# International conflict and strategic games: challenging conventional approaches to mathematical modelling in International Relations

## *Conflitos internacionais e jogos estratégicos: desafios às abordagens convencionais de modelagem matemática em Relações Internacionais*

DOI: 10.21530/ci.v14n1.2019.865

Enzo Lenine[1]

**Abstract**

The pervasiveness of international conflict makes of it one of the main topics of discussion among IR scholars. The discipline has extensively attempted to model the conditions and settings under which armed conflict emerges, at sometimes resorting to formal models as tools to generate hypotheses and predictions. In this paper, I analyse two distinct approaches to formal modelling in IR: one that fits data into mathematical models and another that derives statistical equations directly from a model's assumption. In doing so, I raise the following question: how should maths and stats be linked in order to consistently test the validity of formal models in IR? To answer this question, I scrutinise James Fearon's audience costs model and Curtis Signorino's strategic interaction game, highlighting their mathematical assumptions and implications to testing formal models. I argue that Signorino's approach offer a more consistent set of epistemological and methodological tools to model testing, for it derives statistical equations that respect a model's assumptions, whereas the data-fit approach tends to ignore such considerations.

**Keywords:** Formal Modelling; Empirical Testing; International Conflict; Audience Costs; Strategic Interaction Games.

1 Enzo Lenine Nunes Batista Oliveira Lima is Professor of International Relations at the Institute of Humanities of the University of International Integration of the Afro-Brazilian Lusophony (UNILAB/Malês). His works focus primarily on theory and methodology, hierarchies of knowledge, mathematical modelling and international conflict.

Artigo submetido em 13/11/2018 e aprovado em 08/04/2019.

## Resumo

A prevalência dos conflitos internacionais faz deste um dos principais tópicos de discussão entre os acadêmicos de Relações Internacionais. A disciplina tem tentado extensivamente modelar as condições e configurações sob as quais o conflito armado emerge, às vezes recorrendo a modelos formais como ferramentas para gerar hipóteses e previsões. Neste artigo, analiso duas abordagens distintas para a modelagem formal em RI: uma que encaixa dados em modelos matemáticos e outra que deriva equações estatísticas diretamente das premissas do modelo. Ao fazê-lo, levanto a seguinte questão: como a matemática e a estatística devem ser vinculadas para testar consistentemente a validade dos modelos formais em RI? Para responder esta pergunta, examino o modelo de custos de audiência de James Fearon e o jogo de interação estratégica de Curtis Signorino, destacando suas suposições matemáticas e implicações para testar modelos formais. Argumento que a abordagem de Signorino oferece um conjunto mais consistente de ferramentas epistemológicas e metodológicas para testar modelos, uma vez que deriva equações estatísticas que respeitam as premissas do modelo, enquanto a abordagem de ajuste de dados tende a ignorar tais considerações.


**Palavras-chave:** Modelagem Formal; Teste Empírico; Conflito Internacional; Custos de Audiência; Jogos de Interação Estratégica.

## Introduction

Studies of armed conflicts date back to ancient times, even when International Relations was not known as a distinct field or discipline. Thucydides' account of the Peloponnesian War is perhaps one of the oldest texts dealing with the implications of military conflict under a realist perspective. However, it was in the 20th century that IR thrived as a discipline of its own, becoming known for its intense theoretical debates about the nature of the international system and its effects in the prospects of war and peace. Anarchy characterises the international arena, and the absence of central authority may lead states towards paths of conflict or cooperation.

The theoretical debates in IR attempt to explain state behaviour based on models of power, decision and cooperation. Hans Morgenthau's Politics Among Nations (2003) presents the balance of power model, which underlies the realist theory of IR, and has become one of the most pervasive explanations for state interactions in the international arena. Robert Keohane's and Joseph Nye's Power and Interdependence: World Politics in Transition (2011) offers a more cooperative model of state interaction, being a classic of IR's neoliberal theories. However, these models are not formal in the sense that they contain mathematical expressions, theorems, propositions. Balance of power and complex interdependence are rather discursive constructs, often connected to historical assessments of state behaviour.

Formal modelling per se can be attributed to Lewis Richardson's arms race model (1960) and Thomas Schelling's deterrence model (1960). They allowed further improvement and advances in the literature of international conflict, stimulating the designing of more accurate formal models and the subsequent testing of these models. Furthermore, the construction of datasets on conflicts provided scholars with tools for assessing the validity of their models and the hypotheses they generate.

Most models borrow their assumptions and methodological procedures from Rational Choice Theory (henceforth, RCT): they frequently assume states are rational unitary actors and utility-maximizers. Game theory is the commonest approach to modelling international conflict and cooperation, for they presuppose bargaining, which is more efficiently represented by game-theoretical settings. As one would expect, such RC-oriented approach generates criticisms within the scholarship, with researchers questioning the empirical validity of formal models.

Testing a model in terms of its empirical value is a hard task. There is a tension between fitting data into the model without previous derivation of proper equations, on the one hand; and devising statistical tests directly from the mathematical model, on the other hand. This issue is of uttermost importance if one is willing to assess the explanatory power of a model. The literature addresses empirical testing in different ways, reaching, as consequence, distinct conclusions about a model's validity. There is no straightforward, unique answer to the question about how one should devise empirical tests of formal models, and one of the goals of this paper consists in discussing the different approaches taken by designers of formal models. Many researchers prefer to conduct empirical tests separately, as in data-fit models: building a mathematical model and then checking for statistical significance or historical examples. This procedure opens doors to a variety of questionings on selection bias, proper representation of mathematical assumptions, etc. More recently, some political scientists have devoted efforts towards direct derivation of statistical equations from the model, respecting its mathematical assumptions whenever possible. Computational simulations aid this endeavour by providing a setting where the model can be tested by real-world and computer-generated data.

That said, I propose the following puzzle: how should maths and stats be linked in order to consistently test the validity of formal models in IR? I argue that statistical tests derived directly from the mathematical model provide firmer validity. For the derivation process respects the structure of the model. Throughout the remainder of the paper, I shall scrutinise two examples of both approaches and their epistemological consequences to formal modelling in IR. James Fearon's audience costs model and Curtis Signorino's strategic interaction game will be analysed in depth in order to unravel their underlying rationales.

The paper is divided into four sections. The first discusses the literature on audience costs that has thrived after the publication of Fearon's paper in the American Political Science Review (APSR). It has mostly focused on data-fit tests to test the model. The second section discusses Signorino's extrapolative model of strategic interaction, and the implications to model testing in political science and IR. Finally, the last section sums up the lessons taught by both approaches and assesses their advantages and disadvantages in respect to the empirical testing of models.

## Fitting data into models: audience costs and the crisis game

Since the publication of James Fearon's article in APSR in 1994, the research agenda on international crises has been developing further tests of the audience costs model. As Fearon describes:

> I characterize crises as political contests with two defining features. First, at each moment a state can choose to attack, back down, or escalate the crisis further. Second, if a state backs down, its leaders suffer audience costs that increase as the crisis escalates. These costs arise from the action of domestic audiences concerned with whether the leadership is successful or unsuccessful at foreign policy (FEARON, 1994, p. 577).

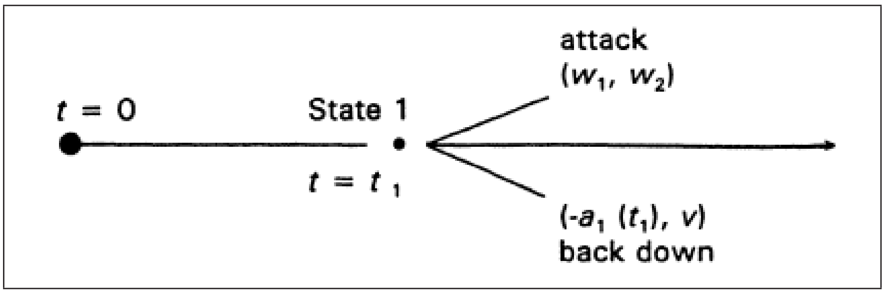
In other words, a leader facing an international crisis (either economic or political) has to deal simultaneously with the complex decision-making process entailed in the crisis itself and domestic reactions in favour or against her performance. The audience costs theory has become pervasive in a variety of fields, such as military crises, economic sanctions, alliances, foreign trade, etc. (TOMZ, 2007). Fearon justifies his game-theoretical approach to the problem by stating that "the major benefit of the formal analysis is a set of comparative statics results that provide insights into the dynamics of international disputes" (FEARON, 1994, p. 577). The international crisis game is framed as follows:

> States in a dispute thus face a dilemma. They have strong incentives to learn whether there are agreements both would prefer to the use of force, but their incentives to misrepresent mean that normal forms of diplomatic communication may be worthless. I argue that international crises are a response to this dilemma. States resort to the risky and provocative actions that characterize crises (i.e., mobilization and deployment of troops and public warnings or threats about the use of force) because less-public diplomacy may not allow them credibly to reveal their own preferences concerning international interests or to learn those of other states. (FEARON, 1994, p. 578)

The main argument underlying the model is as the crisis escalates, audience costs increase, forcing the leader to demonstrate/signal resolve. In democracies, this effect tends to be exacerbated, for the leader must be responsive to the public.

The international crisis game has a simple game tree. The crisis unfolds in continuous time, starting at t = 0. Each point in time constitutes a node where player 1 can choose either to attack, to quit or to escalate. If either player attacks before the other quits, each receives her own expected utilities; if a player quits before the other has quit or attacked, she suffers audience costs, which display linear behaviour (I shall discuss the implications of linearity when analysing Signorino's works) in Fearon's model. The model also sets a time horizon ($t_h$) where war is inevitable, and it is a function of increasing audience costs. The crisis game is depicted in figure 1.

**Figure 1: International crisis game**




Source: FEARON, 1994.

Fearon derives two lemmas and three propositions to solve for the equilibrium in the incomplete information game. The model indicates that there exists a variety of equilibria up to t*, which is the limiting horizon before any player decides to attack. Fearon describes the equilibrium as a war of nerves, based on expectations towards making quiet concessions or escalating and eventually waging a war. As time passes by, however, audience costs increase linearly and $t_h$ is reached. Escalation constrains the courses of action available, making it difficult for a state to back down. Furthermore, the probability density functions, which represent players' initial beliefs, play an important role in defining the outcomes of the game, for they entail the observable capabilities and the interests of each player.

Two questions could be raised about Fearon's model. The first one concerns the very existence of audience costs. The second, assuming that audience costs exist, refers to the behaviour of the $a_i(t)$ function, which is assumed to be linear in the original model. The literature has dealt extensively with the first question, yet there are many contentious issues in that debate. The matter of the linear function might sound as a mathematical technicality, but it offers a window of opportunity for testing the model. If audience costs exist and can be measured, one can collect data points, run a curve-fit model and assess how it changes the equilibrium. Curiously, Fearon (1994) does not provide any explanation why he has chosen the linear form – which would make us assume he did so for matters of mathematical simplicity, yet this is not clear in his work.

Works that followed the publication of Fearon's article have praised his model and have attempted to test its outcomes by deriving hypotheses and fitting data into classical statistical tests. Eyerman and Hart (1996), for example, attempted to test Fearon's model using a Poisson test, and measures of democracy as a proxy to audience costs. Their interest was tightly tied to the theory of democratic peace, which lacks, in their view, a compelling explanatory mechanism. They use SHERFACS phase-disaggregated conflict management dataset to test Fearon's hypothesis, announced as: "the only way to test his hypotheses is to observe the behaviour of democracies and nondemocracies within crises" (EYERMAN; HART, 1996, p. 603). The Poisson model assumes the form of Eq. (1):

$$\textit{Phase Count} = f(\textit{joint democracy, enemies, allies, ethnicity, territory, antagonism}) \quad (1)$$

It is not the goal of this paper to reproduce their findings, but rather what they have not found: any proof of the existence of audience costs. They state: "It appear that bloc dynamics (...) serve to aid communication. Fearon (1994) suggests that this communication may stem from international audience costs in addition to domestic audience costs but that they might be secondary concerns"

(EYERMAN; HART, 1996, p. 611). Eyerman and Hart repeat a similar statement in their conclusion, even though they have not tested for audience costs. Apparently, they assume it is the natural explanation that follows from the outcomes of the Poisson model, yet as it was not derived directly from Fearon's model, one can cast doubts whether the model was correctly specified to suggest the existence of audience costs. Furthermore, as Partell and Palmer (1997, p. 395) point out: "[T]he use of a state's democratic status is problematic because audience costs can be incurred by undemocratic states as well".

In order to solve the flaw in Eyerman and Hart's model, Partell and Palmer (1999) use institutional constraints as a proxy to measure audience costs. They affirm that "the more a leader is constrained in her ability to implement policy on her own, the more reliant she is on others for her position of authority, and thus the more likely it is that she can be removed from office if she fails to perform her duties to the satisfaction of others in the political system" (PARTELL and PALMER, 1999, p. 395). As Fearon's model is based on a principal-agent relationship, where the principals are the voters in democracies and high-ranking generals in most dictatorships, it sounds reasonable to measure audience costs in this way. Nevertheless, the existence of audience costs is assumed and Partell and Palmer fail to make a strong case of why their proxy actually measures audience costs. A measure of audience costs would be more closely related to Tomz's (2007) experiment, which attempts to assess the existence of audience costs based on public opinion surveys. If Tomz is right, the existence of audience costs may be a case solved, but how they generate outcomes is still an open question.

The common feature in the aforementioned works concerns the disconnection between audience costs and the statistical test performed. Authors focused on the outcomes of the model rather than on audience costs, for the tests they had designed were based on data about phases in a crises and measures of democracy (such as Polity and Freedom House). They assume democracies necessarily entail audience costs, never questioning the relevance of foreign policy to the audience. In terms of methodological precision, there is no solid argument to believe that the assumption of audience costs is correct. As Gartzke and Lupu suggest:

> [T]his literature is primarily concerned with testing an implication of Fearon's model, that is, that democracies fare better in certain crisis situations. Yet this implication largely rests on Fearon's assumption that democracies have 'stronger domestic audiences'. If this assumption is incorrect, then there is reason to doubt the specific processes posed in Fearon's model. (GARTZKE; LUPU, 2012, p. 393)

Summing up, Fearon's model could be tested for the existence and the functional form of the audience costs relationship to time. It is tempting to accept Tomz's (2007) findings and Gartzke and Lupu (2012) make an important point about experiments being useful to unravel the mechanisms in play. Nevertheless, at the current state, Fearon's model has only been tested in respect to its outcomes. To be sure, none of the tests performed by the aforementioned authors was strictly derived from the mathematical model. They used data generated in exogenous research contexts and attempted to fit them into the mathematical model. This procedure casts doubts over the validity of those tests – critics of RC models could argue that positive results that corroborate a model's assumptions are just what one would expect from a biased selection of cases (GREEN; SHAPIRO, 1994). In order to avoid such criticisms, one needs to check for the empirical validity of a model's assumptions – meaning that the audience costs assumption should be tested for its existence and linear behaviour – and derive a statistical model directly from the mathematical one.

## Designing structure-oriented tests: the international interaction game

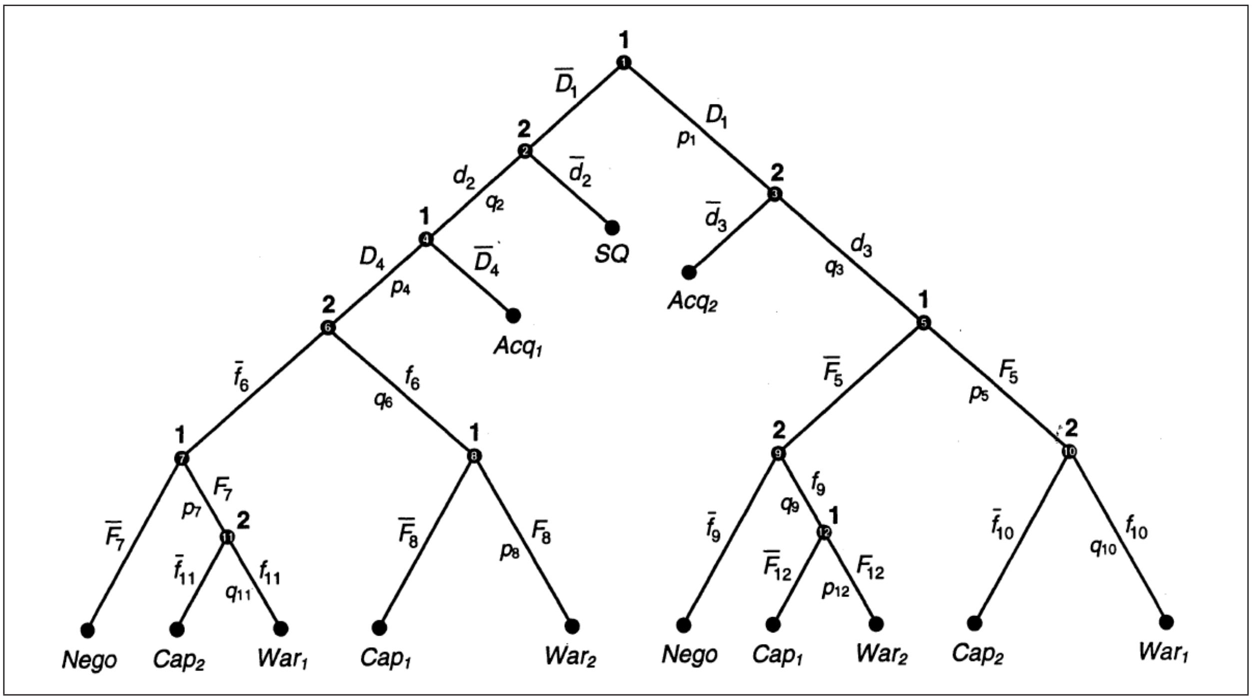
Modelling and testing international conflict is a hard task that demands the construction of a representative game and the derivation of adequate equations to build a bridge between mathematical assumptions and statistical tests. This is precisely where Curtis Signorino's approach offers a different perspective over model testing. Building on Bruce Bueno de Mesquita's and David Lalman's (1992) War and Reason, Signorino attempts to provide a mathematical-statistical framework to test game-theoretical models of strategic interaction in international relations.

Bueno de Mesquita and Lalman (1992) aimed at explaining why states wage wars knowing that they are a costly and risky endeavour. Instead of tackling the problem through the lenses of realist and neorealist accounts of international relations, they resort to formal modelling as a means to directly, clearly and unambiguously state their assumptions (BUENO DE MESQUITA; LALMAN, 1992, p. 21). In addition, they perform statistical tests of the model and examine historical narratives about specific conflicts in their dataset. However, they justify their use of models based on the following reason:

> We model because we believe that how we look at the facts must be shaped by the logic of our generalizations. We are deeply committed to the notion that evidence cannot be both the source of hypotheses and the means of their falsification or corroboration. By approaching our analytic task from a modelling perspective we improve the prospect that our propositions follow from a logical, deductive structure and that the empirical assessments are derived independently from the theorizing (BUENO DE MESQUITA; LALMAN, 1992, p. 20).

Their model assume the game-theoretical form depicted in figure 2 (states are represented by the indices 1 and 2). It is constructed based on the elementary assumptions of RCT: rationality, unitary actor and utility maximization. Initially, it takes the form of a non-cooperative, perfect information game, which is tested to evaluate the fitness of realist/neorealist claims of foreign policy. Once data show that these predictions are not supported by statistical relevance, Bueno de Mesquita and Lalman test for the effects of domestic factors, finding strong statistical significance. They then proceed to analyse the effects of norms and beliefs, as well as the prospects of cooperation.

**Figure 2: Bueno de Mesquita and Lalman's game**




Source: SIGNORINO, 1999.

The authors establish a set of seven assumptions that result in the expected utilities for each terminal node in the model. For each proposition derived from the model, the authors conduct logit statistical tests. They use data of dyadic relations in Europe between 1815 and 1970, a total of 707 observations classified according to the characteristics of each dispute. The dependent variables are coded based on this classification and are named BIGWAR, WAR and STATUSQUO. However, their biggest challenge consists in measuring utilities, which they estimate via alliance portfolios. Alliances, in their view, serve "as a revealed choice measure of national preferences on questions related to security", and they "assume that the more similar the patterns of revealed foreign policy choices of two states, the smaller the utility of any demand that one such state makes on the other, and concomitantly, the smaller the difference between $U_i(\Delta_i)$ and $U_i(\Delta_j)$" (BUENO DE MESQUITA; LALMAN, 1992, p. 288). The Kendall $Tau_b$ correlation is the proxy of alliances portfolios in their analysis. Nevertheless, the authors do not have data on the costs represented by $\alpha$, $\tau$ and $\gamma$ ($\varphi$ is operationalized via the use of force).[2]

Bueno de Mesquita and Lalman's work has been subjected to scrutiny by Curtis Signorino, who has been consistently working on mathematical-statistical models since the publication of his paper in the American Political Science Review in 1999. Such models build bridges between the mathematical part of the model and empirical testing, sometimes drawing valuable insights from computational simulations (especially Monte Carlo)[3] and/or statistical models. The essence of Signorino's argument, which is pervasive in his work, is that formal models can only be properly tested if statistical tests are derived directly from the model itself (BAS; SIGNORINO; WALKER, 2008; SIGNORINO, 1999, 2007; SIGNORINO; YILMAZ, 2003). The challenge of empirical testing of formal models lies precisely in the fact that researchers try to forcefully push data into the model without any consideration for the model's assumptions and the theory underlying them (BAS; SIGNORINO; WALKER, 2008). Tests of such nature cannot validate nor falsify a model, for the mathematical bridge is lacking.[4] Furthermore, in many cases, data

2 In Buenos de Mesquita and Lalman's model, $\alpha$ represents the cost borne by the attacked for fighting away from home in a war; $\tau$ represents the cost borne by the target in a war; $\gamma$ represents the cost borne by a state that gives in after being attacked; and $\varphi$ represents domestic political cost associated with the use of force. The authors provide details of these costs in assumption 6 of their model.

Monte Carlo methods consist of computational algorithms based on randomness used to solve mathematical problems where repeated iterations are necessary. Randomness is introduced artificially and is typically used for: sampling, estimation and optimisation (KROESE et al., 2014). Monte Carlo simulations allow for "exploring and understanding the behaviour of random systems and data" by carrying out "random experiments on a computer and [observing] the outcomes of these experiments" (KROESE et al. 2014, p. 387).

3 By mathematical bridge, I mean the set of equations that link the mathematical part of the formal model and the mathematical part of the statistical test.

4 There is a price that should be paid by using higher-order terms, which entail higher-order derivatives. As Burden and Faires point out: "The Taylor methods (...) have the desirable property of high-order local truncation error,

comes in forms that do not fit directly in the model: this is the case, for example, of binary data on international conflict, which are usually coded as presence or absence of war, which is not directly representative of an interaction game (for the game setting generally assumes three possible outcomes: war, capitulation and status quo) (BAS; SIGNORINO; WALKER, 2008; SIGNORINO; YILMAZ, 2003). According to Signorino (1999), the literature on international conflict relies automatically on logit and probit models to test formal models. He disagrees with this approach, for the strategic interaction entails processes and nonlinearities that are not captured by straightforward application of the aforementioned statistical tests. As Signorino suggests:

> [I]f game theory has taught us anything, it is that the likely outcome of such situations can be greatly affected by the sequence of players' moves, the choices and information available to them, and the incentives they face. In short, in strategic interaction, structure matters. Because of this emphasis on causal explanation and strategic interaction, we would expect that the statistical methods used to analyse international relations theories also account for the structure of the strategic interdependence. Such is not the case. (SIGNORINO, 1999, p. 279)

The interactions entailed in the strategic game are pervaded with uncertainties and subgames which are not captured by the formal structure of a logit functional form (SIGNORINO, 1999, 2003; SIGNORINO; YILMAZ, 2003). Applying logit directly results in loss of information about important steps in the interaction game – not to mention the sources of uncertainty faced either by the players or researcher. Moreover, straightforward application of statistical models without adequate adjustments reduces the strategic game to a dyadic setting, either on the side of outcomes (as mentioned previously), or on the side of the number of players involved in the game (SIGNORINO, 1999). This is rather a mathematical problem of incompatibility between linear statistical tests and nonlinear strategic interaction, a misspecification that is common in much of the literature in political science and IR (SIGNORINO; YILMAZ, 2003; SIGNORINO; TARAR, 2006). In sum:

but the disadvantage of requiring the computation and evaluation of the derivatives", which "is a complicated and time-consuming procedure" (BURDEN; FAIRES, 1989, p. 240). Furthermore, it is important to notice that small errors may be exaggerated by numerical differentiation used for estimating the rate of change of measured data (FAUSETT, 2003). Signorino and Yilmaz (2003) strategically overcame this problem in their model by maintaining the parameters $\beta$ linear, redirecting the effects of nonlinearities solely to the regressors X.

> [A]s implemented, the independence assumptions of the statistical models are often inconsistent with strategic interdependence assumptions of the theories. Indeed, these criticisms apply not only to analyses of international conflict but also to logit and probit analyses of any phenomenon involving strategic interaction in international relations, comparative politics, or American politics. Because of this, we should expect, (...) that logit analysis of strategic interaction can lead to parameter estimates with wrong substantive interpretations: Fitted values and predictions of outcome probabilities can be grossly incorrect, as can calculations of the effects of variables on the changes in outcome probabilities. (SIGNORINO, 1999, p. 280)

The standard rationale for model testing is primarily based on the principle of linearity. Mathematically, linearity implies the principles of additivity and homogeneity, expressed below in Eqs. (2) and (3) respectively, where k is a constant.

$$f(a+b) = f(a) + f(b) \tag{2}$$

$$f(k*a) = k*f(a) \tag{3}$$

Together, additivity and homogeneity constitute the superposition principle of linear algebra. Thanks to superposition, the effects of different independent variables can be computed independently in respect to a dependent variable. In structural engineering, for example, for infinitesimal displacements one can apply the superposition principle and calculate separately the effects of torsion, bending and shear caused by a given load, and then compute the total stress at points of interest in the structure by simply adding the values of each separate effect in that point (BEER et al., 2014; BOWER, 2009). Linearity, therefore, decouples effects resulting from the interactions between variables: it assumes that the variables are independent and do not affect each other.

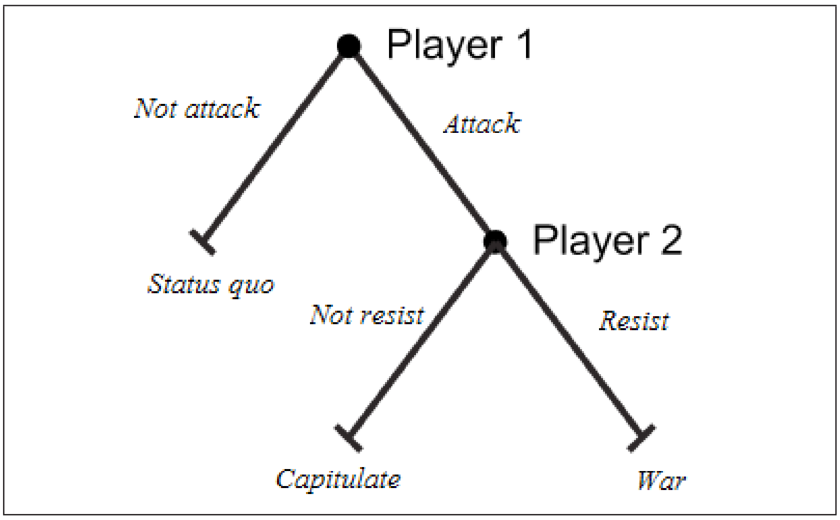
As seducing it is, linearity has become the standard approach in political science and IR. The classical linear regression, for instance, assumes the functional form expressed in Eq. (4), where X is the matrix of regressors, β is the vector of linear parameters, and ϵ the matrix of error variables.

$$y = X\beta + \epsilon \tag{4}$$

$$\text{Where: } y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}, \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}, \text{ and } X = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1j} \\ 1 & X_{21} & \cdots & X_{2j} \\ \vdots & \vdots & \ddots & X \\ 1 & X_{n1} & \cdots & X_{nj} \end{bmatrix} \tag{5}$$

However, the linearity principle entailed in such statistical models fails to capture the effects of dependence between each step of an interaction game and the uncertainties of the decision-making process (SIGNORINO, 2003; SIGNORINO; YILMAZ, 2003). Each branch of the game tree is dependent on the previous node – even the status quo branch – and hence one cannot assume independence between decisions without distorting the game setting. Player 2 makes a decision based on the decision of player 1, entailing thus a sequence of dependent moves, as shown in figure 3.

**Figure 3: Game tree of the sequential interaction game**




Source: Author's design, 2019.

The question concerns how to derive the statistical model whilst preserving the assumptions and structure of the formal model. Signorino and his colleagues have been consistently working on this matter, offering a variety of approaches to solving for the derivation problem. One of the main challenges consists in representing the level of uncertainty entailed in each step of the game tree. A proper model has to be capable of representing the extreme cases (perfect information and complete uncertainty), as well as the cases in-between them.

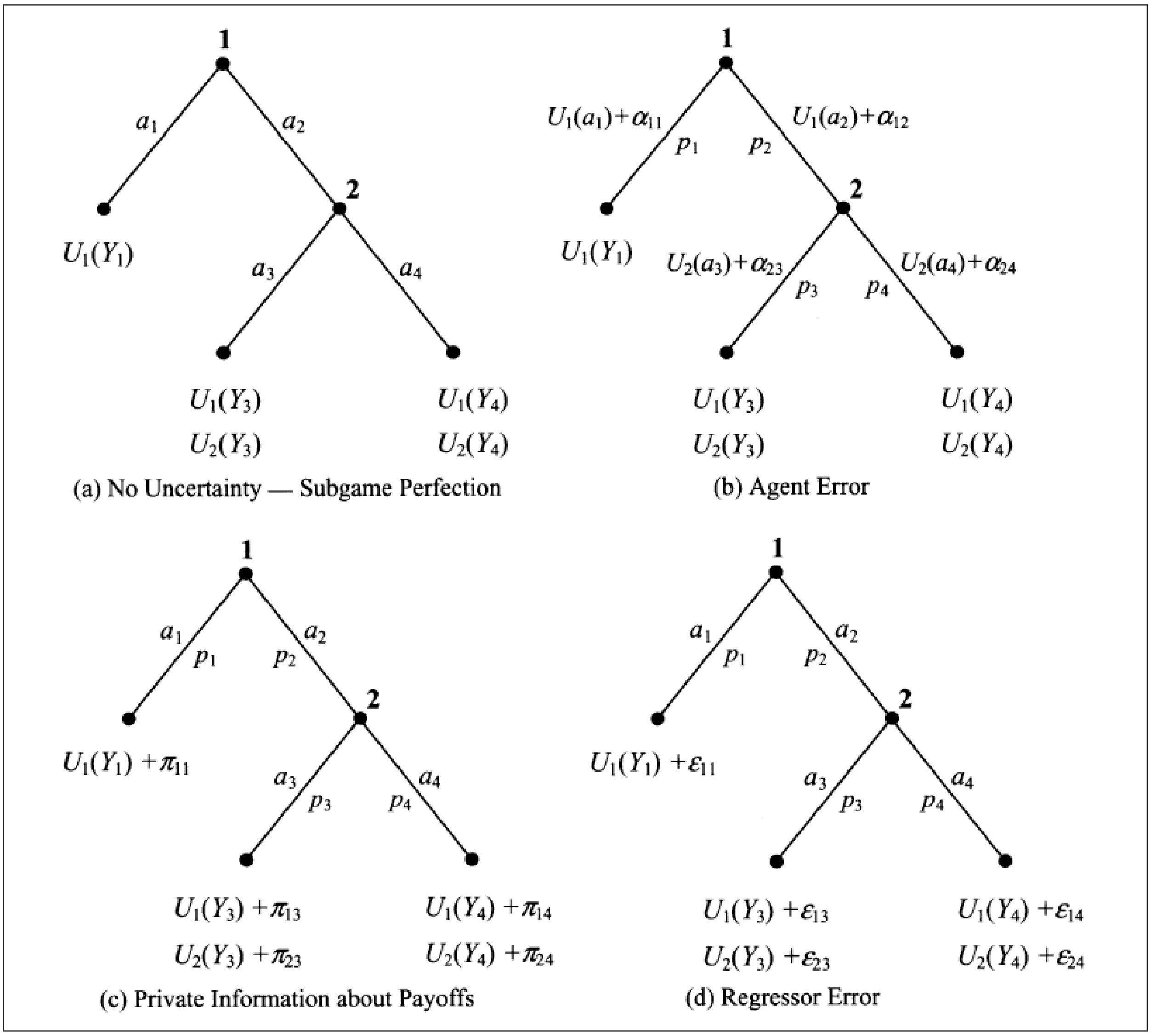
Signorino and his colleagues work extensively with logit and probit models, adjusting them to the formal model of the strategic interaction. Both models deal with binary, categorical data (war, not war; married, not married etc.), and are related to the regression model. In his works, Signorino expresses the utility functions of each player in each branch (either at the final node or on the branch itself) of the game tree via regression, adding error variables that correspond to different theoretical assumptions. The next step consists in injecting these utility functions into the aforementioned models. The logit model [F(x)] implements the regression via the term Y (the regression form expressed in Eq. (4)) in Eq. (6), whereas probit [Pr(Y = 1|X)] does so via Eq. (7), where Φ is the cumulative normal distribution.

$$F(x) = \frac{1}{1+e^{-Y}} \tag{6}$$

$$\Pr(Y = 1|X) = \Phi(X^T\beta) \tag{7}$$

In the strategic interaction model, utility functions are assigned to each player in respect to each possible outcome in the game. There is a component of the utility function that is observable, and this is precisely the component to be regressed (SIGNORINO, 2003). If the model assumes uncertainty, it must be implemented depending on the source of that uncertainty. Signorino (2003) defines three sources of uncertainties: agent error, which assumes that players have bounded rationality and misperceive other players' utilities or that they make erroneous decisions; private information about outcome payoffs, meaning that a player only knows the distribution of others' true utility; and regressor error, which rather reflects the analyst's incapability of modelling players' utilities with the explanatory variables at her disposal. Figure 4 (next page) depicts how the utility functions are implemented in each model.

**Figure 4: Implementation of discrete choice models**




Source: SIGNORINO, 2003. $U_p(Y_k)$ represents each player's observed utilities; α, which is the term for agent error, is implemented on each action branch; π represents the distribution of private information about a player's own outcome payoffs; finally, ε represents the regressor error caused by the analyst's incapability of observing the players' payoffs.

The utility functions in each game specify the source of uncertainty for each case. Based on the example of regressor error (case d), I will explore next how Signorino (2003) derives his model. The utility function is represented by Eq. (8) and the subgame perfect equilibrium is given by Eq. (9).

$$U_m^*(Y_k) = U_m(Y_k) + \varepsilon_{mk} \tag{8}$$

$$y = \begin{cases} Y_1 \; if \; U_2^*(Y_3) > U_2^*(Y_4) \; and \; U_1^*(Y_1) > U_1^*(Y_3) or \\ \quad if \; U_2^*(Y_4) > U_2^*(Y_3) \; and \; U_1^*(Y_1) > U_1^*(Y_4) \\ Y_3 \; if \; U_2^*(Y_3) > U_2^*(Y_4) \; and \; U_1^*(Y_3) > U_1^*(Y_1) \\ Y_4 \; if \; U_2^*(Y_4) > U_2^*(Y_3) \; or \; U_1^*(Y_4) > U_1^*(Y_1) or \end{cases} \tag{9}$$

Remember that in the regressor model, the analyst does not observe the true utilities, and is only capable of making probabilistic statements about the outcomes. Following Signorino (2003), the probability of outcome $Y_1$ is given by Eq. (10), which is the sum of the probabilities comprised by the "or" clause.

$$p_{Y_1} = \Pr[U_2^*(Y_3) > U_2^*(Y_4), U_1^*(Y_1) > U_1^*(Y_3)] + \Pr[U_2^*(Y_4) > U_2^*(Y_3), U_1^*(Y_1) > U_1^*(Y_4)] \quad (10)$$

Eq. (10) can be further clarified by substituting each $U^*_m$ term by its corresponding version of Eq. (9), yielding Eq. (11).

$$p_{Y_1} = \Pr[U_2(Y_3) + \varepsilon_{23} > U_2(Y_4) + \varepsilon_{24}, U_1(Y_1) + \varepsilon_{11} > U_1(Y_3) + \varepsilon_{13}] + \Pr[U_2(Y_4) + \varepsilon_{24} > U_2(Y_3) + \varepsilon_{23}, U_1(Y_1) + \varepsilon_{11} > U_1(Y_4) + \varepsilon_{14}] \quad (11)$$

In order to solve computationally for Eq. (11), it has to be converted into integrals over bivariate normal densities. Signorino does so by denoting the variance of $\varepsilon_{ij}$ as $\sigma^2_{\varepsilon ij}$ and its covariance with $\varepsilon_{ijk}$ as $\sigma_{\varepsilon ijk}$. Also, let $\eta_{ijk} = \varepsilon_{ij} - \varepsilon_{ik}$. Eq. (11), after some reorganising of terms within the brackets, becomes Eq. (12), which can be solved numerically by an appropriate computational routine.

$$p_{Y_1} = \int_{-\infty}^{U_2(Y_3)-U_2(Y_4)} \int_{-\infty}^{U_1(Y_1)-U_1(Y_3)} \Phi(\eta_{243}, \eta_{131})\, d\eta_{131} d\eta_{243} + \int_{-\infty}^{U_2(Y_4)-U_2(Y_3)} \int_{-\infty}^{U_1(Y_1)-U_1(Y_4)} \Phi(\eta_{234}, \eta_{141})\, d\eta_{141} d\eta_{234} \quad (12)$$

This derivation specifies the strategic game depicted in figure [4d]. As one can see, the mathematical work was facilitated by the simple structure of the game, which allowed for a straightforward specification of the subgame perfect equilibrium conditions. However, as Signorino suggests: "[T]he complexity of the underlying game will affect the dimensionality of the integration required for the equilibrium probabilities" (SIGNORINO, 2003, p. 335). Higher dimensionality implies more computational power, which might be time-consuming and too laborious to work out. In this sense, one has to be very careful with modelling in order to make sure that the final model is solvable.

It is worth noting that these models are not constrained to one single source of uncertainty. In other words, it is possible to model agent error, private information and regressor error altogether, yet this would render the model more complicated. The issue here, however, consists in understanding the process of deriving a statistical model from a formal one. The mathematical procedures followed by Signorino (2003) translate the theory into the model and that is precisely where his argument adds to the debate of whether formal models are testable or not. In his words:

> In general, for discrete or continuous dependent variables, one derives a statistical model from a theoretical model using the same general steps: (1) specify the theoretical choice model, (2) add a random component (i.e., source of uncertainty) if none exists, (3) derive the probability model associated with one's dependent variable, and (4) construct a likelihood equation based on that probability model. (SIGNORINO, 2003, p. 318)

Evidently, building a statistical model requires defining further mathematical assumptions, about which the formal model is rather silent. For example, it is commonplace to assume that errors are normally distributed. This is not a flaw of this specific model itself, for the normal distribution is extensively applied in practically all sciences. The question here is that the formal and statistical models are connected via theory, and it is theory that provides the compass to translate a formal model's assumptions into a viable and consistent statistical test (BAS; SIGNORINO; WALKER, 2008; SIGNORINO, 2003).

Another issue of great interest concerns the effects of nonlinearities. To be sure, nonlinear phenomena have been observed in nature since ancient times. Turbulence is perhaps the most widely known phenomenon, and anyone who has ever travelled by plane has already felt the effects of turbulent flow. Eddies and swirls are characteristic features of turbulence and they disrupt a fluid's flow in an irreversible fashion. Turbulence consists of intrinsically disordered movement, caused by small disruptions in the flow. Changes in velocity or pressure might generate turbulent flow, which escalates and propagates as time passes. Likewise, political phenomena are frequently disturbed by noise and nonlinearities, and much of our explanation tries to confer meaning on such complicated features of the political life. That is the case of the strategic interaction game, whose nonlinearities lie precisely on uncertainties.

Overcoming the intrinsic nonlinear aspects of reality is no easy task, and the mainstream in IR would rather pick linear models to deal with their research objects. Signorino and Yilmaz (2003) offer an alternative approach based on a very common procedure in physics and engineering for deriving functions: the Taylor series expansion. A function of one or more variables can be expressed via an infinite sum of terms of increasingly orders, which is useful to approach physical problems, such as the derivation of the equations for pressure field. Taylor series are also useful for analysing error of approximation methods (BURDEN; FAIRES, 1989). The series expansion is represented by Eq. (13) in its original form, which is based on Taylor's theorem.

$$f(x) = P_n(x) + R_n(x) \qquad (13)$$

Where: $P_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k$, and $R_n(x) = \frac{f^{(n+1)}(\xi(x))}{(n+1)!}(x - x_0)^{n+1}$

$P_n(x)$ is the $n^{th}$ Taylor polynomial for f about $x_0$ and $R_n(x)$ is the remainder term associated with the error of expanding f(x) in terms of $P_n(x)$. $R_n(x)$ is also known as the truncation error and it measures the error involved in truncating the infinite expansion at a finite point, i.e. by expanding until a certain term of higher order. As $n \to \infty$, the error decreases.

Signorino and Yilmaz (2003) used Taylor series precisely to account for nonlinear phenomena (via higher-order terms) in their strategic interaction game[5]. They have done so by applying the series expansion to a regression model and rearranging the terms as a linear function of its parameters β. This elegant solution is theoretically more consistent and the results it produces via Monte Carlo simulations seem to be more convincing than the ones generated by the logit model.

Despite its mathematical-statistical consistency, the strategic interaction model has been considered by some as a very complicated model, whose results could be achieved with less sophisticated tools. Carrubba, Yuen and Zorn (2007a, 2007b) challenged Signorino's stochastic approach to strategic decision-making and proposed a return to comparative statics and standard logit and probit testing. They agree with Signorino's argument of placing careful attention on translating a formal model's assumptions into an empirical test. According to them:

> Strategic behaviour will lead to complex parametric relationships, and, as a result, simply including a list of covariates in a linear-form logit is almost certainly a fatal misspecification of the theory. Any well-designed test of a strategic theory must entail accurate operationalization of precisely derived predictions. (CARRUBBA; YUEN; ZORN, 2007a, p. 466)

---

5 Signorino and Yilmaz's model based on Taylor series yields Eq. (14):

$$y^* = -\ln(3) + \tfrac{1}{3}\beta_{13}X_{13} + \tfrac{1}{3}\beta_{14}X_{14} + \tfrac{2}{3}\beta_{24}X_{24} + R + \varepsilon \qquad (14)$$

Where: $R = \frac{-1}{36}X_{13}^2\beta_{13}^2 - \frac{1}{36}X_{14}^2\beta_{14}^2 - \frac{1}{9}X_{24}^2\beta_{24}^2 - \frac{1}{18}X_{13}X_{14}\beta_{13}\beta_{14} - \frac{1}{9}X_{13}X_{24}\beta_{13}\beta_{24} + \frac{2}{9}X_{14}X_{24}\beta_{14}\beta_{24} + \upsilon_3.$

What we can learn from the disagreement between Carruba et al. (2007a, 2007b) and Signorino (2007) is that different statistical models can be used to test a formal model, as long as they are properly derived from the latter. Furthermore, since models are based on mathematics, we may take advantage of this fact and set the boundaries within which they are applicable. Signorino's argument follows this line:

> [A]lthough deterministic models may under certain conditions approximate the relationships in models with uncertainty, in many other situations the predictions will be very different. If one's theoretical model includes uncertainty (e.g., private information or agent error), then the equilibrium conditions should be derived based on the assumed uncertainty. That was actually one of the points of Signorino (2003). If one wants to conduct comparative statics analysis, one should then do so based on the equilibrium conditions for the theoretical model with uncertainty. Similarly, derivation of an estimator, observable implications, or insights for model specification should be based on the equilibrium conditions of the model with uncertainty. (SIGNORINO, 2007, p. 494)

The final question regarding the strategic interaction game and the prospects of empirical testing comes naturally from the Carrubba-Signorino debate: how does a researcher choose between rival models? Luckily, the scholarship has recently been working on feasible tests for comparing the predictions generated by rival models. Clarke (2003, 2007) and Clarke and Signorino (2010) have reached preliminary results with their tests for non-nested discrete choice models, but they do recognise that further research is necessary.

## Assessment: towards structural modelling in IR?

I have explored two models of international bargaining that had become important in recent literature: Fearon's model of audience costs and Signorino's strategic interaction game. Each model entails different rationales when it comes to using statistics as means for testing their assumptions and outcomes. The literature on audience costs has mostly focused on the outcomes of the international crisis game, taking Fearon's assumption of audience costs for granted. Signorino derives his models directly from the game setting to account for mathematical aspects such as nonmonotonicity and nonlinearities, running simulations and performing statistical tests to check the validity of his models.

The standard approach to model-testing in IR follows the lines of the literature on audience costs. Although this approach might render fruitful tests, it does not address the criticisms of selection bias and model-testing properly. It is true that statistical tests based on data alien to the formal model may shed light on certain outcomes of the formal model, but the nature of the test might not suit to validate its results. What I mean by nature is the mathematical construction of the statistical model: it must somehow correspond to the assumptions of the original formal model, for the theorems, lemmas and propositions of the mathematical model are tightly tied to its assumptions. Testing the outcomes might leave crucial questions unanswered, as pointed out in the case of Fearon's model.

In this respect, Signorino's endeavour seems to tackle the issue of model-testing in a more consistent fashion. By deriving a statistical model directly from the assumptions of the formal model, one can be sure about the test's validity once it is confronted with empirical data. Evidently, such derivations require some degree of mathematical manipulation, for many formal models are designed without empirical concerns in mind. In this process, further assumptions may be needed, and having a solid understanding of mathematical concepts will be extremely useful. Furthermore, the pitfalls of linearity should be accounted for, especially when it is known that they are present in the observed phenomenon. As I suggested elsewhere:

> Despite the temptation to trust on linearity, real-world phenomena are pervaded with nonlinear effects. Our brains are wired to think in terms of linear relationships, preferring models where variables behave in a more-or-less linear fashion to those where variables take nonlinear paths. However, nature and society display a variety of phenomena that do not follow the tenets of linearity. Turbulence, fracture propagation, combustion, and conflict escalation are just a few examples of nonlinearities. (LENINE, 2018, p. 93)

For this reason, IR scholars may have to represent nonlinear phenomena in their models, following approaches similitar to Signorino's treatment of uncertainty. Failing to incorporate nonlinearities in a model might affect its explanatory power, more specifically when subjecting it to statistical testing. As we have learned, structure matters, and failing to adequately represent a model's structure may generate incorrect and imprecise outcomes, or even outcomes that are confined to certain boundaries where linearity can be assumed.

It is also important to note that Signorino's bargaining models are more general than Fearon's, for they deal with strategic interaction of any sort. Bargaining takes place not only in the international arena, but practically in all realms of political, social and economic life. Likewise, by recognising the existence of nonlinear behaviour and nonmonotonicity, Signorino shows us the representational gap between the formal model and the statistical test. Such gap might have deleterious effects on the outcomes of the test, leading to implausible conclusions. By drawing our attention to the problem of misspecification error, Signorino demonstrates how fundamental it is to build a mathematically consistent bridge between model and test.

Finally, another approach with which political scientists and IR scholars alike are less familiar is suggested by Rein Taagepera (2008). He recommends shcolars to extrapolate from the classical linear regression equation and look for functionals based on boundary conditions and logical considerations. In mathematical terms, it may lead to the expressing of the observed phenomenon in terms of differential equations, such as Richardson's arms race model. Solving for differential equations results in functional forms that respect the boundary and initial conditions, and the resulting functional is often nonlinear. Nevertheless, basic mathematics does not suffice to deal with such equations, and specific training would be necessary to model phenomena in this fashion. The results, however, would illuminate the understanding of international conflict, for IR scholars would reflect upon the assumptions that should be considered in each situation in a phenomenon-oriented fashion instead of automatically fitting data collected in the real world into a statistical test.

## Conclusion

Throughout the paper, I examined the mathematics behind two approaches to testing formal models, attempting to answer the question of how maths and stats should be combined to provide valid and firm tests. I raised issues on the representational character of tests of formal models, arguing that validity is theoretically and methodologically tied to the structural derivation of statistical tests.

Conventional wisdom in most of IR recommends testing a model for its outcomes. Hypotheses are derived from propositions and theorems, and tested via statistical methods. However, information is usually lost in this process, for statistical methods are standardised, meaning that they do not represent the mathematical structure of the formal model. By resorting to structural derivation, one can solve for the limitations imposed by the standard test whilst respecting the assumptions and phases entailed in the mathematical model. In the particular case of games of international conflict with subgame branches, bridging the maths and stats is quintessential to firmly test the model.

## References

BAS, Muhammet Ali; SIGNORINO, Curtis S.; WALKER, Robert W. Statistical Backwards Induction: A Simple Method for Estimating Recursive Strategic Models. *Political Analysis*, v. 16, n. 1, p. 21-40, 2008.

BEER, Ferdinand Jr.; JOHNSTON, E. Russel; DeWOLF, John; MAZUREK, David. *Mechanics of Mater*ials. New York: McGraw-Hill Education, 2014.

BOWER, Adam. *Applied Mechanics of Solids*. Boca Raton: CRC Press, 2009.

BUENO DE MESQUISTA, Bruce; LALMAN, David. *War and Reason: Domestic and International Imperatives*. New Haven: Yale University Press, 1992.

BURDEN, Richard L.; FAIRES, J. Douglas. *Numerical Analysis*. Boston: PWS-KENT Publishing Company, 1989.

CLARKE, Kevin A. Nonparametric Model Discrimination in International Relations. *Journal of Conflict Resolution*, v. 47, n. 1, p. 72-93, 2003.

CLARKE, Kevin A.A Simple Distribution-Free Test for Nonnested Hypotheses. *Political Analysis*, v. 15, n. 3, 2007, p. 347-363.

CLARKE, Kevin A.; SIGNORINO, Curtis S. Discriminating Methods: Tests for Non-nested Discrete Choice Models. *Political Studies*, v. 58, n. 2, p. 368-388, 2010.

CARRUBBA, Clifford J.; YUEN, Amy; ZORN, Christopher. In Defense of Comparative Statics: Specifying Empirical Tests of Models of Strategic Interaction. *Political Analysis*, v. 15, n. 4, 2007a, p. 465-482.

CARRUBBA, Clifford J.; YUEN, Amy; ZORN, Christopher. Reply to Signorino. *Political Analysis*, v. 15, n. 4, 2007b, p. 502-504.

EYERMAN, Joe; HART, Robert A. Jr. An Empirical Test of the Audience Cost Proposition. *Journal of Conflict Resolution*, v. 40, n. 4, 1996, p. 597-616.

FAUSETT, Laurene V. *Numerical Methods: Algorithms and Applications*. Upper Saddle River: Pearson Education, 2003.

FEARON, James D. Domestic Political Audiences and the Escalation of International Disputes. *American Political Science Re*view, v. 88, n. 3, 1994, p. 577-592.

GARTZKE, Erik; LUPU, Yonatan. Still Looking for Audience Costs. *Security Studies*, v. 21, n. 3, 2012, p. 391-397.

GREEN, Donald; SHAPIRO, Ian. *Pathologies of Rational Choice Theory*. New Haven: Yale University Press, 1994.

KEOHANE, Robert O.; NYE, Joseph S., Jr. *Power and Interdependence: World Politics in Transition*. London: Longman, 2011.

KROESE, Dirk P.; BRERETON, Tim; TAIMRE, Thomas; BOTEV, Zdravko I. Why the Monte Carlo Method is so Important Today. *Computational Statistics*, v. 6, n. 6, 2014, p. 386-392.

LENINE, Enzo. Models, explanation, and the piftalls of empirical testing. *Estudos Internacionais*, v. 6, n. 3, 2018, p. 82-97.

MORGENTHAU, Hans. Política entre as Nações. Brasília: Editora da Universidade de Brasília, 2003.

PARTELL, Peter J.; PALMER, Glenn. Audience Costs and Interstate Crises: An Empirical Assessment of Fearon's Model of Dispute Outcomes. *International Studies Quarterly*, v. 43, n. 2, 1999, p.389-405.

RICHARDSON, Lewis F. *Arms and Insecurity*. Chicago: Quadrangle Books, 1960.

SCHELLING, Thomas C. *The Strategy of Conflict*. Cambridge: Harvard University Press, 1960.

SIGNORINO, Curtis S. Strategic Interaction and the Statistical Analysis of International Conflict. *American Political Science Review*, v. 93, n. 2, 1999, p. 279-297.

SIGNORINO, Curtis S. Structure and Uncertainty in Discrete Choice Models. *Political Analysis*, v. 11, n. 4, 2003, p. 316-344.

SIGNORINO, Curtis S.; YILMAZ, Kuzey. Strategic Misspecification in Regression Models. *American Journal of Political Science*, v. 47, n. 3, 2003, p. 551-566.

SIGNORINO, Curtis S.; TARAR, Ahmer. A Unified Theory and Test of Extended Immediate Deterrence. *American Journal of Political Science*, v. 50, n. 3, 2006, p. 586-605.

SIGNORINO, Curtis S. On Formal Theory and Statistical Methods: A Response to Carrubba, Yuen and Zorn. *Political Analysis*, v. 15, n. 4, 2007, p. 483-501.

TAAGEPERA, Rein. Making Social Sciences More Scientific: The Need for Predictive Models. Oxford: Oxford University Press, 2008.

TOMZ, Michael. Domestic Audience Costs in International Relations: An Experimental Approach. *International Organization*, v. 61, n. 4, 2007, p. 821-840.