

# Non-linear Real Arithmetic Benchmarks derived from Automated Reasoning in Economics

Mulligan, C., Bradford, R., Davenport, J. H., England, M. & Tonks, Z.

**Author post-print (accepted) deposited by Coventry University's Repository**

**Original citation & hyperlink:**

Mulligan, C, Bradford, R, Davenport, JH, England, M & Tonks, Z 2018, Non-linear Real Arithmetic Benchmarks derived from Automated Reasoning in Economics. in Proceedings of the 3rd International Workshop on Satisfiability Checking and Symbolic Computation. CEUR Workshop Proceedings, Third International Workshop on Satisfiability Checking and Symbolic Computation 2018, Oxford, United Kingdom, 7/07/18.

URL: <http://ceur-ws.org/Vol-2189/>

ISSN 1613-0073

ESSN 1613-0073

Publisher: CEUR Workshop Proceedings

**Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.**

**This document is the author's post-print version, incorporating any revisions agreed during the peer-review process. Some differences between the published version and this version may remain and you are advised to consult the published version if you wish to cite from it.**

# Non-linear Real Arithmetic Benchmarks derived from Automated Reasoning in Economics

Casey B. Mulligan<sup>1</sup>, Russell Bradford<sup>2</sup>, James H. Davenport<sup>2</sup>,  
Matthew England<sup>3</sup>, and Zak Tonks<sup>2</sup>

<sup>1</sup> University of Chicago, USA

`c-mulligan@uchicago.edu`

<sup>2</sup> University of Bath, UK

`{R.J.Bradford, J.H.Davenport, Z.P.Tonks}@bath.ac.uk`

<sup>3</sup> Coventry University, UK

`Matthew.England@coventry.ac.uk`

**Abstract.** We consider problems originating in economics that may be solved automatically using mathematical software. We present and make freely available a new benchmark set of such problems. The problems have been shown to fall within the framework of non-linear real arithmetic, and so are in theory soluble via Quantifier Elimination (QE) technology as usually implemented in computer algebra systems. Further, they all can be phrased in prenex normal form with only existential quantifiers and so are also admissible to those Satisfiability Module Theory (SMT) solvers that support the `QF_NRA` logic. There is a great body of work considering QE and SMT application in science and engineering, but we demonstrate here that there is potential for this technology also in the social sciences.

## 1 Introduction

### 1.1 Economic reasoning

A general task in economic reasoning is to determine whether, with variables  $\mathbf{v} = (v_1, \dots, v_n)$ , the hypotheses  $H(\mathbf{v})$  follow from the assumptions  $A(\mathbf{v})$ , i.e. is it the case that

$$\forall \mathbf{v}. A(\mathbf{v}) \Rightarrow H(\mathbf{v})?$$

Ideally the answer would be **True** or **False**, and of course logically it is. But the application being modelled, like real life, can require a more nuanced analysis. It may be that for most  $\mathbf{v}$  the theorem holds but there are some special cases that should be ruled out (additions to the assumptions). Such knowledge is valuable to the economist. Another possibility is that the particular set of assumptions chosen are contradictory<sup>4</sup>, i.e.  $A(\mathbf{v})$  itself is **False**. As all students of an introductory logic class know, this would technically make the implication **True**, but for the purposes of economics research it is important to identify this separately!

---

<sup>4</sup> A lengthy set of assumptions is many times required to conclude the hypothesis of interest so the possibility of contradictory assumptions is real.

**Table 1.** Table of possible outcomes from a potential theorem  $\forall. \mathbf{v}A \Rightarrow H$

	$\neg\exists\mathbf{v}[A \wedge \neg H]$	$\exists\mathbf{v}[A \wedge \neg H]$
$\exists\mathbf{v}[A \wedge H]$	<b>True</b>	<b>Mixed</b>
$\neg\exists\mathbf{v}[A \wedge H]$	Contradictory Assumptions	<b>False</b>

We categorise the situation into four possibilities that are of interest to an economist (Table 1) via the outcome of a pair of fully existentially quantified statements which check the existence of both an example ( $\exists\mathbf{v}[A \wedge H]$ ) and counterexample ( $\exists\mathbf{v}[A \wedge \neg H]$ ) of the theorem. So we see that every economics theorem generates a pair of SAT problems, in practice actually a trio since we would first perform the cheaper check of the compatibility of the assumptions ( $\exists\mathbf{v}[A]$ ).

Such a categorisation is valuable to an economist. They may gain a proof of their theorem, but if not they still have important information to guide their research: knowledge that their assumptions contradict; or information about  $\{\mathbf{v} : A(\mathbf{v}) \Rightarrow H(\mathbf{v})\}$ .

Also of great value to an economist are the options for exploration that such technology would provide. An economist could vary the question by strengthening the assumptions that led to a **Mixed** result in search of a **True** theorem. However, of equal interest would be to weaken the assumption that generated a **True** result for the purpose of identifying a theorem that can be applied more widely. Such weakening or strengthening is implemented by quantifying more or less of the variables in  $\mathbf{v}$ . For example, we might partition  $\mathbf{v}$  as  $\mathbf{v}_1, \mathbf{v}_2$  and ask for  $\{\mathbf{v}_1 : \forall\mathbf{v}_2. A(\mathbf{v}_1, \mathbf{v}_2) \Rightarrow H(\mathbf{v}_1, \mathbf{v}_2)\}$ . The result in these cases would be a formula in the free variables that weakens or strengthens the assumptions as appropriate.

## 1.2 Technology to solve such problems automatically

Such problems all fall within the framework of *Quantifier Elimination* (QE). QE refers to the generation of an equivalent quantifier free formula from one that contains quantifiers. SAT problems are thus a sub-case of QE when all variables are existentially quantified.

QE is known to be possible over real closed fields (real QE) thanks to the seminal work of Tarski [34] and practical implementations followed the work of Collins on the Cylindrical Algebraic Decomposition (CAD) method [11] and Weispfenning on Virtual Substitution [36]. There are modern implementations of real QE in MATHEMATICA [32], REDLOG [14], MAPLE (SYNRAC [21] and the REGULARCHAINS Library [10]) and QEPCAD-B [6].

The economics problems identified all fall within the framework of QE over the reals. Further, the core economics problem of checking a theorem can be answered via fully existentially quantified QE problems, and so also soluble using the technology of Satisfiability Modulo Theory (SMT) Solvers; at least those that support the QF\_NRA (Quantifier Free Non-Linear Real Arithmetic) logic such as SMT-RAT [12], VERiT [18], YICES2 [24], and Z3 [23].

### 1.3 Case for novelty of the new benchmarks

QE has found many applications within engineering and the life sciences. Recent examples include the derivation of optimal numerical schemes [17], artificial intelligence to pass university entrance exams [35], weight minimisation for truss design [9], and biological network analysis [3]. However, applications in the social sciences are lacking (the nearest we can find is [25]).

Similarly, for SMT with non-linear reals, applications seem focussed on other areas of computer science. The original `nlsat` benchmark set [23] was made up mostly of verification conditions from the KEYMAERA [30], and theorems on polynomial bounds of special functions generated by the METITARSKI automatic theorem prover [29]. This category of the SMT-LIB [1] has since been broadened with problems from physics, chemistry and the life sciences [33]. However, we are not aware of any benchmarks from economics or the social sciences.

The reader may wonder why QE has not been used in economics previously. On a few occasions when QE algorithms have been mentioned in economics they have been characterized as “something that is do-able in principle, but not by any computer that you and I are ever likely to see” [31]. Such assertions were based on interpretations of theoretical computer science results rather than experience with actual software applied to an actual economic reasoning problem. Simply put, the recent progress on QE/SMT technology is not (yet) well known in the social sciences.

### 1.4 Availability and further details

The benchmark set consists of 45 potential economic theorems. Each theorem requires the three QE/SMT calls to check the compatibility of assumptions, the existence of an example, and the existence of a counterexample, so 135 problems in total. In all cases the assumption and example checks are SAT, in fact often fully satisfied (any point can witness it) as they relate to a true theorem. Correspondingly, the counterexample is usually UNSAT (for 42/45 theorems) and thus the more difficult problem from the SMT point of view.

The benchmark problems are hosted on the Zenodo data repository at URL <https://doi.org/10.5281/zenodo.1226892> in both the SMT2 format and as input files suitable for REDLOG and MAPLE. The SMT2 files have been accepted into the SMT-LIB [1] and will appear in the next release.

Available from <http://examples.economicreasoning.com/> are MATHEMATICA notebooks<sup>5</sup> which contain commands to solve the examples in MATHEMATICA and also further information on the economic background: meaning of variable names and economic implications of results etc.

---

<sup>5</sup> A MATHEMATICA licence is needed to run them, but static pdf print outs are also available to download.

## 1.5 Plan

The paper proceeds as follows. In Section 2 we describe in detail some examples from economics, ranging from textbook examples common in education to questions arising from current research discussions. Then in Section 3 we offer some statistics on the logical and algebraic structure of these examples. We then give some final thoughts on future and ongoing work in Section 4.

## 2 Examples of Economic Reasoning in Tarski's Algebra

The fields of economics ranging from macroeconomics to industrial organization to labour economics to econometrics involve deducing conclusions from assumptions or observations. Will a corporate tax cut cause workers to get paid more? Will a wage regulation reduce income inequality? Under what conditions will political candidates cater to the median voter? We detail in this section a variety of such examples.

### 2.1 Comparative static analysis

We start with Alfred Marshall's [26] classic, and admittedly simple, analysis of the consequences of cost-reducing progress for activity in an industry. Marshall concluded that, for any supply-demand equilibrium in which the two curves have their usual slopes, a downward supply shift increases the equilibrium quantity  $q$  and decreases the equilibrium price  $p$ .

One way to express his reasoning in Tarski's framework is to look at the industry's local comparative statics: meaning a comparison of two different equilibrium states between supply and demand. With a downward supply shift represented as  $da > 0$  (where  $a$  is a reduction in costs) we have here:

$$\begin{aligned} A &\equiv D'(q) < 0 \wedge S'(q) > 0 \\ &\quad \wedge \frac{d}{da}(S(q) - a) = \frac{dp}{da} \wedge \frac{dp}{da} = \frac{d}{da}D(q) \\ H &\equiv \frac{dq}{da} > 0 \wedge \frac{dp}{da} < 0 \end{aligned}$$

where:

- $D'(q)$  is the slope of the inverse demand curve in the neighborhood of the industry equilibrium;
- $S'(q)$  is the slope of the inverse supply curve;
- $\frac{dq}{da}$  is the quantity impact of the cost reduction; and
- $\frac{dp}{da}$  is the price impact of the cost reduction.

Economically, the first atoms of  $A$  are the usual slope conditions: that demand slopes down and supply slopes up. The last two atoms of  $A$  say that the cost

change moves the industry from one equilibrium to another. Marshall’s hypothesis was that the result is a greater quantity traded at a lesser price.

Hence we set the “variables”  $\mathbf{v}$  to be four real numbers  $(v_1, v_2, v_3, v_4)$ :

$$\mathbf{v} = \left\{ D'(q), S'(q), \frac{dq}{da}, \frac{dp}{da} \right\}.$$

Then, after applying the chain rule,  $A$  and  $H$  may be understood as Boolean combinations of polynomial equalities and inequalities:

$$\begin{aligned} A &\equiv v_1 < 0 \wedge v_2 > 0 \wedge v_3 v_2 - 1 = v_4 \wedge v_4 = v_3 v_1, \\ H &\equiv v_3 > 0 \wedge v_4 < 0. \end{aligned}$$

Thus Marshall’s reasoning fits in the Tarski framework and therefore is amenable to automatic solution. Any of the tools mentioned in the introduction can easily evaluate the two existential problems for Table 1 and conclude easily that  $\forall \mathbf{v}. A \Rightarrow H$  is **True**, confirming Marshall’s conclusion.

In fact, for this example it is straightforward to produce a short hand proof in a couple of lines<sup>6</sup> and so we did not include it in the benchmark set. But it demonstrates the basic approach to placing economics reasoning into a form admissible for QE/SMT tools. A similar approach is used for many of the larger examples forming the benchmark set, such as the next one.

## 2.2 Scenario analysis

Economics is replete with “*what if?*” questions. Such questions are logically and algebraically more complicated, and thereby susceptible to human error, because they require tracking various scenarios. We consider now one such question originating from recent economic debate.

Writing at nytimes.com<sup>7</sup>, Economics Nobel laureate Paul Krugman asserted that whenever taxes on labour supply are primarily responsible for a recession, then wages increase. Two scenarios are discussed here: what actually happens (*act*) when taxes (*t*) and demand forces (*a*) together create a recession, and what would have happened (*hyp*) if taxes on labour supply had been the only factor affecting the labour market.

<sup>6</sup> Subtract the final assumption from the penultimate one and then apply the sign conditions of the other assumptions to conclude  $v_3 > 0$ . Then with this the last assumption provides the other hypothesis,  $v_4 < 0$ .

<sup>7</sup> <https://krugman.blogs.nytimes.com/2012/11/03/soup-kitchens-caused-the-great-depression/>

Expressed logically we have:

$$\begin{aligned}
A \equiv & \left( \frac{\partial D(w, a)}{\partial w} < 0 \wedge \frac{\partial S(w, t)}{\partial w} > 0 \right. \\
& \wedge \frac{\partial D(w, a)}{\partial a} = 1 \wedge \frac{\partial S(w, t)}{\partial t} = 1 \\
& \wedge \frac{d}{dact}(D(w, a) = q = S(w, t)) \\
& \wedge \frac{d}{dhyp}(D(w, a) = q = S(w, t)) \\
& \wedge \frac{dt}{dact} = \frac{dt}{dhyp} \wedge \frac{da}{dhyp} = 0 \\
& \left. \wedge \frac{dq}{dhyp} < \frac{1}{2} \frac{dq}{dact} < 0 \right) \\
H \equiv & \frac{dw}{dact} > 0.
\end{aligned}$$

In economics terms, the first line of assumptions contains the usual slope restrictions on the supply and demand curves. Because nothing is asserted about the units of  $a$  or  $t$ , the next line just contains normalizations. The third and fourth lines say that each scenario involves moving the labour market from one equilibrium (at the beginning of a recession) to another (at the end of a recession). The fifth line defines the scenarios: both have the same tax change but only the *act* scenario has a demand shift. The final assumption / assertion is that a majority of the reduction in the quantity  $q$  of labour was due to supply (that is, most of  $\frac{dq}{dact}$  would have occurred without any shift in demand). The hypothesis is that wages  $w$  are higher at the end of the recession than they were at the beginning.

Viewed as a Tarski formula this has twelve variables,

$$\mathbf{v} = \left\{ \frac{da}{dact}, \frac{da}{dhyp}, \frac{dt}{dact}, \frac{dt}{dhyp}, \frac{dq}{dact}, \frac{dq}{dhyp}, \frac{dw}{dact}, \frac{dw}{dhyp}, \frac{\partial D(w, a)}{\partial a}, \frac{\partial S(w, t)}{\partial t}, \frac{\partial D(w, a)}{\partial w}, \frac{\partial S(w, t)}{\partial w} \right\},$$

each of which is a real number representing a partial derivative describing the supply and demand function or a total derivative indicating a change over time within a scenario.

An analysis of the two existential problems as suggested in Section 1.1. shows that the result is **Mixed**: that is both examples and counterexamples exist. In particular, even when all of the assumptions are satisfied, it is possible that wages actually go down.

Moreover, if  $\frac{\partial D(w, a)}{\partial w}$  and  $\frac{\partial S(w, t)}{\partial w}$  are left as free variables, a QE analysis recovers a disjunction of three quantifier-free formulae. Two of them contradict

the assumptions, but the third does not and can therefore be added to  $A$  in order to guarantee the truth of  $H$ :

$$\frac{\partial S(w, t)}{\partial w} \geq -\frac{\partial D(w, a)}{\partial w} > 0.$$

I.e. assuming that labour supply is at least as sensitive to wages as labour demand is (recall that the demand slope is a negative number) guarantees that wages  $w$  are higher at the end of the recession that is primarily due to taxes on labour supply. See also [27] for further details on the economics.

This example can be found in the benchmark set as Model #0013.

### 2.3 Vector summaries

Economics problems sometimes involve an unspecified (and presumably large) number of variables. Take Sir John Hicks' analysis of household decisions among a large number  $N$  of goods and services [19]. The quantities purchased are represented as a  $N$ -dimensional vector, as are the prices paid.

We assume that when prices are  $\mathbf{p}$  ( $\hat{\mathbf{p}}$ ), the household makes purchases  $\mathbf{q}$  ( $\hat{\mathbf{q}}$ ), respectively:<sup>8</sup>

$$A \equiv (\mathbf{p} \cdot \mathbf{q} \leq \mathbf{p} \cdot \hat{\mathbf{q}}) \wedge (\hat{\mathbf{p}} \cdot \hat{\mathbf{q}} \leq \hat{\mathbf{p}} \cdot \mathbf{q}).$$

Hicks asserted that the quantity impact of the price changes  $\hat{\mathbf{q}} - \mathbf{q}$  cannot be positively correlated with the price changes  $\hat{\mathbf{p}} - \mathbf{p}$ :

$$H \equiv (\hat{\mathbf{q}} - \mathbf{q}) \cdot (\hat{\mathbf{p}} - \mathbf{p}) \leq 0.$$

Hicks' reasoning depends only on the value of vector dot products, four of which appear above, rather than the actual vectors themselves whose length remains unspecified.

Hicks implicitly assumed that prices and quantities are real valued, which places additional restrictions on the dot products. We need then to restrict that the symmetric Gram matrix corresponding to the four vectors  $\hat{\mathbf{q}}, \mathbf{q}, \hat{\mathbf{p}}, \mathbf{p}$  be positive semi-definite. To state this restriction we then need to increase the list of variables from the four dot products that appear in  $A$  and  $H$  to all ten that it is possible to produce with four vectors.

QE technology certifies that there are no values for the dot products that can simultaneously satisfy  $A$  and contradict  $H$ . It follows that Sir John Hicks was correct regardless of the length  $N$  of the price and quantity vectors. In fact, further, it turns out that the additional restriction to real valued quantities is not needed for this example (there are no counter-examples even without this). While those additional restrictions could hence be removed from the theorem we have left them in the statement in the benchmark set because (a) they would

<sup>8</sup> The price change is compensated in the Hicksian sense, which means that  $\mathbf{q}$  and  $\hat{\mathbf{q}}$  are the same in terms of results or "utility", so that the consumer compares them only on the basis of what they cost (dot product with prices).



be standard restrictions an economist would work with and (b) it is not obvious that they are unnecessary before the analysis.

The above analysis could be performed easily with both MATHEMATICA and Z3 but proved difficult for REDLOG. We have yet to find a variable ordering that leads to a result in reasonable time<sup>9</sup>. It is perhaps not surprising that a tool like Z3 excels in comparison to computer algebra here, since the problem has a large part of the logical statement irrelevant to its truth: the search based SAT algorithms are explicitly designed to avoid considering this.

This example can be found in the benchmark set as Model #0078.

## 2.4 Concave production function

Our final example is adapted from the graduate-level microeconomics textbook [22]. The example asserts that any differentiable, quasi-concave, homogeneous, three-input production function with positive inputs and marginal products must also be a concave function. In the Tarski framework, we have a twelve-variable sentence, which we have expressed simply with variable names  $v_1, \dots, v_{12}$  to compress the presentation:

$$\begin{aligned}
A \equiv & v_1 v_{10} + v_2 v_7 + v_3 v_5 = 0 \wedge v_1 v_{11} + v_2 v_8 + v_3 v_7 = 0 \\
& \wedge v_1 v_{12} + v_{10} v_3 + v_{11} v_2 = 0 \\
& \wedge v_1 > 0 \wedge v_2 > 0 \wedge v_3 > 0 \wedge v_4 > 0 \wedge v_6 > 0 \wedge v_9 > 0 \\
& \wedge 2v_{11}v_6v_9 > v_{12}v_6^2 + v_8v_9^2 \\
& \wedge 2v_{10}v_6(v_{11}v_4 + v_7v_9) + v_9(2v_{11}v_4v_7 - 2v_{11}v_5v_6 + v_5v_8v_9) \\
& + v_{12}(v_4^2v_8 - 2v_4v_6v_7 + v_5v_6^2) > v_{10}^2v_6^2 + 2v_{10}v_4v_8v_9 + v_{11}^2v_4^2 + v_7^2v_9^2, \\
H \equiv & v_{12} \leq 0 \wedge v_5 \leq 0 \wedge v_8 \leq 0 \\
& \wedge v_{12}v_5 \geq v_{10}^2 \wedge v_{12}v_8 \geq v_{11}^2 \wedge v_8v_5 \geq v_7^2 \\
& \wedge v_8(v_{10}^2 - v_{12}v_5) + v_{11}^2v_5 + v_{12}v_7^2 \geq 2v_{10}v_{11}v_7.
\end{aligned}$$

In disjunctive normal form (DNF)  $A \wedge \neg H$  is a disjunction of seven clauses. Each atom of  $H$ , and therefore each clause of  $A \wedge \neg H$  in DNF, corresponds to a principal minor of the production function's Hessian matrix. So the clauses in the DNF are  $A \wedge v_{12} > 0$ ,  $A \wedge v_5 > 0 \wedge \dots$  in the same sequence used in  $H$  above. The existential quantifiers applied to  $A \wedge \neg H$  can be distributed across the clauses to form seven smaller QE problems.

As of early 2018, Z3 could not determine whether the Tarski formula  $A \wedge \neg H$  is satisfiable: it was left running for several days without producing a result or

<sup>9</sup> Since the problem is fully existentially quantified we are free to use any ordering of the 10 variables; but there are  $10! = 3,628,800$  such orderings; so we have not tried them all.

error message. For this problem virtual substitution techniques may be advantageous (note the degree in any one variable is at most two). Indeed, REDLOG and MATHEMATICA can resolve it quickly. The problem would certainly be difficult for CAD based tools (which underline the theory solver used by Z3 for such problems). For example, if we take only the first clause of the disjunction (the one containing  $v_{12} > 0$ ), then the Tarski formula has twelve polynomials in twelve variables, and just three CAD projections (using Brown’s projection operator) to eliminate  $\{v_{12}, v_{11}, v_{10}\}$  results in 200 unique polynomials with nine variables still remaining to eliminate!

We note that the problem has a lot of structure to exploit in regards to the sparse distribution of variables in atoms. One approach currently under investigation is to eliminate quantifiers incrementally, taking advantage of the repetitive structure of the sub-problems one uncovers this way (an approach that works on several other problems in the benchmark set). This will be the topic of a future paper.

This example can be found in the benchmark set as Model #0056.

### 3 Analysis of the Structure of Economics Benchmarks

The examples collected for the benchmark set come from a variety of areas of economics. They were chosen for their importance in economic reasoning and their ability to be expressed in the Tarski framework, not on the basis of their computational complexity. The sentences tend to be significantly larger formulae than the initial examples described above. We summarise some statistics on the benchmarks. These statistics are generated for the 45 counterexample checks ( $\exists \mathbf{v}[A \wedge \neg H]$ ) which are representative of the full 135 problems.

#### 3.1 Logical size

Presented in DNF the benchmarks have number of clauses ranging from 7 to 43 with a mean average of 18.5 and median of 18. Of course each clause can be a conjunction but this is fairly rare: the average number of literals in a clause is at most 1.5 in the dataset, with the average number of atoms in a formula only slightly more than the number of clauses at 19.3.

#### 3.2 Algebraic size

The number of polynomials involved in the benchmarks varies from 7 to 43 with both mean average of 19.2 and median average of 19. The number of variables ranges from 8 to 101 with an mean average of 17.2 and median of 14. It is worth noting that examples with this number of variables would rarely appear in the QE literature, since many QE algorithms have complexity doubly exponential in the number of variables [15], [16], with QE itself doubly exponential in the number of quantifier changes [2]. The number of polynomials per variable ranges from 0.4 to 2.0 with an average of 1.2.

### 3.3 Low degrees

While the algebraic dimensions of these examples are large in comparison to the QE literature, their algebraic structure is usually less complex. In particular the degrees are strictly limited. The maximum total degree of a polynomial in a benchmark ranges from 2 to 7 with an average of 4.2. So although linear methods cannot help with any of these benchmarks the degrees do not rise far. The mean total degree of a polynomial in a benchmark is only 1.8 so the non-linearity while always present can be quite sparse.

However, of greater relevance to QE techniques is not total degree of a polynomial but the maximum degree in any one variable: e.g. this controls the number of potential roots of a polynomial and thus cell decompositions in a CAD (with this degree becoming the base for the double exponent in the complexity bound); see for example the complexity analysis in [4]. The maximum degree in any one variable is never more than 4 in any polynomial in the benchmarks. If we take the maximum degree in any one variable in a formula and average for the dataset we get 2.1; if we take the maximum degree in any one variable in a polynomial and average for the dataset we get 1.3.

### 3.4 Plentiful useful side conditions

With variables frequently multiplying each other in a sentence's polynomials, a potentially large number of polynomial singularities might be relevant to a QE procedure. However, we note that the benchmarks here will typically include sign conditions that in principle rule out the need to consider many such singularities. That is, these side conditions will often coincide with unrealistic economic interpretations that are excluded in the assumptions. Of course, the computational path of many algorithms may not necessarily exploit these currently without human guidance.

### 3.5 Sparse occurrence of variables

For each sentence  $\exists \mathbf{v}[A \wedge \neg H]$  (these are the computationally more challenging of the 135) we have formed a matrix of occurrences, with rows representing variables and columns representing polynomials. A matrix entry is one if and only if the variable appears in the polynomial, and zero otherwise. The average entry of all 45 occurrence matrices is 0.15. This sparsity indicates that the use of specialised methods or optimisations for sparse systems could be beneficial.

### 3.6 Variable orderings

Since the problems are fully existentially quantified we have a free choice of variable ordering for the theory tools. It is well known that the ordering can greatly affect the performance of tools. A classic result in CAD presents a class of problems in which one variable ordering gave output of double exponential complexity in the number of variables and another output of a constant size

[7]. Heuristics have been developed to help with this choice (e.g. [13] [5]) with a machine learned choice performing the best in [20]. The benchmarks come with a suggested variable ordering but not a guarantee that this is optimal: tools should investigate further (see also [28]).

## 4 Final Thoughts

### 4.1 Summary

We have presented a dataset of examples from economic reasoning suitable for automatic solution by QE or SMT technology. These clearly broaden the application domains present in the relevant category of the SMT-LIB, and the literature of QE. Further, their logical and algebraic structure are sources of additional diversity.

While some examples can be solved easily there are others that prove challenging to the various tools on offer. The example in Section 2.3 could not be tackled by REDLOG while the one in Section 2.4 caused problems for Z3. Of the tools we tried, only MATHEMATICA was able to decide all problems in the benchmark set. However, in terms of median average timing over the benchmark set MATHEMATICA takes longer with 825ms than REDLOG (50ms) or Z3 (< 15ms).

### 4.2 Ongoing and future work

As noted at the end of Section 2, some of these problems have a lot of structure to exploit and so the optimisation of existing QE tools for problems from economics is an ongoing area of work.

More broadly, we hope these benchmarks and examples will promote interest from the Computer Science community in problems from the social sciences, and correspondingly, that successful automated solution of these problems will promote greater interest in the social sciences for such technology. A key barrier to the latter is the learning curve that technology like Computer Algebra Systems and SMT solvers has, particularly to users from outside Mathematics and Computer Science. In [28] we describe one solution for this: a new MATHEMATICA package called **TheoryGuru** which acts as a wrapper for the QE functionality in the `Resolve` command designed to ease its use for an economist.

## References

1. C. Barrett, P. Fontaine, and C. Tinelli. The Satisfiability Modulo Theories Library (SMT-LIB). [www.SMT-LIB.org](http://www.SMT-LIB.org), 2016.
2. S. Basu, R. Pollack, and M.F. Roy. *Algorithms in Real Algebraic Geometry*. 10 of Algorithms and Computations in Mathematics. Springer-Verlag, 2006.
3. R. Bradford, J.H. Davenport, M. England, H. Errami, V. Gerdt, D. Grigoriev, C. Hoyt, M. Kořta, O. Radulescu, T. Sturm, and A. Weber. A case study on the parametric occurrence of multiple steady states. In *Proc. 2017 ACM International Symposium on Symbolic and Algebraic Computation, ISSAC '17*, pages 45–52. ACM, 2017.

4. R. Bradford, J.H. Davenport, M. England, S. McCallum, and D. Wilson. Truth table invariant cylindrical algebraic decomposition. *Journal of Symbolic Computation*, 76:1–35, 2016.
5. R. Bradford, J.H. Davenport, M. England, and D. Wilson. Optimising problem formulations for cylindrical algebraic decomposition. In J. Carette, D. Aspinall, C. Lange, P. Sojka, and W. Windsteiger, editors, *Intelligent Computer Mathematics*, vol. 7961 of *Lecture Notes in Computer Science*, pages 19–34. Springer Berlin Heidelberg, 2013.
6. C.W. Brown. QEPCAD B: A program for computing with semi-algebraic sets using CADs. *ACM SIGSAM Bulletin*, 37(4):97–108, 2003.
7. C.W. Brown and J.H. Davenport. The complexity of quantifier elimination and cylindrical algebraic decomposition. In *Proc. 2007 International Symposium on Symbolic and Algebraic Computation*, ISSAC '07, pages 54–60. ACM, 2007.
8. B. Caviness and J. Johnson. *Quantifier Elimination and Cylindrical Algebraic Decomposition*. Texts & Monographs in Symbolic Computation. Springer-Verlag, 1998.
9. A.E. Charalampakis and I. Chatzigiannelis. Analytical solutions for the minimum weight design of trusses by cylindrical algebraic decomposition. *Archive of Applied Mechanics*, 88(1):39–49, 2018.
10. C. Chen and M. Moreno Maza. Quantifier elimination by cylindrical algebraic decomposition based on regular chains. In *Proc. 39th International Symposium on Symbolic and Algebraic Computation*, ISSAC '14, pages 91–98. ACM, 2014.
11. G.E. Collins. Quantifier elimination for real closed fields by cylindrical algebraic decomposition. In *Proc. 2nd GI Conference on Automata Theory and Formal Languages*, pages 134–183. Springer-Verlag (reprinted in the collection [8]), 1975.
12. F. Corzilius, U. Loup, S. Junges, and E. Ábrahám. SMT-RAT: An SMT-compliant nonlinear real arithmetic toolbox. In A. Cimatti and R. Sebastiani, editors, *Theory and Applications of Satisfiability Testing - SAT 2012*, vol. 7317 of *Lecture Notes in Computer Science*, pages 442–448. Springer Berlin Heidelberg, 2012.
13. A. Dolzmann, A. Seidl, and T. Sturm. Efficient projection orders for CAD. In *Proc. 2004 International Symposium on Symbolic and Algebraic Computation*, ISSAC '04, pages 111–118. ACM, 2004.
14. A. Dolzmann and T. Sturm. REDLOG: Computer algebra meets computer logic. *SIGSAM Bull.*, 31(2):2–9, 1997.
15. M. England, R. Bradford, and J.H. Davenport. Improving the use of equational constraints in cylindrical algebraic decomposition. In *Proc. 2015 International Symposium on Symbolic and Algebraic Computation*, ISSAC '15, pages 165–172. ACM, 2015.
16. M. England and J.H. Davenport. The complexity of cylindrical algebraic decomposition with respect to polynomial degree. In V.P. Gerdt, W. Koepf, W.M. Werner, and E.V. Vorozhtsov, editors, *Computer Algebra in Scientific Computing: 18th International Workshop, CASC 2016*, vol. 9890 of *Lecture Notes in Computer Science*, pages 172–192. Springer International Publishing, 2016.
17. M. Erascu and H. Hong. Real quantifier elimination for the synthesis of optimal numerical algorithms (Case study: Square root computation). *Journal of Symbolic Computation*, 75:110–126, 2016.
18. P. Fontaine, M. Ogawa, T. Sturm, and X.T. Vu. Subtropical satisfiability. In C. Dixon and M. Finger, editors, *Frontiers of Combining Systems (FroCoS 2017)*, vol. 10483 of *Lecture Notes in Computer Science*, pages 189–206. Springer International Publishing, 2017.

19. J.R. Hicks. *Value and Capital*. Clarendon Press, 2nd edition, 1946.
20. Z. Huang, M. England, D. Wilson, J.H. Davenport, L. Paulson, and J. Bridge. Applying machine learning to the problem of choosing a heuristic to select the variable ordering for cylindrical algebraic decomposition. In S.M. Watt, J.H. Davenport, A.P. Sexton, P. Sojka, and J. Urban, editors, *Intelligent Computer Mathematics*, vol. 8543 of *Lecture Notes in Artificial Intelligence*, pages 92–107. Springer, 2014.
21. H. Iwane, H. Yanami, and H. Anai. SyNRAC: A toolbox for solving real algebraic constraints. In H. Hong and C. Yap, editors, *Mathematical Software – Proc. ICMS 2014*, vol. 8592 of *Lecture Notes in Computer Science*, pages 518–522. Springer Heidelberg, 2014.
22. G.A. Jehle and P.J. Reny. *Advanced Microeconomic Theory*. Pearson, 2011.
23. D. Jovanovic and L. de Moura. Solving non-linear arithmetic. In B. Gramlich, D. Miller, and U. Sattler, editors, *Automated Reasoning: 6th International Joint Conference (IJCAR)*, vol. 7364 of *Lecture Notes in Computer Science*, pages 339–354. Springer, 2012.
24. D. Jovanovic and B. Dutertre. LibPoly: A library for reasoning about polynomials. In M. Brain and L. Hadarean, editors, *Proc. 15th International Workshop on Satisfiability Modulo Theories (SMT 2017)*, number 1889 in CEUR-WS, 2017.
25. X. Li and D. Wang. Computing equilibria of semi-algebraic economies using triangular decomposition and real solution classification. *Journal of Mathematical Economics*, 54:48–58, 2014.
26. A. Marshall. *Principles of Economics*. MacMillan and Co., 1895.
27. C.B. Mulligan. *The Redistribution Recession: How Labor Market Distortions Contracted the Economy*. Oxford University Press, 2012.
28. C.B. Mulligan, J.H. Davenport, and M. England. TheoryGuru: A Mathematica package to apply quantifier elimination technology to economics. In: J.H. Davenport, M. Kauers, G. Labahn and J. Urban editors, *Mathematical Software – Proc. ICMS 2018*, vol. 10931 of *Lecture Notes in Computer Science*, pages 369–378. Springer, 2018.
29. L.C. Paulson. Metitarski: Past and future. In L. Beringer and A. Felty, editors, *Interactive Theorem Proving*, vol. 7406 of *Lecture Notes in Computer Science*, pages 1–10. Springer, 2012.
30. A. Platzer, J.D. Quesel, and P. Rümmer. Real world verification. In R.A. Schmidt, editor, *Automated Deduction (CADE-22)*, vol. 5663 of *Lecture Notes in Computer Science*, pages 485–501. Springer Berlin Heidelberg, 2009.
31. C. Steinhorn. *Tame Topology and O-Minimal Structures*, pages 165–191. Springer Berlin Heidelberg, 2008.
32. A. Strzeboński. Cylindrical algebraic decomposition using validated numerics. *Journal of Symbolic Computation*, 41(9):1021–1038, 2006.
33. T. Sturm and A. Weber. Investigating generic methods to solve Hopf bifurcation problems in algebraic biology. In K. Horimoto, G. Regensburger, M. Rosenkranz, and H. Yoshida, editors, *Algebraic Biology*, pages 200–215. Springer, 2008.
34. A. Tarski. *A Decision Method For Elementary Algebra And Geometry*. RAND Corporation, Santa Monica, CA (reprinted in the collection [8]), 1948.
35. Y. Wada, T. Matsuzaki, A. Terui, and N.H. Arai. An automated deduction and its implementation for solving problem of sequence at university entrance examination. In G.-M. Greuel, T. Koch, P. Paule, and A. Sommese, editors, *Mathematical Software – Proc. ICMS 2016*, vol. 9725 of *Lecture Notes in Computer Science*, pages 82–89. Springer International Publishing, 2016.
36. V. Weispfenning. The complexity of linear problems in fields. *Journal of Symbolic Computation*, 5(1/2):3–27, 1988.