CESifo WORKING PAPERS

7528
2019

February 2019

How to test technical conditions

Suehyun Kwon

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 7528*
Category 12: Empirical and Theoretical Methods

How to test technical conditions

Abstract

This paper identifies technical conditions that are common in contract-theory models or that matter for dynamic structural models but not necessarily static structural models. When possible, tests for the conditions are provided as well. Conditions include the Spence-Mirrlees condition, orthogonality, projection in vector spaces and equilibrium selection among others.

Keywords: Spence-Mirrlees condition, structural estimation, vector projection, de-trending, equilibrium selection.

Suehyun Kwon
Department of Economics
University College London
30 Gordon Street
United Kingdom – London, WC1H 0AX
suehyun.kwon@gmail.com

First draft: February 18, 2019

1 Introduction

This paper studies technical assumptions in contract-theory models together with econometric assumptions that can change estimates in dynamic structural models. The latter set of assumptions includes orthogonality in vector spaces as opposed to independence or covariance zero, de-trending without knowing the underlying structure of the equilibrium of a dynamic model, continuity of strategies or beliefs. Since each property pertains to a specific type of issue, I will describe the issue and provide a test when it can be tested.

For contract-theory models, there are a few key assumptions that most models make. For instance, in any adverse-selection model, agents' utility functions typically satisfy the Spence-Mirrlees condition, and the type distribution satisfies the monotone hazard rate. It was pointed out by Mirrlees in the 1970s that there is no commonly known distribution that satisfies both CDFC and FOSD, but when it comes to the Spence-Mirrlees condition and the monotone hazard rate, have they been tested? If so, which datasets satisfy these? The Spence-Mirrlees condition is a type of single-crossing condition; do we know in general which variables or datasets satisfy the single-crossing condition? What about single peak?

When it comes to game-theory models, they don't require as many assumptions on the utility function as contract-theory models. Game-theory models require belief hierarchy or common knowledge in most cases, but this is a different type of modelling assumption. Furthermore, many adverse-selection models have quasilinear utilities, linear with respect to transfers, and many moral-hazard models have additively separable utility with respect to consumption and the cost of effort. Both adverse selection and moral hazard require additive separability of money and allocation/effort, then whether each component is linear, convex or concave needs to be specified.

Once the standard assumptions are given, one can ask which datasets satisfy these assumptions, but one can also ask why other branches of economics don't make the same set of assumptions. I will focus on the test of these assumptions in this paper, but it is worthwhile to point out what these assumptions imply and which branches of economics make different sets of assumptions for what reason.

First of all, most of risk-averse agents exhibit wealth effect, except in cases as CARA utility. The additive separability in contract-theory models implies that the wealth effect is essentially shut down; with moral hazard, the consumption utility can be concave, but the disutility of working has no wealth effect. This is in contrast to risk-averse agents in macro.

Second, quasilinear utility in adverse selection also implies that agents are risk-neutral with respect to money. Transfers have zero welfare effect within a population. Again, this is in contrast to standard macro models.

Leaving aside the difference from macro papers, a few obvious properties of utility functions in contract theory are (i) additive separability in money and allocation/effort, (ii) risk neutrality with respect to money in adverse selection, (iii) the Spence-Mirrlees condition, or more generally single-crossing property, and (iv) the monotone hazard rate property. Moral hazard models also typically assume monotone likelihood ratio property, but given that this is more commonly satisfied by well-known distribution functions, I will not focus on matching the right distribution to the dataset among well-known distributions. However, when one doesn't know the shape of the distribution, this still requires some discussion as to what one can assume about the structure of a randomly given dataset; consider Taylor expansion for continuous functions in real numbers.

I should mention that I'm not designing an experiment, whether it is a lab experiment, a field experiment, or an RCT. I will also not discuss data collection methods.

Among the four properties mentioned above, how can one test additive separability, risk neutrality for money and the Spence-Mirrlees condition? The monotone hazard rate is with respect to the distribution of types, which in the end is the belief of the mechanism designer. Even additively separability and risk neutrality for money are related to decision theory papers. I will discuss when one can confirm the properties from the dataset if one cannot design a test in the process of generating this dataset, but the real challenge is the

Spence-Mirrlees condition. How can one test the Spence-Mirrlees condition on a dataset? To see why these tests might be relevant, consider empirical papers studying contracts. It is unlikely that a dataset obtained from some company will include tests of these properties on the utility functions of individuals in the dataset. With a big company, it might be possible to test on a fraction, but how feasible is it to obtain a CEO-level dataset together with tests on their utility functions? But if we don't know whether the premise of the theory is validated on the dataset, can we go ahead and estimate parameters?

When it comes points (i) and (ii), there's a difference between any expected-utility maximizer exhibiting additive separability and risk neutrality and there being at least one utility function exhibiting both and explaining the dataset. This is related to implementation literature that requires every equilibrium to have the same property or robust mechanism that requires the same property with any prior.

With single-crossing properties, if one knows the function is differentiable and knows the first-partial derivatives, then one can test with the usual inequality for the Spence-Mirrlees condition. One can detect any violations from the dataset, but just because the current dataset doesn't violate the conditions, it doesn't mean the conditions are satisfied; for example, if the data points are sufficiently apart from each other, when can one detect discontinuity?

As for econometric assumptions, each condition I characterize is about a specific issue rather than all results in this category belonging to the same class of models. I discuss that orthogonality and the order of projection in vector spaces matter for dynamic structural models; within the same period, counterfactual or extrapolation doesn't depend on which one of non-orthogonal independent variables picks up the effect in its estimate. One can split the independent variables as sums of orthogonal vectors, and as long as the final objective is linear in all orthogonal components of independent variables, the final objective is the same. It doesn't quite work the same in dynamic models or with any objective function that is not globally linear. It doesn't work when some independent variables are endogenous and they have different costs.

I also provide a set of conditions on when one can test some of the technical conditions that are typically assumed instead of being tested; these include i.i.d., endogeneity and the length of Markov chains. Conditions on local linearity and continuity are also provided.

The last set of results are on equilibrium selection and what commonly used estimators implicitly assume about the equilibrium of the sample data.

The rest of the paper is organized as follows. Instead of setting up one model to point out every issue, I will point out each issue and discuss relevant tests in section 2. Section 3 concludes.

2 Results

I first discuss the Spence-Mirrlees condition in section 2.1; other results are in section 2.2.

2.1 Spence-Mirrlees Condition

Formally, the Spence-Mirrlees condition itself doesn't require quasi-linear utility fuctions. However, it effectively requires an one-dimensional type space for each agent. When the utility function is twice differentiable, the condition requires that the marginal rate of substitution increases with the type at every (allocation, transfer); it also requires the ordering of types to be identical for every pair of allocation and transfer. When the utility function is not twice differentiable, or more generally, one can test for the single-crossing property for (allocation, transfer, type) directly. As mentioned earlier, one of the problems is that just because the points in the current dataset doesn't violate the single-crossing property, it doesn't necessarily mean that the underlying utility function does satisfy the single-crossing property. Fewer the data points, bigger the problem.

There are a few possibilities here; one can try to show that there exists at least one utility function consistent with the data and satisfies the single-crossing property. One can ask whether every utility function consistent with the data should satisfy the single-crossing property, and if it requires a few more assumptions, what are the additional assumptions? Or one can just show that one cannot reject the hypotheses that the utility function does not satisfy the single-crossing property. Regardless, the "effective" one-dimensional type space must be satisfied and can be used to reject the single-crossing property.

In the textbook adverse-selection models, they typically assume one-dimensional type space. However, if one were to think about what one knows from any dataset on contracts, it involves different categories. It might be an obvious step, but for the purpose of satisfying the Spence-Mirrlees condition so that the usual approach for static adverse selection models works, it is sufficient to have an one-dimensional ordering that is independent of any allocation and transfer.

If one were to think more seriously about where the Spence-Mirrlees condition is used in the proof of static adverse-selection models, there are four types of constraints with a continuum of types which one can reduce to local IC constraints at each type, IR constraint for the lowest type, and the monotonicity condition on the allocation. Then one can solve for the relaxed problem without the monotonicity condition and verify that the optimal contract satisfies it. This actually suggests that if the marginal rate of substitution increases with the type at every (allocation, transfer) in the optimal contract, the proof should work. As for the reason why this is not the same as requiring the Spence-Mirrlees condition at every pair of (allocation, transfer) even though the optimal menu is continuous in type, the optimal menu is a one-dimensional subset of $\mathbb{R}^2$.

Suppose one has a dataset. It is unlikely that the dataset contains the marginal rate of substitution for every type at every option on the menu. When can one test whether the current contract is optimal? If we don't know whether the contract is optimal, can we test for the Spence-Mirrlees condition from the existing dataset? Can we come up with an optimal contract based on estimates? What assumptions need to be made to satisfy the Spence-Mirrlees condition for the optimal contract we find?

Theorem 1. *In order for a static contract to be optimal, it is necessary that the agent's types, ordered by the marginal rate of substitution between allocation* *and transfer, must be identical for every (allocation,transfer) in the contract.*

One cannot reject the hypothesis that there exists at least one utility function that satisfies the Spence-Mirrlees condition and is consistent with the data if the ordering of types by the marginal rate of substitution between allocation and transfer given in the data is identical at every option in the contract.

In order to guarantee that every utility function consistent with the data makes the contract optimal, the ordering of types by the marginal rate of substitution at the options in the contract has to be identical.

In order for every utility function consistent with the data to satisfy the Spence-Mirrlees condition, one needs to assume the ordering of types by the marginal rate of substitution is identical at every feasible (allocation, transfer).

2.2 Other Conditions

This section aims at providing mathematical foundation for certain standard assumptions in econometric theory. Before I list the topics, let me be clear that "independent" as in the sense of independent variables vs. dependent variables in a regression is completely unrelated to "independent"-ly distributed random variables; it will become clearer once I describe the problem in more detail. The issues I address include independence of random variables, orthogonality and the order of projection in a vector space, continuity and extrapolation, existence of unique equilibrium of structural models among others. As is always the case, if one only points out a problem without an alternative, it doesn't improve the outcome necessarily, but an immediate consequence of getting the endogeneity wrong with extrapolation or finding counterfactuals is that independent variables can be endogenous and individuals in the population can invest in those; getting the orthogonality wrong doesn't hurt extrapolation within a given period, but together with knowledge dissemination, one can easily mislead the population into investing in one independent variable over another.

This is a preliminary draft, and the longer version will include tests on when random variables are independent and orthogonal, respectively. The first theorem is on orthogonalization of independent variables. If the dataset includes two variables that one might guess are independent but have not tested, when can one include these as independent variables? When can one be confident about the validity of an instrument? Typical curriculum includes a math camp then a probability course followed by OLS, IV and so forth. However, if one does not check why each technical condition matters for the estimator, merely running a STATA code or assuming conditions must be met doesn't lead to the correct estimates. Furthermore, if with each estimator, one cares so much about the bias and the speed of conversion, why checking technical conditions on what are supposed to be independent variables doesn't get as much attention? With vectors in $\mathbb{R}^n$, how many data points does one need to confirm the hypothesis that these variables are independent? Orthogonal? If the tensor is non-zero, does it matter for estimators? Which ones are affected the most?

Theorem 2 (Independence and Orthogonality of Random Variables in $\mathbb{R}^n$)**.** *With any estimator that assumes local linearity and continuity in the neighborhood of the estimates, one can always decompose independent variables as sums of orthogonal vectors, and as long as the objective function is linear, any counterfactual or extrapolation within the same period doesn't depend on how one orthogonalizes independent variables.*

The axes of orthogonalization also doesn't matter once certain types of affine transformation are allowed between sets of axes. One can alternatively find all estimates, which will be infinitely many but can be represented as a set.

Any proof that is not provided immediately after each theorem is in the appendix.

After orthogonality, the next issue I want to point out is correlation across random variables. To be more precise, getting the independent variables and dependent variables correct or the other way around, i.e., assuming that dependent variables are independent, is one of the central issues in any applied work, and I don't need to emphasize it much further. However, whether to assume certain variables are i.i.d. can benefit from further discussion. Particularly, in a static model, this is more of an assumption that one needs to make about the data-generating process than one that can be tested on the current dataset. Independently and identically distributed by definition means that there is a theoretical probability that N data points will be the one we have in the data. This means that we need to assume both the underlying distribution itself and the data-generating procesd to be i.i.d. However, when the dataset is dynamic, i.i.d. now means not only across different agents or categories but also across different time periods, which makes it easier to test. One can compute the theoretical probability as before with a given distribution, but the time dimension gives more information that one can potentially test without assuming a particular distribution to start with.

Theorem 3 (Test for Data-generating Process)**.** *Any data-generating process requires both the underlying distribution and the correlation structure across data points. Any dynamic data-generating process requires the correlation across the time-dimension as well, i.e., fully persistent, partially persistent or i.i.d..*

Without assuming any of the underlying distribution, correlation within period and persistence across different time periods, one can take the dataset as the distribution itself which maximizes the fit, but which one of the three should be assumed and at which point one should consider this might not be such a small-probability event and one of three assumptions might be wrong needs further discussion.

Next, one can also ask whether any of the independent variables are endogenous in a dynamic structural model. De-trending without knowing the equilibrium strategies is a related issue that also deserves some attention. But before discussing any particular equilibrium and strategies, endogeneity of independent variables doesn't matter in a static model. However, it does matter in a dynamic setting; it makes a difference when a variable is exogenously given or endogenous. Again, this has been discussed in macro models but not to the extent of discussing a particular equilibrium strategy and what to do with multiple equilibria, particularly if there is multiplicity for the same payoff vector. Both theorem 3 and the discussion on endogeneity should be taken into account in robustness checks, but things can be far more subtle in a dynamic model, and yet at the same time, one can test better as well.

A related issue to the endogeneity of variables is the length or the order of Markov chains, e.g., $AR(1)$, $AR(2)$, $\cdots$, $AR(k)$. Ideally, one should be able to test both the endogeneity and the length of the Markov chain together; how many periods of each player's (private) history matter for the parameters of the continuation game. What I'm referring to as paremeters, which are typically just exogenously given state variables, are sometimes called the payoff-relevant states or the state of the world in theory models. Also, what I'm referring to as the continuation game can be defined with respect to any equilibrium notion of repeated games or stochastic games. But one can only estimate with respect to a particular equilibrium, and this is where de-trending without knowing the underlying equilibrium I mentioned becomes relevant.

The last few paragraphs are all related to the issue of what assumptions can be made without loss of generality; theorem 2 is more about the estimated values with commonly used estimators, which as mentioned already doesn't matter for counterfactuals or extrapolating within the same period with local linearity. Theorem 2 also starts to matter in a dynamic model. Overall, the approach I'm taking in section 2.2 is related to Bergemann and Morris (2013), but their paper in the end is about Bayes correlated equilibrium.

The next set of results are more about the equilibrium one is estimating, as opposed to estimators or data-generating processes as in theorem 2. The few paragraphs above after theorem 2 can be considered as technical assumptions in the model, but it is equally likely be an equilibrium strategy of a dynamic game. I was referring to equilibrium selection and continuation games earlier, but the particular equilibrium players in the given dataset is playing matters in a static model as well. To be more precise, data-generating processes for exogenous variables, heterogeneous characteristics, private information and equilibrium strategies all should be taken into account in any applied work, but multiplicity of equilibria and in particular, multiple equilibria that all correspond to the same payoff vector in any repeated game or stochastic game have not gotten enough attention in structural models. By definition, reduced form abstracts away from many of the issues I mention here, and if the estimates don't depend on the underlying model, multiplicity of equilibria or equilibrium selection, then structural estimation with robustness checks will yield the same estimates at the end of the day; it is just a different way of estimating the parameters.

As for the prevalence of multiplicity of equilibria, think of different versions of folk theorem. In most cases, the statement is with respect to players being infinitely patient, and there are estimated discount factors, particularly in most macro models, but a fixed discount factor still doesn't mean unique equilibrium necessarily. Robustness when it comes to equilibrium selection is related to implementation in mechanism design literature, but if multiple equilibria corresponding to the same payoff vector can lead to different estimates of paremeter values, one needs to be more careful.

Before discussing how to select an equilibrium, find the one players are playing, or in some cases how to test who are the players of the game, whether players know the equilibrium they are playing, why any outsider must know the equilibrium itself, I will discuss a few more technical assumptions that one needs from the equilibrium strategies in order to estimate anything. As mentioned, robustness checks at the end of any applied paper should address at least some of the issues I point out here, but unfortunately, testing a few more estimators with or without all of the independent variables doesn't address all the issues. R^2 and *** don't check everything either.

An alternative way of paraphrasing the previous paragraph is some papers make certain assumptions for the lack of an alternative, but these are not stated as explicitly in the paper. First of all, in order to estimate a parameter, one needs (local) continuity. Without continuity, counterfactual or extrapolation doesn't make sense; or one could plug in the estimates in a "mathematical" equation, but these equations are supposed to represent something, and I yet need to see an equation for counterfactual or extrapolation that is discontinuous at the point of estimated values. If these equations are data-generating processes or exogenous, one just needs to assume, or prove, continuity. However, if these objects are strategies or beliefs, as they should be at least in some intermediate steps of dynamic structural models, then one needs to check when beliefs and strategies are continuous. I will first state the conditions on exogenous variables then discuss beliefs and strategies in the paragraph after.

I mentioned local continuity in the paragraph above, but in reality, most estimators require local linearity. In case anyone has forgotten, any estimator that uses matrix inversion in its proof assumes local linearity.

Theorem 4 (Which Estimators Assume Local Linearity)**.** *Pretty much any commonly used estimators require local linearity near the point of estimation including OLS and IV.*

Local linearity is satisfied with any function that is differentiable at that point but not with any continuous function.

When it comes to beliefs, any discrete signals or pieces of information change the belief discontinuously. One can hope to estimate the time when a player receives a signal and what is the value of the signal, but the beliefs themselves are discontinuous. If the econometrician knows the information structure, then one can still estimate the beliefs before and after the jump, but now the estimator needs to account for discontinuity. For example, when the belief is an independent variable in some intermediate step, e.g., for the strategy as dependent variable, one could do regression discontinuity, but if the beliefs are private information, then one needs to estimate the timing, signal, beliefs before and after the signal arrives, and then the effect on the dependent variable. When it comes to strategies, there is no reason why any player should be playing a pure strategy in a repeated game, and with mixed strategies, a realized action is by definition discontinuous. Then there is another layer of whether each mixed strategy should be continuous in the parameter. Certain models yield more structures in any equilibrium strategy, and global games are one such example, but the number of dynamic global games papers or dynamic games with supermodularity are not many. In this class of models, one at the very least knows that strategies are monotone with respect to the parameter, and one could get continuity depending on further technical assumptions in payoffs. Given that these are all independent from the issue of equilibrium selection and yet relevant for any equilibrium, these need to be resolved first.

Theorem 5 (Continuity Conditions for Beliefs and Strategies)**.** *A player's belief is discontinuous with any discrete signal. Continuous signals include Brownian motion.*

A player's equilibrium strategy need not be continuous in any primitive of the model. However, it is monotone in any supermodular game and continuius with further assumptions in supermodular games.

There is still one more issue one can and needs to address before getting into equilibrium selection. De-trending is a common step in time series data, and one can think about the usual ways of treating panel data, but there could be potential complications. I should point out that some assumptions in structural models or macro models have been independently tested, and the assumptions, if not the values of parameters, are borrowed from other papers. I do not attempt at arguing about every single assumption, particularly when they have been tested, but in some cases, one needs to be a bit careful; theorem 2 for example addresses an issue of potentially getting the estimate wrong, which doesn't matter for counterfactuals or extrapolation within the paper but can't be taken out as an estimate out of context.

Theorem 6 (De-trending before Equilibrium Selection)**.** *De-trending of time-series data on valuation is the same as focusing on private values of players and leaving out the common-value part.*

However, if the time-series data is not on types or values and instead on beliefs and strategies, de-trending without knowing the equilibrium strategies can cause a problem.

One can de-trend without equilibrium selection if the time-series data is on strategies that are globally linear in players' types.

In the final version of this paper, there will be one more theorem on equilibrium selection, including what are implicitly assumed in commonly-used estimators and what one can further estimate by making an assumption on the equilibrium strategy.

3 Conclusion

I characterized technical conditions that matter for (dynamic) structural models and are often assumed instead of being tested. The first set of conditions are on standard static adverse selection models, and in particular, on the Spence-Mirrlees condition. Theorem 1 characterizes when one can test whether the data is consistent with the Spence-Mirrlees condition instead of just assuming the contract in the data must be optimal.

The second set of results are more on econometric assumptions, consisting of three sub-types. The first subcase is on orthogonalizing independent variables in a regression, which doesn't matter locally in the same period but matters in a dynamic model unless the objective function is globally linear, none of the independent variables are endogenous and so forth. The second subcase is more on standard modelling assumptions including i.i.d., endogeneity, the length of Markov chains, local linearity, and continuity or monotonicity, particularly with supermodularity. Lastly, the third subcase is equilibrium selection, whether it be implicit assumptions underlying common practice or what one could further estimate by making an additional assumption.

Some of these conditions might be known to anyone who's worked on dynamic structural models. However, when an assumption is made for technical reasons or for the lack of alternative, they are not often stated explicitly so in the paper, and not every paper tests every technical assumption that could potentially be tested. Some of the results can also help in testing whether an assumption in the existing paper can be dispensed with, i.e., it can be confirmed in the data instead of being assumed away.

I also hope these results can shed light on when one can take an estimate from a existing paper and use as an exogenous parameter.