



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
Main Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2007

---

## Approximate generalizations and computational experiments

Kübler, Felix

**Abstract:** In this paper I demonstrate how one can generalize finitely many examples to statements about (infinite) classes of economic models. If there exist upper bounds on the number of connected components of one-dimensional linear subsets of the set of parameters for which a conjecture is true, one can conclude that it is correct for all parameter values in the class considered, except for a small residual set, once one has verified the conjecture for a predetermined finite set of points. I show how to apply this insight to computational experiments and spell out assumptions on the economic fundamentals that ensure that the necessary bounds on the number of connected components exist. I argue that these methods can be fruitfully utilized in applied general equilibrium analysis. I provide general assumptions on preferences and production sets that ensure that economic conjectures define sets with a bounded number of connected components. Using the theoretical results, I give an example of how one can explore qualitative and quantitative implications of general equilibrium models using computational experiments. Finally, I show how random algorithms can be used for generalizing examples in high-dimensional problems.

DOI: <https://doi.org/10.1111/j.1468-0262.2007.00779.x>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-95953>

Journal Article

Published Version

Originally published at:

Kübler, Felix (2007). Approximate generalizations and computational experiments. *Econometrica*, 75(4):967-992.

DOI: <https://doi.org/10.1111/j.1468-0262.2007.00779.x>

## APPROXIMATE GENERALIZATIONS AND COMPUTATIONAL EXPERIMENTS

BY FELIX KUBLER<sup>1</sup>

In this paper I demonstrate how one can generalize finitely many examples to statements about (infinite) classes of economic models. If there exist upper bounds on the number of connected components of one-dimensional linear subsets of the set of parameter values for which a conjecture is true, one can conclude that it is correct for all parameter values in the class considered, except for a small residual set, once one has verified the conjecture for a predetermined finite set of points. I show how to apply this insight to computational experiments and spell out assumptions on the economic fundamentals that ensure that the necessary bounds on the number of connected components exist.

I argue that these methods can be fruitfully utilized in applied general equilibrium analysis. I provide general assumptions on preferences and production sets that ensure that economic conjectures define sets with a bounded number of connected components. Using the theoretical results, I give an example of how one can explore qualitative and quantitative implications of general equilibrium models using computational experiments. Finally, I show how random algorithms can be used for generalizing examples in high-dimensional problems.

KEYWORDS: Computational economics, general equilibrium, o-minimal structures.

### 1. INTRODUCTION

COMPUTATIONAL METHODS are widely used as a tool to study economic models that do not admit closed-form solutions. An important drawback of these methods is that they seem to provide information only for the specific parameter values for which the computations have been carried out. In particular, a computation can almost never prove that a model has a given property for all parameter values. The purpose of this paper is to show that, for a very wide class of economic models, to prove that a certain conclusion holds for a set of parameters that has Lebesgue measure close to 1, it is sufficient to verify the conclusion for a sufficiently large finite set of parameters. These results rest on three ideas. (1) Economic models whose descriptions involve only functions in a particular class (which is very large and contains all utility functions and production functions commonly used in applied work) give rise to sets that have very special mathematical properties. (2) In particular, deep results in algebraic geometry provide simple mechanical procedures for bounding the number of connected components of sets of parameters for which the conclusion holds. (3) From these bounds, the volume of the set can be bounded below if the set is known to contain all the points in a particular grid.

<sup>1</sup>I thank seminar participants at various universities and conferences, and especially Don Brown, Dave Cass, Ken Judd, Narayana Kocherlakota, Mordecai Kurz, George Mailath, Marcel Richter, Klaus Ritzberger, Ilya Segal, co-editor, and four anonymous referees for helpful discussions and comments.

To describe these basic ideas a bit more formally, suppose the unknown set of parameters is a compact subset of Euclidean space  $E \subset \mathbb{R}^l$ . The economic conjecture is correct for an unknown set of parameters  $\Phi \subset E$ . Although it is not possible to use computational methods to determine that  $\Phi = E$ , it is often the case that for any given  $\bar{e} \in E$ , computational methods can determine whether  $\bar{e} \in \Phi$ . The question is under which conditions one can estimate the Lebesgue measure of  $\Phi$  from checking that  $F \subset \Phi$  for some large but finite set  $F \subset E$ . Obviously this is trivial if  $\Phi$  is known to be convex: if a collection of points is known to lie in  $\Phi$ , their convex hull must be a subset of  $\Phi$ . Although it is almost never the case that  $\Phi$  is convex, one can often bound the number of connected components of  $\Phi$ . Koiran (1995) showed that from knowing upper bounds on the number of connected components of the intersection of arbitrary axes-parallel lines and the set  $\Phi$ , one can construct lower bounds on the size of the set  $\Phi$  by verifying that the conjecture holds on a prespecified grid  $F \subset E$ . The problem of proving that conjectures hold approximately thus reduces to finding bounds on the number of connected components of the set defined by the economic statement. I will argue in Section 4 that these bounds can be obtained rather mechanically from the mathematical formulation of the conjecture.

One important complication arises from the fact that numerical methods often find only approximate solutions to economic problems and that, therefore, it is often not possible to determine whether given  $e \in E$  in fact lies in  $\Phi$ . However, Kubler and Schmedders (2005) argued that in many equilibrium problems, one can perform a backward error analysis and can infer from the computations that there exists a  $\tilde{e}$  in a small neighborhood of  $e$  that in fact lies in  $\Phi$ . To use this information to bound the volume of  $\Phi$ , I state and prove a modified version of Koiran’s result.

To describe the general method, a little more notation is needed. I assume that the set of unknown parameters  $E$  is  $[0, 1]^l$  and that the economic conjecture holds true for a Lebesgue measurable set of parameters  $\Phi \subset \mathbb{R}^l$ , which can be written in the form

$$(1) \quad \Phi = \{x_0 | Q_1 x_1 Q_2 x_2 \cdots Q_n x_n ((x_0, x_1, \dots, x_n) \in X)\},$$

where  $Q_i \in \{\exists, \forall\}$ ,  $x_i \in \mathbb{R}^{l_i}$ , and  $X$  is a finite union and intersection of sets of the form

$$\{(x_0, \dots, x_n) : g(x) > 0\} \quad \text{or} \quad \{(x_0, \dots, x_n) : f(x) = 0\}$$

for functions  $f$  and  $g$  in some specified class.

For a positive integer  $N$ , define  $F$  to be the set of evenly spaced grid points with distance  $1/N$ , i.e.,  $F = \{1/N, 2/N, \dots, 1\}^l$ . Suppose that for a given  $\varepsilon$ ,  $1/N > \varepsilon \geq 0$ , the computational experiment verifies for each  $e \in F$  that there is  $\tilde{e}$  with  $\| \tilde{e} - e \| \leq \varepsilon$  and with  $\tilde{e} \in \Phi$ .

Although in general one cannot rule out that there exist some  $e \in E$  for which the statement is false, it might often be useful to find bounds on the size of sets of variables for which the conjecture might be wrong. Theorem 1 below shows that a bound on the number of connected components of certain subsets of  $\Phi$  can be used to make a statement on the Lebesgue measure of  $\Phi$ ,  $\text{vol}(\Phi)$ . The main result of this paper is that

$$\text{vol}(\Phi) \geq 1 - \left(2\varepsilon + \frac{l}{N}\right)\lambda,$$

where  $\lambda$  is an upper bound on the number of connected components of the intersection of  $\Phi$  and any axes-parallel cylinder.

Moreover, if the conjecture cannot be verified at some grid points (either because it is false at these points or because of limitations of the numerical methods used to verify the conjecture), but if the fraction of points in  $F$  at which the conjecture can be verified is some  $\nu < 1$ , the result implies that  $\text{vol}(\Phi) \geq \nu - (2\varepsilon + \frac{l}{N})\lambda$ . In these cases, the method can still be used to estimate the volume of  $\Phi$ .

The complement of  $\Phi$  in  $[0, 1]^l$  is called the exceptional set and Theorem 1 bounds its volume from above. The method will say nothing about where this set might be located and does not give lower bounds on the size of the exceptional set; in particular, it might be empty.

The resulting conclusion is, of course, much weaker than showing that the statement is true for all elements of  $E$ , but the point is that in many applications this is just not possible. The philosophy is somewhat related to the idea underlying genericity analysis for smooth economies. There one is concerned with showing that the exceptional set has measure zero. It might very well be possible that *all* of the economically relevant specifications fall into the residual set of measure zero, but it is simply not true (or cannot be shown) that the residual set is empty. Of course, there is a huge quantitative difference between showing that the residual set has small positive measure and showing that it has zero measure; in this respect generic results are much stronger than the target results in this paper.

In Section 2, I present a simple example that illustrates the basic idea of the paper. In the following two sections, I then generalize this example along two dimensions. First, the simple example assumes that the set of unknown parameters is one dimensional. In this case, it is easy to see that the number of connected components of a set tells us something about its size if one knows sufficiently many equispaced points in the set. In several dimensions this is obviously no longer true. Instead, I show in Section 3 that it suffices to work with the number of connected components of the intersection of the set and axes-parallel lines or axes-parallel cylinders.

Second, the example is constructed so that the economic conjecture can be characterized by polynomial inequalities. Although it turns out that for many

economic problems the relevant functions are not always all polynomials, I will argue in Section 4 that they are very often Pfaffian (see, e.g., Khovanskii (1991) for definitions and motivations) and I show how the available bounds from Gabrielov and Vorobjov (2004) on the number of connected components of the set of solutions to Pfaffian equations can be used to derive upper bounds on the number of connected components of sets that are relevant for computational experiments.

In Section 5, I discuss the applicability of these methods to general equilibrium analysis. Are there general assumptions on preferences and technologies that guarantee that all interesting statements about a given class of general equilibrium models can be tackled with the methods in this paper? Are there other classes of functions, besides Pfaffians, that guarantee the required “finiteness property”? It turns out that a necessary and sufficient conditions for these methods to be applicable to a general equilibrium model is that preferences and technology are *definable in an “o-minimal” structure*, as discussed by Blume and Zame (1993) or Richter and Wong (2000). In this context, it follows from a mathematical result on o-minimal structures that it is not possible to give a complete direct characterization of the class of functions for which sets defined by (1) have finitely many connected components. In this section, I also give a more elaborate example to illustrate the potential applicability of approximate generalizations to applied equilibrium analysis.

Finally, in Section 6, I discuss the computational feasibility of the method. Even with a fixed number of connected components, it turns out that the number of examples one has to compute grows exponentially with the dimension of  $E$ . Already for medium-sized problems, the methods are, therefore, often not directly applicable. An alternative is to use a random algorithm and to make statements about the size of the set of interest that are correct with high probability (see Judd (1997) or Blum, Cucker, Shub, and Smale (1998)). Using random numbers that are generated by physical processes, one can randomly draw values for the exogenous parameters and after sufficiently many draws, for any given  $\delta$ , my results then imply bounds on the probability that the true residual set is less than  $\delta$ . These random algorithms are applicable even for problems for which known bounds on the number of connected components are relatively large, as long as they are orders of magnitude smaller than the errors in the computations.

## 2. A SIMPLE EXAMPLE

The most basic comparative statics exercise in a pure exchange economy asks what happens to equilibrium prices as individual endowments change (see, e.g., Nachbar (2002) for a general analysis of the problem). I consider a simple example of this exercise that is supposed to illustrate the three main ideas of the methods introduced in this paper: (1) economic models often give rise to sets that are defined by polynomial inequalities, (2) one can find bounds on the

number of connected components of the set of parameters for which an economic conjecture holds, and (3) these bounds imply that the set has Lebesgue measure close to 1, once one has verified that it contains all points in a finite grid.

Suppose there are two commodities and two households with endowments  $e^1, e^2$  and with constant elasticity of substitution (CES) utility functions

$$u^1(x) = -x_1^{-2} - 64x_2^{-2}, \quad u^2(x) = -64x_1^{-2} - x_2^{-2}.$$

Consider the conjecture that for all economies with individual endowments  $e^1 = (50, e), e^2 = (e, 50)$ , and  $e \in [0, 1]$ , there exist competitive equilibria for which the equilibrium price ratio of good 2 to good 1 is (locally) decreasing in  $e$ . In these economies, there is always one competitive equilibrium for which the price ratio is equal to 1. As will become clear below, the example is constructed so that for all  $e \in [0, 1]$  there are in fact three competitive equilibria, one of which exhibits a decreasing price of good 2. Suppose, however, for the sake of the example, that the only thing that is known is that for many points in  $[0, 1]$ , an algorithm finds one equilibrium at which the price is locally decreasing in endowments. This paper shows that it is possible to infer from this that the price must be decreasing in endowments for a large set of parameters  $e$ .

Normalizing the price of good 1 to one, equilibrium can be characterized by the requirement that aggregate excess demand for the first good is 0. Defining  $q$  to be the third root of the price of good 2 and multiplying out, one obtains that this is equivalent to

$$(2) \quad (8e + 25)q^3 + (-2e - 100)q^2 + (2e + 100)q - 8e - 25 = 0.$$

For the price to be decreasing in  $e$ , by the implicit function theorem, it must hold that

$$3(8e + 25)q^2 - 2(2e + 100)q + 2e + 100 \neq 0, \\ - \frac{8q^3 - 2q^2 + 2q - 8}{3(8e + 25)q^2 - 2(2e + 100)q + 2e + 100} < 0.$$

It will turn out to be useful to write this equivalently as

$$(3) \quad -(8q^3 - 2q^2 + 2q - 8)(3(8e + 25)q^2 - 2(2e + 100)q + 2e + 100) < 0.$$

The conjecture thus defines a set  $\Phi \subset E = [0, 1]$  as

$$\Phi = \{e \in [0, 1] : \exists q [(8e + 25)q^3 + (-2e - 100)q^2 + (2e + 100)q - 8e - 25 = 0 \text{ and } -(8q^3 - 2q^2 + 2q - 8)(3(8e + 25)q^2 - 2(2e + 100)q + 2e + 100) < 0]\}.$$

This paper addresses the question of whether one can bound the Lebesgue measure of this set by computing finitely many examples, i.e., by verifying  $\{0, 1/N, \dots, 1\} \subset \Phi$  for some finite integer  $N$ .

Note that although it is true that for almost any  $\bar{e} \in (0, 1)$ , if there exists a  $\bar{q}$  that satisfies (2) and (3), then it must also be true in some neighborhood of  $\bar{e}$ , there is no easy way to determine the size of this neighborhood. Therefore, it is not straightforward to use continuity arguments to generalize finitely many examples and to bound the size of the set  $\Phi$ . In fact, it is well known in numerical analysis that zeros of high-dimensional polynomials often behave extremely sensitively with respect to small changes in the coefficients (see, e.g., Wilkinson (1984) for a famous example).

The main idea of this paper is as follows. Suppose that for some reason, one can obtain an upper bound,  $\kappa$ , on the number of connected components of  $\Phi$ . Then given that in one dimension connected components must be convex, it suffices to verify that  $\bar{e} \in \Phi$  for all  $\bar{e} \in \{0, 1/N, \dots, 1\}$  to know that the Lebesgue measure of  $\Phi$  is at least  $(1 - 1/N(\kappa - 1))$ . The set for which the conjecture is wrong can at most be the union of  $\kappa - 1$  intervals of the form  $(i/N, (i + 1)/N)$ ,  $0 \leq i \leq N - 1$ . Once one knows  $\kappa$ , one can therefore verify that the conjecture is “approximately correct” by checking it at finitely many points. Furthermore, if the conjecture can be verified only at  $M$  of the  $N + 1$  points in the grid, the Lebesgue measure of  $\Phi$  can still be bounded to be  $M/(N + 1) - 1/N(\kappa - 1)$ .

Why should it be any easier to find bounds on the number of connected components of  $\Phi$  than to bound  $\Phi$  by more direct arguments? The answer lies in the fact that one can bound the number of zeros of a polynomial system of equations by simply knowing the degree of the polynomials: a univariate polynomial of degree  $d$  has at most  $d$  zeros; the classical Bézout’s theorem generalizes this to higher dimensions.

In Section 4, I will give rather mechanical recipes for bounding the number of connected components. For illustrative purposes, I now show in some detail how such a bound can be obtained in this example from the simple fact that a univariate polynomial of degree  $d$  has at most  $d$  zeros. It is also possible to apply the results from Section 4.

The first observation is that by the definition of  $\Phi$ , equilibrium prices change monotonically in  $e$  for all  $e \in \Phi$ . Therefore, the number of connected components of  $\Phi$  is bounded by 1 plus the number of real 0’s of the two equations

$$\begin{aligned} (8e + 25)q^3 - (2e + 100)q^2 + (2e + 100)q - 8e - 25 &= 0, \\ -(8q^3 - 2q^2 + 2q - 8) \\ &\times (3(8e + 25)q^2 - 2(2e + 100)q + 2e + 100) = 0. \end{aligned}$$

Moreover, by symmetry we know that for any  $e \in \Phi$  there exists an equilibrium with  $q = 1$  at which prices do not change. Therefore, we can factor  $(q - 1)$

in both of the above equations and obtain the system

$$\begin{aligned} 8eq^2 + 6eq + 8e + 25q^2 - 75q + 25 &= 0, \\ (24eq^2 - 4eq + 2e + 75q^2 - 200q + 100)(4q^2 + 3q + 4) &= 0. \end{aligned}$$

For all  $q > 0$ , we can isolate  $e$  in the first equation and substitute it into the second to obtain the equation only in  $q$ :

$$(4) \quad q^4 - 2q^3 + 2q - 1 = 0.$$

Because this equation has at most four zeros, the number of connected components of  $\Phi$  is bounded by 5. This implies that in this example, by computing equilibrium at 101 equi-spaced points and verifying that at each computed equilibrium the price is decreasing in the endowment, one can prove that the Lebesgue measure of endowments in  $[0, 1]$  for which this must be true is no smaller than 0.96.

Now what happens if one can only approximate the solution to Equation (2), in the sense that one finds a  $\tilde{q}$  for which aggregate excess demand is approximately equal to 0, i.e., for which

$$|(8e + 25)\tilde{q}^3 + (-2e - 100)\tilde{q}^2 + (2e + 100)\tilde{q} - 8e - 25| = \varepsilon$$

for some small  $\varepsilon > 0$ . Although one cannot, in general claim that there exists a true equilibrium close to  $\tilde{q}$ , one can claim that  $\tilde{q}$  is an exact equilibrium for some  $\tilde{e}$  close to  $e$ . In fact,

$$\tilde{e} = \frac{25\tilde{q}^3 - 100\tilde{q}^2 + 100\tilde{q} - 25 \pm \varepsilon}{8 - 100\tilde{q} + 100\tilde{q}^2 - 8\tilde{q}^3}.$$

Given  $\tilde{q}$ , it is straightforward to compute bounds on  $|\tilde{e} - e|$ . Therefore, even if equilibrium cannot be computed exactly, one can use computational methods to verify that there are  $e_0, \dots, e_N$  with  $e_i \in \Phi$  and  $\|e_i - i/N\| < \delta$  for some small  $\delta$ ,  $i = 0, \dots, N$ . This suffices to apply the method above and to bound the volume of  $\Phi$ . It is easy to see that the argument goes through as before with the only modification being that now there can be four intervals of the form  $(i/N - \delta, (i + 1)/N + \delta)$  that might not be subsets of  $\Phi$ . Therefore, for  $N = 100$ , the lower bound on the volume of  $\Phi$  is now  $0.96 - 8\delta$ .

### 3. CONNECTED COMPONENTS IN SEVERAL DIMENSIONS

The goal is to give good lower bounds on the size (Lebesgue measure) of  $\Phi$  as defined by Equation (1) in the [Introduction](#). In this section, I consider an arbitrary (Lebesgue measurable) set  $\Phi \subset \mathbb{R}^l$  and assume that for some reason one can obtain bounds on the number of connected components of the intersection of this set and arbitrary axes-parallel lines in  $\mathbb{R}^l$ . I will then show in

Section 4 how these bounds arise from Equation (1) if one limits the functions that define  $\Phi$  to be of a particular class.

Throughout, fix  $\|\cdot\|$  to denote the 2-norm. Define a generalized indicator function  $\mathfrak{S}^\varepsilon(x)$  to be 1 if there is a  $y \in \Phi$  with  $\|y - x\| \leq \varepsilon$  and 0 otherwise. For  $\varepsilon = 0$ , this is the simple indicator function and the Lebesgue measure of  $\Phi$  is given by  $\int_{[0,1]^l} \mathfrak{S}^0(x) dx$ .

For  $x \in F$ , define a cylinder of radius  $\varepsilon$  centered around  $(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_l)$  by

$$C_{-i}^\varepsilon(x) = \{y \in \mathbb{R}^l : \|y_i - x_i\| \leq \varepsilon \text{ for } i \neq \bar{i}\}.$$

Note that  $C_{-i}^0(x)$  is simply a line parallel to the  $x_i$  axis passing through the point  $x$ . For a set  $A$ , denote by  $\kappa(A)$ , the number of its connected components. The following lemma generalizes Lemma 2 in Koiran (1995).

LEMMA 1: Given  $\bar{x} \in F$ , define  $Q = C_{-1}^\varepsilon(\bar{x}) \cap \Phi$ . Then

$$\left| \int_0^1 \mathfrak{S}^\varepsilon(y, \bar{x}_2, \dots, \bar{x}_l) dy - \frac{1}{N} \sum_{i=1}^N \mathfrak{S}^\varepsilon\left(\frac{i}{N}, \bar{x}_2, \dots, \bar{x}_l\right) \right| \leq \frac{\kappa(Q)}{N}.$$

PROOF: The number of connected components of the set of  $x \in C_{-1}^0$  for which  $\mathfrak{S}^\varepsilon(x) = 1$  is not larger than the number of connected components of  $\Phi \cap C_{-1}^\varepsilon$ . Therefore, it can be written as the union of  $K$  disjoint connected pieces with  $K \leq \kappa$ , i.e., there exist  $a_1 < b_1 < \dots < a_K < b_K$  such that

$$\{x \in C_{-1}^0 : \mathfrak{S}^\varepsilon(x) = 1\} = \bigcup_{k=1}^K \{x \in C_{-1}^0, x_1 \in [a_k, b_k]\}.$$

Then

$$\begin{aligned} & \left| \frac{1}{N} \sum_{i=1}^N \mathfrak{S}^\varepsilon\left(\frac{i}{N}, \bar{x}_2, \dots, \bar{x}_l\right) - \int_0^1 \mathfrak{S}^\varepsilon(y, \bar{x}_2, \dots, \bar{x}_l) dy \right| \\ & \leq \sum_{k=1}^K \left| \int_{a_k}^{b_k} \mathfrak{S}^\varepsilon(y, \bar{x}_2, \dots, \bar{x}_l) dy - \frac{1}{N} \sum_{i: a_k \leq i/N \leq b_k} \mathfrak{S}^\varepsilon\left(\frac{i}{N}, \bar{x}_2, \dots, \bar{x}_l\right) \right|. \end{aligned}$$

The definition of  $\mathfrak{S}^\varepsilon$  implies that for all  $k$ ,

$$\left| \int_{a_k}^{b_k} \mathfrak{S}^\varepsilon(y, \bar{x}_2, \dots, \bar{x}_l) dy - \frac{1}{N} \sum_{i: a_k \leq i/N \leq b_k} \mathfrak{S}^\varepsilon\left(\frac{i}{N}, \bar{x}_2, \dots, \bar{x}_l\right) \right| \leq \frac{1}{N}.$$

The result follows directly from this by adding up the  $K$  pieces. Q.E.D.

This lemma is now extended to several dimensions. The underlying idea is to bound the number of connected components of the intersection of  $\Phi$  and any axes-parallel cylinder. For this, define  $\lambda$  to be the maximal number of connected components across all intersections of  $\Phi$  with all possible cylinders  $C^\varepsilon$ , i.e.,

$$(5) \quad \lambda = \sup_{i=1, \dots, l; x \in [0,1]^l} \kappa(C_{-i}^\varepsilon(x) \cap \Phi).$$

In Section 5.2 below, I will characterize the sets for which  $\lambda < \infty$ . Note that it does not suffice to consider bounds on the number of connected components only across lines that connect grid points. As will become clear subsequently, the method crucially rests on the existence of a uniform bound across all possible cylinders.

The following theorem is the main tool for the analysis in this paper.

**THEOREM 1:** *Given a bound on connected components  $\lambda$ , one can estimate the size of  $\Phi$  by verifying that the grid  $F = \{1, \dots, N\}^l \subset \Phi$  as*

$$(6) \quad \left| \frac{1}{N^l} \sum_{i_1, \dots, i_l}^N \mathfrak{S}^\varepsilon\left(\frac{i_1}{N}, \dots, \frac{i_l}{N}\right) - \int_{[0,1]^l} \mathfrak{S}^0(x) dx \right| \leq \left(\frac{l}{N} + 2\varepsilon\right)\lambda.$$

**PROOF:** The theorem is proved by induction. For  $l = 1$ , one only needs to modify the last step of the proof of Lemma 1 to obtain

$$\left| \frac{1}{N} \sum_{i=1}^N \mathfrak{S}^\varepsilon\left(\frac{i}{N}\right) - \int_{[0,1]} \mathfrak{S}^0(x) dx \right| \leq \lambda \left(\frac{1}{N} + 2\varepsilon\right).$$

For  $l > 1$ , the induction goes as follows. Adding and subtracting the term  $\int_{[0,1]} \frac{1}{N^{l-1}} \sum_{i_2, \dots, i_l}^N \mathfrak{S}^\varepsilon(x_1, i_2/N, \dots, i_l/N) dx_1$  to the left-hand side of Equation (6), one obtains

$$\begin{aligned} & \left| \frac{1}{N^l} \sum_{i_1, \dots, i_l}^N \mathfrak{S}^\varepsilon\left(\frac{i_1}{N}, \dots, \frac{i_l}{N}\right) - \int_{[0,1]^l} \mathfrak{S}^0(x) dx \right| \\ & \leq \left| \int_{[0,1]} \frac{1}{N^{l-1}} \sum_{i_2, \dots, i_l}^N \mathfrak{S}^\varepsilon\left(x_1, \frac{i_2}{N}, \dots, \frac{i_l}{N}\right) dx_1 - \int_{[0,1]^l} \mathfrak{S}^0(x) dx \right| \\ & \quad + \left| \frac{1}{N^l} \sum_{i_1, \dots, i_l}^N \mathfrak{S}^\varepsilon\left(\frac{i_1}{N}, \dots, \frac{i_l}{N}\right) \right. \\ & \quad \left. - \int_{[0,1]} \frac{1}{N^{l-1}} \sum_{i_2, \dots, i_l}^N \mathfrak{S}^\varepsilon\left(x_1, \frac{i_2}{N}, \dots, \frac{i_l}{N}\right) dx_1 \right|. \end{aligned}$$

Assuming that (6) holds for  $l - 1$ , one obtains that for all  $x_1 \in [0, 1]$ ,

$$\left| \frac{1}{N^{l-1}} \sum_{i_2, \dots, i_l}^N \mathfrak{S}^\varepsilon \left( x_1, \frac{i_2}{N}, \dots, \frac{i_l}{N} \right) - \int_{[0,1]^{l-1}} \mathfrak{S}^0(x_1, \tilde{x}) d\tilde{x} \right| \leq \lambda \left( 2\varepsilon + \frac{l-1}{N} \right).$$

By Lemma 1,

$$\left| \frac{1}{N^l} \sum_{i_1, \dots, i_l}^N \mathfrak{S}^\varepsilon \left( \frac{i_1}{N}, \dots, \frac{i_l}{N} \right) - \int_{[0,1]} \frac{1}{N^{l-1}} \sum_{i_2, \dots, i_l}^N \mathfrak{S}^\varepsilon \left( x_1, \frac{i_2}{N}, \dots, \frac{i_l}{N} \right) dx_1 \right| \leq \frac{\lambda}{N}.$$

The result then follows by integrating the first term over  $[0, 1]$  and adding the result to the second expression. *Q.E.D.*

Koiran (1995) considered the (important) special case  $\varepsilon = 0$ . With bounds on the number of connected components of the intersection of  $\Phi$  with axes-parallel lines, this provides a method for bounding the measure of  $\Phi$ . In practice, these bounds are often orders of magnitude better than bounds on connected components of the intersection with general cylinders  $C^\varepsilon$ . However, these bounds are applicable only in cases where the economic model can be solved exactly at the prespecified points in  $F$ ; I give an example in Section 5. It is unclear, under which conditions the bounds in the theorem are tight and whether the choice of the grid points is optimal. In particular, the question of whether one can find locations of points in higher dimension that do not require the number of points to grow at the exponential rate of Equation (6) is subject to further research.

Figure 1 illustrates the basic idea behind the theorem. For simplicity, consider the case  $\varepsilon = 0$  and  $l = 2$ . To make the idea more transparent, it is useful to assume that the grid is in fact  $\{0, 1/N, \dots, 1\}^2$ , i.e., includes points with one coordinate being 0.<sup>2</sup> Suppose  $N = 3$ , i.e., the conjecture can be verified on a grid of  $4 \times 4$  points in  $\mathbb{R}^2$  (the black dots in the figure) and suppose  $\lambda = 2$ . Clearly, along each horizontal line that connects grid points, there are at most two points that are not connected. The upper part of the figure depicts a generic example. No matter where the exceptional set is located, it is either the case that  $2/3$  of arbitrary vertical lines cut at least three horizontal lines (as in the figure) or that  $1/3$  of vertical lines cut four horizontal lines. The fact that

<sup>2</sup>This makes the formal proof more complicated, but helps make this specific example understandable.

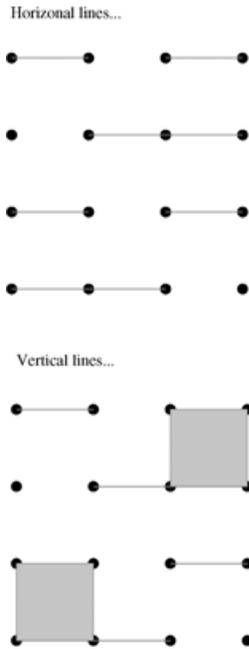


FIGURE 1.—An illustration of Theorem 1.

the number of connected components along each vertical line is at most 2 now implies that there must be a set of Lebesgue measure not smaller than  $2/9$  for which the conjecture holds. This is the crucial step of the argument: It seems that just from the knowledge that the conjecture holds at the grid points one can say nothing about an arbitrary vertical line that does not pass through any grid points. However, given that any horizontal line that passes through the grid points has at most one “opening” (i.e., one interval of length  $1/3$  that does not lie in  $\Phi$ ), it follows that a large fraction of vertical lines must pass through  $\Phi$ . The lower part of the figure depicts the “worst-case scenario” where the measure is, in fact, equal to  $2/9$ .

Although in one dimension, there is a clear relationship between convexity of a set and the set consisting of only one connected component, this is no longer true in higher dimensions. The theorem and the example show that the correct generalizations in higher dimensions consider the number of connected components along arbitrary axes-parallel lines.

In applying these method, the “only” challenge is to find reasonable bounds on  $\lambda$ . It turns out that computational experiments in economics usually consider very specific mathematical environments, for which it is easy to obtain bounds.

#### 4. BOUNDING THE NUMBER OF CONNECTED COMPONENTS IN ECONOMIC APPLICATIONS

So far, it has been assumed that bounds on the number of connected components exist and can be computed relatively easily from (1). Of course, there are many functions  $f, g$  for which the number of connected components of a set defined as in (1) might be infinite (consider, for example, the set  $\{x \in (0, 1) : \sin(1/x) = 0\}$ ) or for which it is not easily possible to compute bounds on the number of connected components.

However, in many economic application, the functions  $f$  and  $g$  in Equation (1) can be written as so-called Pfaffian functions. These are classes of functions for which it can be shown that  $\Phi$  has finitely many connected components. In fact, there is a fairly large literature in mathematics now that considers the problem of finding reasonable bounds on the number of connected components of sets defined by Pfaffian functions (see, e.g., Gabrielov and Vorobjov (2004) for an overview).

##### 4.1. Pfaffian Functions

The following definition is from Khovanskii (1991), who showed that these functions maintain many of the finiteness properties of polynomials.

**DEFINITION 1:** A Pfaffian chain of order  $r \geq 0$  and degree  $\alpha \geq 1$  in an open domain  $G \subset \mathbb{R}^n$  is a sequence of analytic functions  $f_1, \dots, f_r$  on  $G$  that satisfy differential equations

$$df_j(x) = \sum_{1 \leq i \leq n} g_{ij}(x, f_1(x), \dots, f_j(x)) dx_i$$

for  $1 \leq j \leq r$ . The  $g_{ij}$  are polynomial in  $x = (x_1, \dots, x_n), y_1, \dots, y_j$  of degree not exceeding  $\alpha$ . A function  $f(x) = p(x, f_1(x), \dots, f_r(x))$ , with  $p$  being a polynomial of degree  $\beta$ , is called a Pfaffian function of order  $r$  and degree  $(\alpha, \beta)$ .

Polynomials are included in this definition as Pfaffian functions of order 0. The following simple facts about Pfaffian functions are easy to verify.

- The expression  $\exp(x)$  is a Pfaffian function of order 1 and degree  $(1, 1)$  in  $\mathbb{R}$ ;  $f(x) = \log(x)$  is a Pfaffian function of order 2 and degree  $(2, 1)$  on  $\mathbb{R}_{++}$  because  $f'(x) = 1/x$  and  $f''(x) = -(f'(x))^2$ . Similarly,  $f(x) = x^\alpha$  is a Pfaffian function of order 2 because  $f'(x) = \alpha 1/x f(x)$ .
- Given two Pfaffian functions of order  $r$  with the same underlying chain and degrees  $(\alpha_1, \beta_1)$  and  $(\alpha_2, \beta_2)$ , respectively, the sum is a Pfaffian function of order  $r$  and degree  $(\max(\alpha_1, \alpha_2), \max(\beta_1, \beta_2))$ . The product of the two functions is Pfaffian of order  $r$  and degree  $(\max(\alpha_1, \alpha_2), \beta_1 + \beta_2)$ .
- A partial derivative of a Pfaffian function of order  $r$  and degree  $(\alpha, \beta)$  is a Pfaffian function with the same Pfaffian chain of order  $r$  and degree  $(\alpha, \alpha + \beta - 1)$ .

These facts show that all commonly used utility functions and production functions (e.g., CES), as well as first order conditions for agents' optimality, can be written as Pfaffian functions.

#### 4.2. Bounds

Gabriellov, Vorobjov, and Zell (2003) showed how to compute bounds on the number of connected components of sets of the form (1). However, in their general approach it is often rather difficult to obtain good bounds because they, in fact, bound the sum of all Betti numbers  $\Phi$ , while the number of connected components equals the zeroth Betti number. It is therefore useful to consider the special case of Equation (1), which is often relevant in economic problems, i.e.,

$$(7) \quad \Phi = \{x_0 | \exists x_1 (g(x_0, x_1) > 0 \text{ and } f(x_0, x_1) = 0)\},$$

where  $x_0 \in \mathbb{R}^{l_0}$ ,  $x_1 \in \mathbb{R}^{l_1}$ ,  $f : \mathbb{R}^{l_0+l_1} \rightarrow \mathbb{R}^{J_1}$ , and  $g : \mathbb{R}^{l_0+l_1} \rightarrow \mathbb{R}^{J_2}$ , are Pfaffian functions.

Because projection is continuous, the number of connected components of  $\Phi$  is bounded by the number of connected components of

$$\{(x_0, x_1) : g(x_0, x_1) > 0 \text{ and } f(x_0, x_1) = 0\}.$$

Strict inequalities can be turned into equalities in the following way. Given  $J_1$  inequalities  $g_1(x) > 0, \dots, g_{J_1}(x) > 0$  and a system of equations  $f(x) = 0$ , the number of connected components of

$$\{x : f(x) = 0 \text{ and } g_1(x) > 0 \text{ and } \dots \text{ and } g_{J_1}(x) > 0\}$$

is bounded by the number of connected components of

$$\{x : f(x) = 0 \text{ and } g_1(x) \neq 0 \text{ and } \dots \text{ and } g_{J_1}(x) \neq 0\},$$

which is bounded by the number of connected components of

$$\left\{ (x, \gamma) : f(x) = 0 \text{ and } 1 - \gamma \prod_{j=1}^{J_1} g_j(x) = 0 \right\}.$$

Given these results, it is interesting to obtain bounds on the number of connected components of sets of the form

$$S = \{x : f(x) = 0\} \subset \mathbb{R}^l, \quad f : \mathbb{R}^l \rightarrow \mathbb{R}^n.$$

Suppose all  $f_i, 1 \leq i \leq n$ , are Pfaffian functions on a domain  $G \subset \mathbb{R}^l$ , with either  $G = \mathbb{R}^l$  or  $G = \mathbb{R}^l_{++}$ , having common Pfaffian chain of order  $r$  and de-

gresses  $(\alpha, \beta_i)$ , respectively. Let  $\beta = \max_i \beta_i$ . Then the number of connected components of  $\{x : f_1(x) = \dots = f_n(x) = 0\}$  does not exceed

$$(8) \quad 2^{(r(r-1))/2+1} \beta (\alpha + 2\beta - 1)^{l-1} ((2l - 1)(\alpha + \beta) - 2l + 2)^r.$$

This bound is from Gabrielov and Vorobjov (2004). It grows exponentially fast in the length of the underlying Pfaffian chain and in the dimension. There is a large gap between these upper bounds and known lower bounds for the general case, but these are, to the best of my knowledge, the best currently known bounds.

Much better bounds are known for the special case where all  $f_i$  are polynomials (i.e.,  $r = 0$ ). In many economic applications, it is often sufficient to consider polynomials and it is, therefore, of practical importance to have good bounds for this case.

The following bound is from Rojas (2000). Suppose  $f_1, \dots, f_n$  are polynomial and  $G = \mathbb{R}^l$ . Consider the convex hull of the union of the  $l$  unit vectors in  $\mathbb{R}^l$  together with the origin and the exponents of all monomials in the equalities that define  $S$  (i.e., for the monomial  $x_1^{\alpha_1} \dots x_l^{\alpha_l}$  one would take the vector  $\alpha \in \mathbb{R}^l$ ). For a set  $Q \subset \mathbb{R}^l$ , denote by  $\text{vol}(Q)$  the  $l$ -dimensional volume that is standardized to obtain volume 1 for the  $l$ -dimensional simplex. Then the number of connected components of  $S$ ,  $\kappa(S)$ , can be bounded as

$$(9) \quad \kappa(S) \leq 2^{l-1} \text{vol}(Q).$$

## 5. COMPUTATIONAL EXPERIMENTS IN GENERAL EQUILIBRIUM ANALYSIS

In applied general equilibrium analysis, numerical methods are routinely used to investigate quantitative features of general equilibrium models. It is therefore interesting to investigate the extent to which the methods in this paper can contribute to the current state of the art in this field.

I first discuss the extent to which computational methods can be used to verify a conjecture for a given specification of a general equilibrium model, taking into account computational errors. I then describe conditions on the fundamentals of the economy that ensure that the methods of this paper are applicable to general equilibrium models. Finally, I give an application of the methods to an example from the literature.

### 5.1. Approximate Competitive Equilibrium

It is useful to discuss in some detail one special case of (1). I assume that the economic statement of interest for a given specification of exogenous variables,  $e$ , can be written as

$$(10) \quad \exists (x_1, \dots, x_k) \in \mathbb{R}^n, \quad h_1(x_1, e) = 0,$$

$$\begin{aligned} & \vdots \\ & h_k(x_k, e) = 0, \\ & \psi(x_1, \dots, x_k, e) > 0. \end{aligned}$$

For each  $i = 1, \dots, k$ , the (possibly multivariate) function  $h_i$  should be understood to summarize the equilibrium conditions for a given economy, i.e., they consist of necessary and sufficient first order conditions together with market clearing or consist simply of the aggregate excess demand functions. The vector  $x_i$  is supposed to contain all endogenous variables (e.g., allocations and prices) for this economy. Different  $h_i$  correspond to different specifications of the economy, for example,  $h_1$  could summarize the equilibrium conditions for an economy without taxes, while  $h_2$  could indicate that some taxes are introduced to the economy. The function  $\psi$  makes the comparative statics comparisons that are of interest in the particular application.

In Section 5.3, I give a concrete example from applied general equilibrium analysis. For now, I want to focus on the abstract mathematical problem. To determine whether, for a given  $\bar{e} \in E$ , the statement is true, one now has to compute a solution to the nonlinear system of equations.

Most existing algorithms used in practice find only one of possibly several solutions; often the algorithms find only an approximate solution. The fact that the algorithms find only one solution limits the economic statements one can consider (for example, one can generally not make statement about all solutions), but is irrelevant for the approximate generalizations suggested in this paper.

In most cases, only approximate solutions can be obtained. As an example, consider a pure exchange economy with  $I$  agents that have endowments and utility functions  $\mathcal{E} = (e^i, u^i)_{i=1}^I$ . For the associated aggregate excess demand function  $z_{\mathcal{E}}(\cdot)$ , Scarf's (1967) algorithm finds, for any given  $\varepsilon > 0$ , a  $\tilde{p}$  such that  $\|z_{\mathcal{E}}(\tilde{p})\| < \varepsilon$ . Evidently,  $\tilde{p}$  might not be a good approximation for an exact equilibrium price. However, if  $z_{\mathcal{E}}(\cdot)$  is the aggregate excess demand function for a given profile of individual endowments  $e^1, \dots, e^I$  and if  $e^1$  is sufficiently large, it follows from Walras' law that  $z_{(e^1+z_{\mathcal{E}}(\tilde{p}), e^2, \dots, e^I, (u^i))}(\tilde{p}) = 0$ . Given that  $\tilde{p} \cdot z_{\mathcal{E}}(\tilde{p}) = 0$ , adding  $z_{\mathcal{E}}(\tilde{p})$  to an agent's endowments does not change his individual demand, but only his excess demand. In other words,  $\tilde{p}$  is the exact equilibrium price for a close-by economy. This observation, which has been known at least since Postlewaite and Schmeidler (1981) (and follows from Debreu (1970)), has been tied to "backward error analysis" used in numerical analysis by Kubler and Schmedders (2005). They showed that the idea is applicable to a wide variety of general equilibrium models, including models with production, uncertainty, and, possibly, incomplete financial markets.

The fact that only approximate solutions can be obtained then means that the computational experiment can only determine an  $\varepsilon > 0$  such that there exists an  $\tilde{e}$  with  $\|\tilde{e} - \bar{e}\| \leq \varepsilon$  and  $\tilde{e} \in \Phi$ .

In some cases, if the functions  $h_i$  are all polynomials, one can apply Smale’s so-called alpha method to bound the difference between true equilibrium prices and allocations, and computed prices and allocations.

5.1.1. *Smale’s alpha method*

Because it is not very well known in economics, Smale’s method is summarized for completeness. The following results are from Blum, Cucker, Shub, and Smale (1998, Chapter 8).

Let  $D \subset \mathbb{R}^n$  be open and let  $f : D \rightarrow \mathbb{R}^n$  be analytic. For  $z \in D$ , define  $f^{(k)}(z)$  to be the  $k$ th derivative of  $f$  at  $z$ . This is a multilinear operator that maps  $k$ -tuples of vectors in  $D$  into  $\mathbb{R}^n$ . Define the norm of an operator  $A$  to be

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}.$$

Suppose that  $f^{(1)}(z)$ , the Jacobian of  $f$  at  $z$ , is invertible and define

$$\gamma(z) = \sup_{k \geq 2} \left\| \frac{(f^{(1)}(z))^{-1} f^{(k)}(z)}{k!} \right\|^{1/(k-1)}$$

and

$$\beta(z) = \|(f^{(1)}(z))^{-1} f(z)\|.$$

**THEOREM 2:** *Given a  $\bar{z} \in D$ , suppose the ball of radius  $(1 - \frac{\sqrt{2}}{2})/\gamma(\bar{z})$  around  $\bar{z}$  is contained in  $D$  and that*

$$\beta(\bar{z})\gamma(\bar{z}) < 0.157.$$

*Then there exists a  $\tilde{z} \in D$  with*

$$f(\tilde{z}) = 0 \quad \text{and} \quad \|\bar{z} - \tilde{z}\| \leq 2\beta(\bar{z}).$$

Note that the result holds for general analytic functions. However, it is only applicable for polynomials, because, for general analytic functions, it is difficult or impossible to obtain bounds on  $\sup_{k \geq 2} \|((f^{(1)}(z))^{-1} f^{(k)}(z))/k!\|^{1/(k-1)}$ . In Section 5.3, I give a trivial example where the method is applicable.

5.2. *Generalizable Economies*

In general equilibrium analysis, the following question arises naturally: What assumptions on fundamentals guarantee that there are bounds on the number of connected components of sets defined as in Equation (1)?

If the economic conjecture can be generalized from finitely many examples to a set of large volume, I say that the economic model allows for approximate generalizations. Although it is true that CES utility and production functions are Pfaffian, one would ideally hope for assumptions on preferences and technologies that are sufficient for approximate generalizations and that are a bit more general than assuming Pfaffians. Furthermore, the question arises whether there are necessary conditions on preferences and technologies that have to hold so that the techniques in this paper are applicable and the economy allows for approximate generalizations.

One possible characterization, which will turn out to be both necessary and sufficient, is that the underlying classes of economies are *definable in an o-minimal structure*. I give a brief explanation of what this means and then discuss its implications.

5.2.1. *o-minimal structures*

The following definitions are from van den Dries (1999). Define a structure on  $\mathbb{R}$  to be a sequence  $\mathcal{S} = (\mathcal{S}_m)_{m \in \mathbb{N}}$  such that, for each  $m \geq 1$ , the following statements hold:

- (S1)  $\mathcal{S}_m$  is a Boolean algebra of subsets of  $\mathbb{R}^m$ .
- (S2) If  $A \in \mathcal{S}_m$ , then  $\mathbb{R} \times A$  and  $A \times \mathbb{R}$  belong to  $\mathcal{S}_{m+1}$ .
- (S3) There exists  $\{(x_1, \dots, x_m) \in \mathbb{R}^m : x_1 = x_m\} \in \mathcal{S}_m$ .
- (S4) If  $A \in \mathcal{S}_{m+1}$ , then  $\pi(A) \in \mathcal{S}_m$ , where  $\pi: \mathbb{R}^{m+1} \rightarrow \mathbb{R}^m$  is the projection map on the first  $m$  coordinates.

A set  $A \subset \mathbb{R}^m$  is said to be definable in  $\mathcal{S}$  if it belongs to  $\mathcal{S}_m$ . A function  $f: \mathbb{R}^m \rightarrow \mathbb{R}^n$  is said to be definable in  $\mathcal{S}$  if its graph belongs to  $\mathcal{S}_{m+n}$ .

An o-minimal structure on  $\mathbb{R}$  is a structure such that the following statements hold:

- (O1) There exists  $\{(x, y) \in \mathbb{R}^2 : x < y\} \in \mathcal{S}_2$ .
- (O2) The sets in  $\mathcal{S}_1$  are exactly the finite unions of intervals and points.

It can be easily verified that a set  $\Phi$  as defined in (1) belongs to an o-minimal structure  $\mathcal{S}$  if all functions  $f, g$  are definable in  $\mathcal{S}$ .

It is beyond the scope of this paper to discuss the assumption of o-minimality in detail. Theorem 2 below makes it clear that this assumption is very useful for the analysis. For a thorough reference on o-minimal structures, see van den Dries (1998). A well-known example of an o-minimal structure is induced by the ordered field of real numbers; definable sets are the semialgebraic sets. In formulation (1), the functions  $f$  and  $g$  are then all polynomials. Wilkie (1996) proved that the structure generated by Pfaffian functions is also o-minimal.

The following two theorems are important for our analysis. The first theorem is a standard result for o-minimal structures (see, e.g., van den Dries (1999)).

**THEOREM 3:** *Let  $\Phi \subset \mathbb{R}^l$  be a definable set in an o-minimal structure on  $\mathbb{R}$ . There is a uniform bound  $B$  such that for any affine set  $L \subset \mathbb{R}^l$ , the set  $\Phi \cap L$  has at most  $B$  connected components.*

A set is affine if it can be defined by a system of linear equations.

The fact that the uniform bounds exist is interesting from a theoretical perspective, because it implies that for any set definable in an o-minimal structure,  $\lambda$  as defined in Equation (5) is finite. In practice, however, how to obtain actual bounds when the sets cannot be described by either polynomial or Pfaffian functions is an open question.

The second result follows from the cell-decomposition theorem (see van den Dries (1999) for a statement and proof of the cell-decomposition theorem).

**THEOREM 4:** *If  $\mathcal{S}$  is an o-minimal structure on  $\mathbb{R}$ , all definable sets are Lebesgue measurable.*

### 5.2.2. o-minimal economies

Given an o-minimal structure  $\mathcal{S}$ , preferences  $\succeq$  over consumption bundles in some definable set  $X$  are called definable if all better sets are definable, i.e., for all  $x \in X$ ,  $\{y: y \succeq x\}$  is definable in  $\mathcal{S}$ . Richter and Wong (2000) proved that definable preferences can be represented by definable utility functions. It is easy to see that definable utility functions give rise to definable best response correspondences and that in pure exchange economies the equilibrium manifold is definable if preferences are definable. Blume and Zame (1993) applied o-minimality to consumer theory and general equilibrium analysis, and proved a definable analogue of Debreu's theorem on generic local uniqueness.

Both Blume and Zame and Richter and Wong argued that the assumption that preferences and technologies are definable in an o-minimal structure is very natural and satisfied in almost all (finite) applied general equilibrium models.

Given a class of o-minimal economies, any statement that gives rise to a definable set  $\Phi$  can be approximately generalized using the methods in this paper. As mentioned above, a set  $\Phi$  defined by (1) is definable in an o-minimal structure  $\mathcal{S}$  if the functions  $f$  and  $g$  are definable. Moreover, any first order sentence about definable economies defines a set  $\Phi$  that admits bounds on the number of connected components.

It is clear that the assumption of o-minimality of the classes of economies considered is both necessary and sufficient for the applicability of approximate generalizations. Theorem 2 shows sufficiency; necessity follows directly from the condition (O2). If an economy is not o-minimal, there exist sets with infinitely many connected components.

### 5.2.3. A complete characterization?

Although o-minimality is necessary and sufficient for approximate generalizations, the assumption is a bit unsatisfactory in that it only provides an indirect characterization of preferences and technologies. This leads to the question whether one can derive a largest class of utility and production functions

that guarantee that the underlying economy is definable in an o-minimal structure.

Surprisingly, one can show that this is impossible. Rolin, Speissegger, and Wilkie (2003) constructed a pair of distinct o-minimal structures on the reals that are not both reducts of a common o-minimal expansion. This result implies that there cannot be one largest class of utility and production functions that gives rise to o-minimal economies. Instead, the assumption on o-minimality is an assumption on the entire economy. If some agents' preferences are definable in one o-minimal structure while others are definable in another o-minimal structure, it is not guaranteed that there exists a larger structure that is still o-minimal and in which all preferences are definable.

In this sense, o-minimality provides the best characterization of finiteness one can hope for.

### 5.3. An Example

I reexamined a well-known example from applied general equilibrium analysis. Following Shoven (1976), I asked about the output effects of capital taxation in the two-sector model of the U.S. economy. The example is intended to give an illustration of the methods and is, therefore, held as simple as possible.

In a static economy, two consumption goods are produced by two sectors,  $j = 1, 2$ , using as input capital and labor. Production functions are Cobb–Douglas and of the form

$$\begin{aligned} f_1(y_{1l}, y_{1k}) &= \gamma_1 y_{1l}^{2/3} y_{1k}^{1/3}, \\ f_2(y_{2l}, y_{2k}) &= \gamma_2 y_{2l}^{1/2} y_{2k}^{1/2}. \end{aligned}$$

Two individuals,  $i = 1, 2$ , are endowed with capital and labor  $(k_i, l_i)$ . Let  $K = k_1 + k_2$  and  $L = l_1 + l_2$  denote aggregate endowments. Utilities are Cobb–Douglas,  $u^i(x_1, x_2) = x_1^{\xi_i} x_2^{1-\xi_i}$ ,  $0 < \xi_i < 1$ . Prices are  $(p_1, p_2, p_k, p_l)$  and I normalize  $p_l = 1$  throughout.

In the benchmark equilibrium, there is a tax on capital in sector 1 and the revenue  $T = \tau p_k y_{1k}$  is distributed equally among the two agents. Following Shoven (1976), I assume that in the benchmark, equilibrium prices, as well as total output per sector are observable. The economic conjecture is that removal of this tax will increase total output, measured at new equilibrium prices, by at least 5 percent. I want to illustrate how to use the methods in this paper to prove that the conjecture is true for a large set of exogenous parameters for which the model is consistent with the benchmark equilibrium. The following five steps are necessary:

1. Formulate the economic statement as a system of equations and inequalities.
2. Identify the set of exogenous parameters that one wants to consider. This might involve adding additional constraints on parameters to match quantities in the benchmark equilibrium.

3. Formulate a system of equations that defines the set  $\Phi$  and that are Pfaffian or polynomial.
4. Perform the computations on a grid of parameter values. For each parameter on the grid, either prove that it is in  $\Phi$  (e.g., if Smale's method is applicable) or show that there are close-by parameters that lie in  $\Phi$ .
5. Find bounds on the number of connected components and apply Theorem 1.

In step 1, note that competitive equilibrium is characterized by the following market clearing and firms' optimality conditions.

$$(11) \quad \xi_1 \frac{p_k k_1 + l_1 + T/2}{p_1} + \xi_2 \frac{p_k k_2 + l_2 + T/2}{p_1} = \gamma_1 y_{1l}^{2/3} y_{1k}^{1/3},$$

$$(12) \quad (1 - \xi_1) \frac{p_k k_1 + l_1 + T/2}{p_2} + (1 - \xi_2) \frac{p_k k_2 + l_2 + T/2}{p_2} = \gamma_2 y_{2l}^{1/2} y_{2k}^{1/2},$$

$$(13) \quad y_{1k} + y_{2k} = k_1 + k_2,$$

$$(14) \quad y_{1l} + y_{2l} = l_1 + l_2,$$

$$(15) \quad p_1 \frac{2\gamma_1}{3} y_{1l}^{-1/3} y_{1k}^{1/3} = 1,$$

$$(16) \quad p_1 \frac{\gamma_1}{3} y_{1l}^{2/3} y_{1k}^{-2/3} = p_k + \tau,$$

$$(17) \quad p_2 \frac{\gamma_2}{2} y_{2l}^{-1/2} y_{2k}^{1/2} = 1,$$

$$(18) \quad p_2 \frac{\gamma_2}{2} y_{2l}^{1/2} y_{2k}^{-1/2} = p_k.$$

In addition, the economic conjecture is that

$$(19) \quad \gamma_1 p_1 y_{1l}^{2/3} y_{1k}^{1/3} + \gamma_2 p_2 y_{2l}^{1/2} y_{2k}^{1/2} > 1.05 \times \text{benchmark output}.$$

Note that all functions are semialgebraic and that one can rewrite these equations as a system of polynomial equations.

In step 2, the fact that output per sector is observable together with the firms' optimality conditions uniquely determine the parameters in the production functions as well as firms' factor demand. For concreteness, suppose total outputs are 60 and 40, and that benchmark prices are all equal to 1. This implies  $\gamma_1 = \gamma_2 = 2$ ,  $L = 60$ , and  $K = 36.875$  with a tax rate around 18.5 percent (to match the equilibrium exactly, with  $\gamma_1 = \gamma_2 = 2$  one needs to set  $\tau = (\frac{32,767}{19,683})^{1/3}$ ), which corresponds to a tax revenue of  $T = 3.125$ . The observations impose the following restrictions on agents' preference parameters and the distribution of capital and labor endowments across agents:

$$(20) \quad \xi_1 (k_1 + l_1 + T/2) + \xi_2 (T/2 + K + L - k_1 - l_1) = 60.$$

Without loss of generality, one can assume that agent 1 holds less capital than agent 2. Furthermore, I assume that he has larger labor endowments and a propensity to consume commodity 1 of less than 1/2. Therefore, I define the set of admissible exogenous variables to be

$$E = [0, 10] \times [30, 55] \times [0.05, 0.5].$$

For each  $(k_1, l_1, \xi_1) \in E$ , the  $\xi_2$  that solves Equation (20) turns out to lie between 0 and 1.

In step 3, given the definition of  $E$ , one now needs to find a system of polynomial equations that characterizes the set  $\Phi$ , i.e., the set of all  $e \in E$  such that  $\exists(p_1, p_2, p_k), (\bar{y}_{jk}, \bar{y}_{jl})_{j=1,2}$  that satisfy Equations (11)–(18) with  $\tau = T = 0$ ,  $l_2 = L - l_1, k_2 = K - k_1$  as well as Equations (19) and (20).

To find good bounds on the number of connected components, it is now useful to rewrite these equations as polynomial equations, substituting for as many variables as possible. In general, computer algebra systems such as Maple or Mathematica are ideally suited for this task. In this simple example, it can be easily done by hand. Using the fact that

$$p_k = \frac{y_{1l}}{2y_{1k}} = \frac{y_{2l}}{y_{2k}},$$

the system (11)–(18) and (19) can be rewritten as a system of polynomial equations and inequalities in  $y = y_{1l}$  and in the unknown parameters. After some relatively straightforward algebra, one obtains that  $(k_1, l_1, \xi_1) \in \Phi$  if there exist  $y > 0$  and such that

$$\begin{aligned} & -59l_1^2\xi_1 + 165l_1k_1\xi_1 + \frac{12}{5}l_1k_1y - 192l_1k_1 + \frac{17,405}{16}l_1\xi_1 - \frac{177}{2}l_1y \\ & + 7,080l_1 - 96k_1^2\xi_1 + \frac{12}{5}k_1^2y - 2,670k_1\xi_1 - \frac{1,299}{4}k_1y + 11,520k_1 \\ & + \frac{22,125}{2}\xi_1 + \frac{278,775}{32}y - 424,800 = 0, \\ & 15 - \frac{1}{2}y > 0. \end{aligned}$$

In the fourth step, one needs to verify that the conjecture is correct on a grid of points. For concreteness, I take  $F$  to consist of  $1,000 \times 1,000 \times 1,000$  equispaced points,

$$\begin{aligned} F = \{0.01, 0.02, \dots, 10\} \times \{30.025, 30.050, \dots, 55\} \\ \times \{0.05045, 0.0509, \dots, 0.5\}. \end{aligned}$$

Because  $y$  is a linear function in the parameters, it is trivial to find good error bounds. The fact that if  $a + bx = \varepsilon$ , there is an  $\bar{x}$  with  $a + b\bar{x} = 0$  and  $|\bar{x} - x| = \frac{\varepsilon}{b}$

can be seen as a trivial special case of Smale’s formula. The fraction of points that can thus be shown to lie in  $\Phi$  turns out to be greater than 0.999998.

Last, in the fifth step of the procedure, to bound the size of  $\Phi$ , one then needs a bound on the number of connected components. For this, one can use Equation (8) from Section 4.2. How large is the number of connected components of  $\Phi$  along a line parallel to the  $l_1$  axes? To bound this, one needs to fix an arbitrary  $\bar{\xi}_1, \bar{k}_1$  and consider the set of all  $y, l_1, t$  such that

$$\begin{aligned}
 & -59l_1^2\bar{\xi}_1 + 165l_1\bar{k}_1\bar{\xi}_1 + \frac{12}{5}l_1\bar{k}_1\bar{y} - 192l_1\bar{k}_1 + \frac{17,405}{16}l_1\bar{\xi}_1 - \frac{177}{2}l_1y \\
 & + 7,080l_1 - 96\bar{k}_1^2\bar{\xi}_1 + \frac{12}{5}\bar{k}_1^2\bar{y} - 2,670\bar{k}_1\bar{\xi}_1 - \frac{1,299}{4}\bar{k}_1y + 11,520\bar{k}_1 \\
 & + \frac{22,125}{2}\bar{\xi}_1 + \frac{278,775}{32}y - 424,800 = 0, \\
 & 1 - t\left(15 - \frac{1}{2}y\right) = 0.
 \end{aligned}$$

The volume of the convex hull of the three unit vectors together with  $(2, 0, 0)$ ,  $(1, 1, 0)$ , and  $(0, 1, 1)$  is less than 2. Therefore, one obtains a bound on the number of connected components of  $2^{3-1} \times 2 = 8$ .

For lines along the  $\xi_1$  or  $k_1$  axes, the argument is similar and one obtains  $\lambda \leq 8$ . To apply Theorem 1, note that  $\lambda \frac{l}{N} = \frac{24}{1,000}$ . Therefore, the normalized volume of  $\Phi$  is greater than 0.975.

### 6. A RANDOM ALGORITHM

Suppose one has access to a random number generator and can draw uniformly and independently random  $\tilde{e} \in \{1/N, \dots, 1\}^l$ . See, e.g., L’Ecuyer (2004) for a discussion on generating random and quasi-random numbers. There are now a few web sites that offer sources of random numbers that are generated by physical processes. For example, at [www.randomnumbers.info](http://www.randomnumbers.info), the user can download numbers that are generated by a physical random number generator that exploits an elementary quantum optics process. A precise description of the physical principles that underlie the method can be obtained at that site.

Both random and pseudo-random numbers are naturally integer valued (see Blum, Cucker, Shub, and Smale (1998) for a more elaborate discussion on *probabilistic machines*) and, therefore, lie on a grid. It is not possible to draw random numbers uniformly over an interval, but it is possible to generate random draws from a finite set. Suppose as before that  $E = [0, 1]^l$  and that  $F = \{1/N, \dots, 1\}^l$ . In this formulation,  $N$  is now the number of digits of the random numbers and can be thought of as relatively large (however, one should keep in mind that the cost of generating random numbers increases

with the size of the numbers; it is not reasonable to assume that  $N$  is arbitrarily large).

Suppose one has  $M \times l$  random numbers drawn independently and identically distributed from  $\{1, \dots, N\}$ . Scaled appropriately, this gives  $M$  random vectors  $\tilde{e}^1, \dots, \tilde{e}^M \in F$ . If, for each  $i = 1, \dots, M$ , there is an  $\bar{e}$  with  $\|\bar{e} - \tilde{e}^i\| \leq \varepsilon$  and with  $\bar{e} \in \Phi$ ,  $\mathfrak{S}^\varepsilon(\bar{e}^i) = 1$  for all  $i$ , then by the binomial formula one obtains that the probability of the event that the fraction of points  $x \in F$  for which  $\mathfrak{S}^\varepsilon(x) = 0$  is greater than  $\delta$  must be less than or equal to  $(1 - \delta)^M$ . Therefore,

$$\text{Prob} \left[ \frac{1}{N^l} \sum_{i_1, \dots, i_l} \mathfrak{S}^\varepsilon \left( \frac{i_1}{N}, \dots, \frac{i_l}{N} \right) < 1 - \delta \right] \leq (1 - \delta)^M.$$

Using Theorem 1, one can now infer probabilistic statements about the size of  $\Phi$  from probabilistic statements about the number of points in the finite grid for which the statement is true. If  $\varepsilon < 1/N$  and, as before, letting  $\lambda$  denote a bound on the maximal number of connected components of  $\Phi \cap C_i^\varepsilon$ , the fact that

$$\int_{[0,1]^l} \mathfrak{S}^0(x) dx \geq \frac{1}{N^l} \sum_{i_1, \dots, i_l} \mathfrak{S}^\varepsilon \left( \frac{i_1}{N}, \dots, \frac{i_l}{N} \right) - \left( 2\varepsilon + \frac{l}{N} \right) \lambda$$

implies that

$$(21) \quad \text{Prob} \left[ \int_{[0,1]^l} \mathfrak{S}^0(x) dx < 1 - \delta - \left( 2\varepsilon + \frac{l}{N} \right) \lambda \right] \leq (1 - \delta)^M.$$

See Koiran (1995) or Blum, Cucker, Shub, and Smale (1998, Chapter 17.4) for the case  $\varepsilon = 0$  and a discussion of the result.

Note that the number of points needed is independent of the dimension, which enters only through bounds on the number of connected components. In the example in Section 5.3, one had to verify the conjecture at a billion points to bound the size of the residual set to be less than 0.025. Using the probabilistic approach, if one draws 1,000 random points with five significant digits (i.e.,  $N = 100,000$ ) from the set of admissible parameters and takes  $\delta = 0.0049$ , one obtains that the probability that the set  $\Phi$  is greater than  $0.996 - \frac{3\lambda}{100,000}$  is at least  $1 - (1 - 0.0049)^{1,000} = 0.9926$ . It is, therefore, easy to verify that with high probability (at least 0.9926) the set  $\Phi$  is greater than 0.995. Note that the bound on  $\lambda$  obtained in Section 5.3 is crucial to this argument. In the example, 1,000 points suffice to show that the set is greater than 0.995 with high probability, but one needs a billion points to show that this is true with certainty.

Although the random method is much more efficient, it is not clear how to interpret a statement like “29 is a prime number with probability 0.9926.” Even though it is well known in theoretical computer science that random algorithms

often reduce the complexity of the problem considerably, these algorithms usually solve a specific problem and it can often be checked that the candidate solution produced by the algorithm is an actual solution (without probabilities attached). One possible interpretation of Equation (21) is the following. Suppose *nature* draws randomly a vector of parameters  $e$  uniformly from  $[0, 1]^l$ . Equation (21) implies that the overall probability that this parameter will lie in  $\Phi$  is at least  $(1 - \delta - (2\varepsilon + \frac{l}{N})\lambda)(1 - (1 - \delta)^N)$ . This, therefore, allows for statistical statements about how likely it is that the conjecture is true for randomly selected parameters.

Note that the number of connected components can be fairly large if  $N$  is sufficiently large and  $\varepsilon$  is very small. However, in practice a bound for the volume arises naturally from the precision with which equilibria can be computed, i.e., with  $\varepsilon$ . Because Theorem 1 is only valid if  $1/N > \varepsilon$ , these methods are applicable if and only if the number of connected components is orders of magnitude smaller than  $1/\varepsilon$ .

A naive application of the random algorithm without knowledge of bounds on  $\lambda$  as defined in Equation (5) does not allow for any statements about the size of  $\Phi$ . If, for example,  $\Phi \subset [0, 1]$  consists of all irrational numbers, even without any computational error, because any random number will certainly be rational (in fact integer valued), the method predicts the volume of  $\Phi$  to be 0, while in reality it is equal to 1. Only when one can bound  $\lambda$  (even if this bound turns out to be fairly large) can one make meaningful probabilistic statements about the size of  $\Phi$  by randomly sampling it.

## 7. CONCLUSION

Computational experiments that make statements about one specific example economy can be generalized to infinite classes of economies when the economic fundamentals are definable in an o-minimal structure. Theorem 1, the main theoretical result, precisely specifies the conditions under which finitely many examples suffice to make statements about sets of parameters with positive Lebesgue measure.

I argue that this theoretical insight can be fruitfully put to work in applied general equilibrium analysis. For all commonly used specifications of utility and production functions, one can easily compute how many examples are needed to make statements about *large* sets of parameters. These statements are possible even if equilibria cannot be computed exactly. The methods are directly applicable to models whose solutions can be characterized by finite systems of equations; this includes stationary equilibria and steady states in infinite horizon models used in modern macroeconomics and public finance.

However, it turns out that in large problems the number of examples needed is astronomically high and it is, therefore, not feasible to make general statements using a deterministic algorithm. A random algorithm can be used to make statements about the probability that a given conjecture holds for a set of relative size  $1 - \delta$ .

Computing numerous random examples and then using statistical inference to summarize the findings is not a new idea (see, e.g., Judd (1997)), but it has not previously been formalized to take into account finite precision arithmetics of actual computations. For this method the most important practical insight of this paper is about the interplay of errors in computation,  $\varepsilon$ , the size of the random numbers used,  $N$ , and the number of connected components. One can estimate the Lebesgue measure of the set  $\Phi$  by randomly drawing examples if the number of connected components is orders of magnitude smaller than  $1/\varepsilon$ . Otherwise, it is not possible to make even probabilistic statements about the size of  $\Phi$ .

The methods introduced in this paper are obviously not the only ones that can be used to show that a given formula holds for a rich class of parameters. Because the real closed field is decidable, one can apply algorithmic quantifier elimination and use an algorithm to verify whether a given semialgebraic statement of interest is true for *all* parameters in a given (semialgebraic) set (see, e.g., Basu, Pollack, and Roy (2003)). However, for more complicated structures, decidability is an open problem and there are certainly no algorithms available for quantifier elimination at this time. Moreover, even for the semi-algebraic case, the methods in this paper are much more tractable than quantifier elimination.

*Department of Economics, 3718 Locust Way, University of Pennsylvania, Philadelphia, PA 19104, U.S.A.; kubl@sas.upenn.edu.*

*Manuscript received May, 2005; final revision received December, 2006.*

#### REFERENCES

- BASU, S., R. POLLACK, AND M.-F. ROY (2003): *Algorithms in Real Algebraic Geometry*. New York: Springer-Verlag. [991]
- BLUM, L., F. CUCKER, M. SHUB, AND S. SMALE (1998): *Complexity and Real Computation*. New York: Springer Verlag. [970,982,988,989]
- BLUME, L., AND W. R. ZAME (1993): "The Algebraic Geometry of Competitive Equilibrium," in *Essays in General Equilibrium and International Trade: In Memoriam Trout Rader*, ed. by W. Neufeind. New York: Springer-Verlag, 53–66. [970,984]
- DEBREU, G. (1970): "Economies with a Finite Set of Equilibria," *Econometrica*, 38, 387–392. [981]
- GABRIELOV, A., AND N. VOROBOV (2004): "Complexity of Computation with Pfaffian and Noetherian Functions," in *Normal Forms, Bifurcations and Finiteness Problems in Differential Equations*. Dordrecht: Kluwer, 211–250. [970,978,980]
- GABRIELOV, A., N. VOROBOV, AND T. ZELL (2003): "Betti Numbers of Semialgebraic and Sub-Pfaffian Sets," *Journal London Mathematical Society*, 69, 27–43. [979]
- JUDD, K. (1997): "Computational Economics and Economic Theory: Complements or Substitutes?" *Journal of Economic Dynamics and Control*, 21, 907–942. [970,991]
- KHOVANSKII, A. (1991): *Fewnomials, Translations of Mathematical Monographs*, Vol. 88. Providence, RI: American Mathematical Society. [970,978]
- KOIRAN, P. (1995): "Approximating the Volume of Definable Sets," in *Proceedings of the 36th IEEE Symposium on Foundations of Computer Science*. New York: IEEE, 134–141. [968,974, 976,989]

- KUBLER, F., AND K. SCHMEDDERS (2005): "Approximate versus Exact Equilibria in Dynamic Economies," *Econometrica*, 73, 1205–1235. [968,981]
- LECUYER, P. (2004): "Random Number Generation," in *Handbook of Computational Statistics*, ed. by J. E. Gentle, W. Haerdle, and Y. Mori. New York: Springer-Verlag, 35–70. [988]
- NACHBAR, J. (2002): "General Equilibrium Comparative Statics," *Econometrica*, 70, 2065–2074. [970]
- POSTLEWAITE, A., AND D. SCHMEIDLER (1981): "Approximate Walrasian Equilibrium and Nearby Economies," *International Economics Review*, 22, 105–111. [981]
- RICHTER, M. K., AND K.-C. WONG (2000): "Definable Utility in O-Minimal Structures," *Journal of Mathematical Economics*, 34, 159–172. [970,984]
- ROJAS, J. M. (2000): "Some Speed-ups and Speed Limits in Real Algebraic Geometry," *Journal of Complexity*, 16, 552–571. [980]
- ROLIN, J.-P., P. SPEISSEGER, AND A. WILKIE (2003): "Quasianalytic Denjoy–Carleman Classes and O-Minimality," *Journal of the American Mathematical Society*, 16, 751–777. [985]
- SCARE, H. (1967): "On the Computation of Equilibrium Prices," in *Ten Studies in the Tradition of Irving Fisher*. New York: Wiley, Chapter 6. [981]
- SHOVEN, J. B. (1976): "The Incidence and Efficiency Effects of Taxes on Income from Capital," *Journal of Political Economy*, 84, 1261–1283. [985]
- VAN DEN DRIES, L. (1999): *Tame Topology and O-Minimal Structures*. Cambridge, U.K.: Cambridge University Press. [983,984]
- WILKIE, A. J. (1996): "Model Completeness Results of Restricted Pfaffian Functions and the Exponential Function," *Journal of the American Mathematical Society*, 9, 1051–1094. [983]
- WILKINSON, J. H. (1984): "The Perfidious Polynomial," in *Studies in Numerical Analysis, Studies in Mathematics*, Vol. 24, ed. by G. H. Golub. Washington DC: Mathematical Association of America, 1–28. [972]