfect collinearity
  - Nonlinearity
  - Simultaneity
  - Endogeneity
  - Unobservability

- **Incompleteness** arises in models where the relationships among variables are not fully specified

- **Perfect collinearity** is the familiar problem in linear regression that one cannot separately identify the coefficients in a linear regression like

$$Y_i = a + b\, X_i + c\, Z_i + e_i$$

- **Nonlinearity** can cause nonidentification by allowing equations to have multiple solutions
- **Simultaneity** is the familiar source of nonidentification that arises from X and Y being determined jointly or simultaneously, as in the case of price and quantity in a market
- **Endogeneity** is the general problem of regressors being correlated with errors
- Many models contain **unobserved** heterogeneity, which typically takes the form of nonadditive or non-separable error terms

3.6. Control Variables

- **"I controlled for that."** This is perhaps the commonest response to a potential identification question in econometric modeling
- The solution of adding a control variable refers to the inclusion of another variable Z in the model to fix this problem
- There are two reasons why simply including covariates intended to act as controls may not fix these identification problems, and indeed can potentially make them worse:
  - The first reason is functional form
  - The second reason is that Z itself could be endogenous, and the problems resulting from adding an endogenous Z regressor to the model could be worse than the confounding issue

- In the causal diagram literature (see, e.g., Pearl 2000, 2009), a distinction is made between "**confounders**" and "**colliders**."

- A similar argument applies to difference-in-difference models

- These issues with potential controls are closely related to the Berkson (1946) and Simpson (1951) paradoxes in statistics

3.7. Identification by Functional Form

- **Identification** by functional form is when identification holds when we assume some functions in the model have specific parametric or semiparametric forms, but where identification may fail to hold without these parametric or semiparametric restrictions

- As noted in **section 3.5**, models that are nonlinear in parameters may fail to be identified because nonlinear equations can have multiple solutions

- Suppose we continue with the classical Cowles model considered in example 4 in **sections 3.3** and **3.5**, except that now, while the demand curve is still

$$Y = bX + cZ + U,$$

we let the supply curve be

$$Y = dX^2 + aX + \varepsilon$$

- Formally proving identification entails showing that the equations

$$E(Y - dX^2 - aX|Z = z) = 0 \text{ and } E(Y - bX - cZ|Z = z) = 0$$

  for all z on the support of Z can be uniquely solved for a, b, c, and d

- Historically, identification by functional form assumed completely parameterized models with no unknown functions

- Still another example of identification by functional form is the model

$$Y = a + bX + U$$

- In particular, **identification** based on functional form, such as constructed instruments, can be used to provide overidentifying information for model tests and robustness checks (see the next section for the definition of overidentification)

3.8. Over-, Under-, and Exact Identification, Rank and Order Conditions

- Models often contain collections of equalities involving θ

- Common examples are conditional or unconditional moments, i.e., equations of the form

$$E[g(W, \theta)] = 0 \text{ or}$$
$$E[g(W, \theta) \,|Z] = 0$$

- We then say that parameters θ are *exactly* ***identified*** if removing any one these equalities causes θ to no longer be point identified

- The parameters are ***overidentified*** when θ can still be point identified after removing one or more of the equalities, and they are ***underidentified*** when we do not have enough equalities to point identify θ

- Generally, when parameters are overidentified, it is possible to test validity of the moments used for identification

4. Coherence, Completeness, and Reduced Forms

- Although often ignored in practice, consideration of coherence and completeness of models should logically precede the study of identification

- Incoherent or incomplete models arise in some simultaneous games

- Entry games are an example of a system of equations involving discrete endogenous variables

- To illustrate, consider the simple model:

$$Y_1 = I(Y_2 + U_1 \geq 0),$$
$$Y_2 = \theta Y_1 + U_2$$

- if we replace the above model with

$$Y_1 = I(DY_2 + U_1 \geq 0),$$
$$Y_2 = (1 - D)\, \theta Y_1 + U_2$$

5. Causal Reduced-form versus Structural Model Identification

- Among economists doing empirical work, recent years have seen a rapid rise in the application of so-called reduced-form or causal inference methods, usually based on randomization

- Proponents of these methods often refer to their approach as a reduced-form methodology

- Two key characteristics of causal methods are:

  (i) a focus on identification and estimation of treatment effects rather than deep parameters, and

  (ii) an emphasis on natural or experimental randomization as a key source of identification

5.1. Randomized Causal or Structural Modeling? Do Both

- Before getting into details regarding the two methodologies, it should be pointed out that the perceived conflict between proponents of **causal**, **reduced-form** methods versus **structural modeling** approaches is somewhat artificial

- For both **identification** and **estimation**, the strengths of both approaches may be combined in many ways, including these:

  (i) Causal analyses based on randomization can be augmented with structural econometric methods to deal with identification problems caused by data issues such as attrition, sample selection, measurement error, and contamination bias

  (ii) It is not just reduced-form methods that require instrument independence

- For both **identification** and **estimation**, the strengths of both approaches may be combined in many ways, including these:

  (iii) Identifiable causal effects can provide useful benchmarks for structural models

  (iv) Economic theory and structure can provide guidance regarding the external validity of causal parameters

  (v) One can use causal methods to link randomized treatments to observable variables and use structure to relate these observables to more policy relevant treatments and outcomes

  (vi) Big data analyses on large data sets can uncover promising correlations

  (vii) Structural type assumptions can be used to clarify when and how causal effects may be identified

5.2. Randomized Causal versus Structural Identification: An Example

- An obstacle to comparing **causal** versus **structural** analyses is that these methods are usually described using different notations

- The example structural model considered here will be the linear regression model

$$Y = a + bT + e$$

- What makes one model or analysis structural and another causal?

- We begin with a general triangular model, where Y is determined by T along with error terms and T is determined by Z and errors

- Because both T and Z are binary, we can without loss of generality write this model as a linear random coefficients model

$$Y = U_0 + U_1 T \text{ and}$$
$$T = V_0 + V_1 Z$$

- The treatment effect for an individual is defined as $Y(1)$ - $Y(0)$, or equivalently as U_1, which is the difference between the outcome one would have if treated versus not treated

- A common assumption in the causal inference literature is the stable unit treatment value assumption (SUTVA)

- With just the assumptions we have so far, the parameter c satisfies

$$c = \frac{\operatorname{cov}(Z, Y)}{\operatorname{cov}(Z, T)} = \frac{\operatorname{cov}(Z, U_0 + U_1 T)}{\operatorname{cov}(Z, (V_0 + V_1 Z))}$$

$$= \frac{\operatorname{cov}(Z, U_0 + U_1(V_0 + V_1 Z))}{\operatorname{cov}(Z, (V_0 + V_1 Z))} = \frac{\operatorname{cov}(Z, U_1 V_1 Z)}{\operatorname{cov}(Z, V_1 Z)}$$

$$= \frac{E(U_1 V_1)\operatorname{var}(Z)}{E(V_1)\operatorname{var}(Z)} = \frac{E(U_1 V_1)}{E(V_1)}.$$

- The difference between our particular causal and structural models will consist only of different assumptions regarding the equation

$$c = E(U_1\, V_1)/E(V_1)$$

- Note from previous slide that

$$\begin{aligned} c &= \frac{E(U_1 V_1)}{E(V_1)} \\ &= \frac{E(U_1)\, E(V_1) + \mathrm{cov}(U_1, V_1)}{E(V_1)} \\ &= E(U_1) + \frac{\mathrm{cov}(U_1, V_1)}{E(V_1)} \end{aligned}$$

- Now consider a causal identification argument
- Let P_v denote the probability that $V_1 = v$. Then, by definition of expectations

$$c = \frac{E(U_1V_1)}{E(V_1)}$$

$$= \frac{E(U_1V_1|V_1=1)P_1 + E(U_1V_1|V_1=0)P_0 + E(U_1V_1|V_1=-1)P_{-1}}{E(V_1|V_1=1)P_1 + E(V_1|V_1=0)P_0 + E(V_1|V_1=-1)P_{-1}}$$

$$= \frac{E(U_1|V_1 = 1)P_1 - E(U_1|V_1 = -1)P_{-1}}{P_1 - P_{-1}}.$$

- Interestingly, if one imposes both our structural assumption cov(U_1, V_1)=0 and our reduced-form "no defiers" restriction $P_{-1} = 0$, then it can be shown that

$$0 = \left[E(U_1|V_1 = 0) - E(U_1|V_1 = 1)\right](1 - P_1)$$

- Even with binary treatment, it is possible to relax the above listed assumptions in both the structural and the causal framework

- Given our independence and exclusion assumptions, the only difference between the structural and causal assumptions we have made here is the following: the causal analysis assumes nobody has V_1 = -1, identifies LATE (which is ATE for compliers), and identifies nothing about people who
are not compliers

- It is important to recall that these particular assumptions and models are not universal or required features of causal versus structural methods

- Another limitation of the reduced-form methodology is how it extends to more general treatments

- A related limitation of LATE is that the definition of a complier depends on the definition of the instrument Z

5.3. Randomized Causal versus Structural Simultaneous Systems

- Suppose that instead of a treatment affecting an outcome, where the direction of causality is assumed, we had a simultaneous system of equations, say

$$Y = U_0 + U_1 X \text{ and}$$
$$X = H(Y, Z, V)$$

- As before, let us again analyze the meaning of

$$c = \text{cov}(Z, Y)/\ \text{cov}(Z, X)$$

- In contrast, a causal analysis of this system is possible, but much more complex

- Another limitation of applying causal methods to simultaneous systems is that the counterfactual notation itself rules out some types of structural models

- A final limitation in applying causal analyses to simultaneous systems is the SUTVA restriction discussed earlier

5.4. Randomized Causal versus Structural Identification: Conclusions

- One great **advantage** of causal-based methods is their long history of success in the hard sciences
- Another virtue of **causal methods** is the fundamental nature of treatment effects as interpretable estimands
- Structural models can also cope with many data issues that cause difficulties for causal analyses
- Where are the deep parameters that structural models have uncovered?
- What are their widely agreed-upon values?
- The main **disadvantage** of imposing behavioral restrictions for identification is that reality is complicated, so every structural model we propose is likely to be oversimplified and hence misspecified

- One great **advantage** of causal-based methods is their long history of success in the hard sciences
- Another virtue of **causal methods** is the fundamental nature of treatment effects as interpretable estimands
- Structural models can also cope with many data issues that cause difficulties for causal analyses
- Where are the deep parameters that structural models have uncovered?
- What are their widely agreed-upon values?

- The main **disadvantage** of imposing behavioral restrictions for identification is that reality is complicated, so every structural model we propose is likely to be oversimplified and hence misspecified

- A big issue for both structural and causal models is external validity, that is, if the environment changes even slightly, how would an identified parameter or treatment effect change?

- Another limitation of causal methods is that economic policy is often concerned with **characteristics** that cannot be directly observed, like utility, risk aversion, noncognitive skills, bargaining power, expectations, or social welfare

6. Identification of Functions and Sets

- The first two subsections below discuss two modern literatures
  - nonparametric or semiparametric identification, and
  - set identification

- The third subsection below describes the role of normalizations in identification

6.1. Nonparametric and Semiparametric Identification

- In **section 3.2**, we defined ***nonparametric identification*** as the case where θ consists of functions or infinite sets

- Recall that ***parametric identification*** was defined as the case where θ is a finite set of constants, and all the different possible values of ϕ also correspond to different values of a finite set of constants

- Semiparametric identification can also be used to refer to identification of a vector of constants that are of interest in a nonparametric model

- The difference between parametric and semiparametric identification can be somewhat arbitrary

- More generally, econometric models often involve moments, which take the form of integrals

6.2. Set Identification

- Recall that we let θ denote the parameters we wish to identify, Θ is the set of all possible values of θ, and θ_1 is the unknown true value of θ

- ***Partial identification*** broadly refers to the analysis of situations where ϕ provides some information about parameters θ, but not enough information to actually point identify θ

- An important tool used for studying **set identification** is the theory of random sets

- One reason why parameters may be set rather than point identified is incompleteness of the underlying model

- Khan and Tamer (2010) define ***non-robust identification*** as the situation where an otherwise point-identified parameter loses even set identification when an identifying assumption is relaxed

6.3. Normalizations in Identification

- Nonparametric or semiparametric identification results often require so-called normalizations

- The proof is by construction:

$$\beta = [E(XX')]^{-1}E[Xg^{-1}(E(Y|X))]$$

- Is β still point identified when the function g is unknown?

- To have any chance of point identifying β, we need to impose a restriction on the model that rules out all values of $c \neq 1$

- Scale restrictions, whether they are free normalizations or not, are commonly needed to point identify semiparametric models

- Consider a threshold-crossing binary choice model, that is,

$$Y = I(\alpha + X'\beta + e \geq 0)$$

- Another common use of normalizations for identification are in non-separable error models

- Returning to the example of threshold-crossing models, when derived from utility maximization such models embody an additional normalization to location and scale

- Suppose $\alpha_y + X' \beta_y + e_y$ is the utility one receives from making the choice $Y = y$ for y equal to zero or one

- Utility maximization then means choosing

$$Y = I(\alpha + X'\beta + e \geq 0)$$

6.4. Examples: Some Special Regressor Models

- Return now to example 5, from **section 3.3**, regarding identification of a latent error distribution

- Here we consider the same model, except now we allow for the presence of additional covariates Z

- In this model

$$
\begin{aligned}
E(Y|X = x, Z = z) &= \Pr(X + U > 0 \mid X = x, Z = z) = \Pr(x + U > 0 \mid Z = z) \\
&= 1 - \Pr(U \leq -x \mid Z = z) = 1 - F_{U|Z}(-x|z)
\end{aligned}
$$

- *Example: Set Identification of the Latent Mean*
  In this example we let Z be empty and consider identification of $\theta = E(U)$

- *Example: General Binary Choice*
  Suppose we continue to have IID observations of Y, X, Z with $Y = I(X + U > 0)$ where $U \perp X|Z$, but now in addition assume that $U = g(Z) + e$ with $g(Z) = E(U|Z)$, so

$$Y = I(X + g(Z) + e > 0)$$

- *Example: Binary Choice with Random Coefficients*
  Before considering binary choice, consider first the simpler linear random coefficients model

7. Limited Forms of Identification

7.1. Local and Global Identification

- A necessary condition for **global identification**, and one that is often easier to verify in practice, is local identification

- Formally, ***local identification*** of θ_0 means that there exists a neighborhood of θ_0 such that no $\theta \in \Theta$ exists in this neighborhood that is both unequal to θ_0 and observationally equivalent to θ_0

- To illustrate the difference between local and global identification, suppose $m(x)$ is a known continuous function

- **Case 1:** Suppose we know $m(x)$ is strictly monotonic
- **Case 2:** Suppose m is known to be a Jth order polynomial for some integer J
- **Case 3:** Suppose all we know about m is that it is continuous
- Local identification may be sufficient in practice if we have enough economic intuition about the estimand to know that the correct θ should lie in a particular region

7.2. Generic Identification

- **Generic identification** is a weaker condition than point identification, is a necessary condition for point identification, and is often easier to prove than point identification

- To interpret what generic identification means, imagine that nature chooses a value θ_0 by randomly picking an element of Θ

- In models that are systems linear equations (as in Wright–Cowles identification), generic identification is closely related to the order condition for identification

- Generic identification is sometimes seen in social interactions models

- The term generic identification is sometimes used more informally to describe situations in which identification holds except in special or pathological cases, but where it might be difficult to explicitly describe all such cases

8. Identification Concepts That Affect Inference

- For the most part, identification is treated as a precursor to estimation
- In this section we summarize these identification concepts that affect inference
- However, it should be noted that some previously discussed concepts are also related to inference

8.1. Weak versus Strong Identification

- Informally, **weak identification** arises in situations that are, in a particular way, close to being not point identified
- The usual source of weak identification is low correlations among variables used to attain identification
- The key feature of weakly identified parameters is not that they are imprecisely estimated with large standard errors
- **Nonparametric regressions** are also typically imprecisely estimated, with slower than parametric convergence rates and associated large standard errors
- Weak identification resembles multicollinearity, which in a linear regression would correspond to $E(XX')$ instead of $E(\hat{X}X')$ being ill-conditioned

8.2. Identification at Infinity or Zero; Irregular and Thin Set Identification

- Based on Chamberlain (1986) and Heckman (1990), **identification** at **infinity** refers to the situation in which identification is based only on the joint distribution of data at points where one or more variables go to infinity

- Khan and Tamer (2010) and Graham and Powell (2012) use the term ***irregular identification*** to describe cases where thin set identification leads to slower than root-*n* rates of estimation

- It is easy to confuse irregular identification with weak identification, but they are not the same

- The difference is that asymptotic theory for weakly identified parameters is based on models where true parameter values are assumed to vary with the sample size, in order to obtain good approximations to the true precision with which they can be estimated in moderately sized samples

8.3. Ill-Posed Identification

- Suppose that parameters θ are point identified
- Problems of ill-posedness arise when the connection from ϕ to θ is not sufficiently
- When identification is ill posed, construction of a consistent estimator requires "regularization," that is, some way to smooth out the discontinuity in g
- Nonparametric estimation of a probability density function is an example of an ill-posedness problem
- Depending on the application, the degree of ill-posedness can range from mild to moderate to severe

8.4. Bayesian and Essential Identification

- Two more names for the same concept that appear in the literature are ***frequentist identification*** and ***sampling identification***
- These terms are used to contrast the role of identification in frequentist statistics from its role in Bayesian statistics
- A parameter vector θ is defined to be *Bayes identified* if its posterior distribution differs from its prior distribution
- Typically, a parameter that is point identified will also be Bayes identified
- Parameters that are set rather than point identified are also generally Bayes identified

9. Conclusions

- Identification is a rapidly growing area of research within econometrics, as the ever-expanding zooful of different terms for identification indicates
- Unlike statistical inference, there is not a large body of general tools or techniques that exist for proving identification
- Finally, one might draw a connection between identification and big data
- This paper has considered over two dozen different identification related concepts, as listed in the introduction
- Given the increasing recognition of its importance in econometrics, the identification zoo is likely to keep expanding