János Zoltán VARGA

MODELLING APPROACHES IN ECONOMICS

Equation-based, agent-based and data models

Theses of the Doctoral Dissertation

Szeged, 2023.

University of Szeged
Faculty of Economics and Business Administration
Doctoral School in Economics

MODELLING APPROACHES IN ECONOMICS

Theses of the Doctoral Dissertation

Supervisors:

Dr. habil Gábor Dávid KISS

Egyetemi docens

Szegedi Tudományegyetem

Gazdaságtudományi Kar

Dr. György KOVÁCS

Egyetemi docens

Szegedi Tudományegyetem

Gazdaságtudományi kar

Szeged, 2023

Table of Content

Table of Content 3
1. Introduction 4
   1.1 The relevance of the topic and the goals of the thesis 4
   1.2 Hypotheses 5
2. Research Methodology 7
3. Theoretical background and methodology 9
   3.1 Modeling and philosophy of science 9
   3.2 Data models, equation-based and agent-based models 14
      3.2.1 Data models 14
      3.2.2 Equation-based and agent-based models 16
   3.3 Child as a positional good – related theory 17
   3.4 Social groups, in-group bias 18
4. Main results 23
   4.1 Data model 23
   4.2. Equation-based model 26
   4.3. Agent-based model 27
References 29
Publications related to the thesis 32

1. Introduction

The first section of the thesis booklet describes the relevance of the topic of the thesis and presents the hypotheses.

1.1 The relevance of the topic and the goals of the thesis

Model-building is an integral part of economic and scientific activity in general, and the investigated phenomena can usually only be grasped through some intermediary model. The development of modern economic thinking can also be roughly traced by examining the defining models (e.g. François Quesnay's Tableau économique, Alfred Marshall's supply-demand model, Arrow-Debreu's general equilibrium model, Keynes' (and (Hicks)'s) IS-LM model, the model of capital market exchange rates, Black-Scholes option pricing, dynamic-stochastic general chain equilibrium (DSGE) models, etc. The question arises, which factors make a (economic) model decisive? Are the empirical results, the mathematical elegance, the simplicity, a character that enables further explanation, perhaps the less precisely defined preference of influential economists and economic schools, or perhaps some other circumstance, reason singles out certain representations and models from the entire range? With what methods and models can various economic phenomena be investigated; do they have certain inherent properties modeling approaches that make them more or less applicable?

The topic of this thesis is modeling, the relationship between the model and reality, the advantages and disadvantages of different types of models. Of course, all of this is primarily examined from the point of view of economics. For this purpose, modeling is first placed within the philosophy of science, and then the three modeling procedures most often used by economics are presented, each through a specific example. In this way, the characteristics of model building that I consider important are explained with the help of approaches that differ in substance from each other. The models used by economics can basically be classified into three groups:

Data models: the data model is used to describe data recorded during observations, because the results of experiments and observations are primarily manifested in the form of a set of data, which must then be interpreted by the scientist. In essence, the data model covers the statistical-econometric tools used during the empirical work.

Equation-based model (EBM): the most important methodological tool for theoretical work. In general, it can be said that in an investigated system there are two distinguishable entities: individuals and states. Individuals make up the system, and they are clearly distinguishable both from each other and from the environment (e.g. households, companies, etc.). States or state variables are measurable properties of a system that require interest (e.g. prices, quantity, GDP, etc.). Equation-based modeling focuses on observable states, the basic unit of the model in the equation-based case is the equation itself, which tries to capture the relationship between states and state variables,

Agent-based modeling (ABM): in contrast to EBM, ABM focuses on individuals and the relationships between them. The unit of analysis is the individual, which is represented by an agent. It is used to study complex systems, when the behavior of a large number of heterogeneous individuals and the observable interactions between them shape the macro behavior of the system. In general, it can be difficult to examine with equation-based models, because excessive simplifications have to be used in order to be solvable. In some ways, ABM stands between theoretical and empirical research, because we examine the empirical behavior of the system simulated by the agent-based model.

The thesis provides an example of all three types of models, which enable the examination of different phenomena.

1.2 Hypotheses

In this section, I will briefly present the most important findings of the thesis.

> **Hypothesis 1:** A representation is a *sign* that is created by stipulation according to the creator's intentions, and a scientific representation is a representation that can be considered a scientific model, if the scientific community uses it, accepts it, and recognizes it.

The model usually tries to represent some investigated phenomenon, but it is not clear what properties a model must have in order to help scientific research and make it possible to formulate correct conclusions about the investigated phenomenon. This hypothesis summarizes the position of the dissertation in relation to scientific models (the thesis uses the words model, representation and theory as synonyms, which are also explained there).

> **Hypothesis 2.** In the case of financial time series, the sampling method affects the stochastic properties of the resulting data series.

A significant part of economic publications is of an empirical nature, i.e. a presentation of some kind of observation, data collection, or experiment, with the use of appropriate and accepted statistical-econometric tools. The internal construction and internal structure of the instrument used for observation also includes the prior knowledge and assumptions of the researcher or, more broadly, the scientific community. The method of sampling is therefore crucial: due to the use of an inappropriate procedure, the statistical properties of the resulting sample make drawing conclusions and extracting information from the sample difficult or even prevent it. All this increases the uncertainty of conclusions made during empirical work. The hypothesis is tested on financial time series, examining the effects of alternative sampling procedures on the statistical properties of the obtained data, as well as their role in risk management.

> **Hypothesis 3:** A child is partly a positional good, and for this reason, depending on the social environment, income can have a negative effect on fertility.

Positional goods are among those goods whose evaluation by the consumer depends in some way on the environment. According to the basic idea, certain goods are socially scarce, their consumption depends negatively on the number of consumers, and they also cause external effects (Frank 2005). Social scarcity refers to the fact that the scarcity of certain goods does not occur due to physical resource limitations, but due to social factors. For example, the number and number of players on the field in a soccer team is determined by the rules and the coach's decisions, while the number of leadership positions within a company or other organization is a consequence of the organizational structure (Vatiero 2008). Having a child has an aspect that depends on the environment, i.e. the child can be considered partly as a positional good, i.e. the usefulness that the parent realizes partly depends on how he performs compared to his peers, and what social status the child later achieves.

To test the hypothesis, a microeconomic, equation-based model is introduced (constrained profit maximization), which enables the formal capture of positional considerations and the theoretical support of the argument.

> **Hypothesis 4.** The ability of individuals to join more than one social group is key to sustaining large-scale cooperation.

Groups play a fundamental role in building the structure of human societies: humans are social being; he has a basic need to belong to a group, he derives part of his personality and self-esteem from group membership (Tajfel and Turner 2004; Baumeister and Leary 1995). As the community became larger and more complex, people perceive and distinguish more and more social categories and groups in order to cope with the computational and cognitive challenges of the social world (Macrea - Bodenhausen 2000). As a result, the individual can be a member of several social groups at the same time. Consequently, social groups compete for the members' resources (primarily for their time) in order to maintain the cultural customs, norms, and symbolic features that distinguish them from other social groups. The hypothesis is tested using the agent-based model presented in chapter 5 of the thesis, and I point out that competition between social groups for the resources of individuals can have a detrimental effect on large-scale cooperation at the entire population level, and one of the remedies for this problem can be recategorization or creation of new groups covering the entire population.

2. Research Methodology

The examination of the 1st hypothesis is carried out on the one hand with literature review: in the 2nd chapter of the thesis, the related concepts and theories of the philosophy of science, the defining theories regarding modeling and representation, and the most important findings related to economic modeling are presented. This is followed by an explanation of one's position. On the other hand, the methodology and results used to examine the next three hypotheses of the thesis can be considered as supporting the first hypothesis.

The examination of the 2nd hypothesis is tested on financial market data: a statistical examination of data series produced by traditional (time-based) and alternative methods of sampling (autocorrelation, heteroscedasticity, normality) is performed.

Hypothesis 3 is tested using a constrained optimization problem built to model households' decisions about having children.

To test the 4th hypothesis, I created an agent-based model, in which the agents are persons and social groups, the model uses both game theory and network theory tools.

Table 1. Research hypotheses and methods

No. of hypothesis	Hypothesis	Method
1	A representation is a *sign* that is created by stipulation according to the creator's intentions, and a scientific representation is a representation that can be considered a scientific model, if the scientific community uses it, accepts it, and recognizes it.	Literature review
2	In the case of financial time series, the sampling method affects the stochastic properties of the resulting data series.	Data model - comparison of the statistical properties of data series produced by traditional and alternative sampling procedures (autocorrelation, heteroskedasticity, normality).
3	A child is partly a positional good, and for this reason, depending on the social environment, income can have a negative effect on fertility.	Nonlinear programming, equation-based model
4	The ability of individuals to join more than one social group is key to sustaining large-scale cooperation.	Agent-based model, network theory, game theory

Source: own calculations.

3. Theoretical background and methodology

The theoretical background of the thesis can be divided into 4 parts: the first is the philosophy of science approaches related to modeling, the emerging questions regarding the relationship between the model and reality, and the most influential findings of economics related to modeling. In addition, it discusses the most important theoretical and philosophical properties of the three applied modeling tools (data model, equation-based and agent-based). Furthermore, the thesis presents the knowledge related to positional goods, which are a starting point for the equation-based examination of the positional aspects of childbearing. Finally, the connections between an individual's social group membership and cooperation are explored.

3.1 Modeling and philosophy of science

Below, I briefly review the most important and popular approaches to representation. One group of trends focuses on the relationship between the model and the target system. Using the tools of set theory and mathematical logic, the structural concept intends to describe the model and the target system as some formally definable structure. And the representation is created by the fact that an isomorphism can be established between the two structures. By structure we typically mean a mathematical object consisting of a non-empty set U, which represents the object's constituent elements, and n number of relations R={r_1,..,r_n} interpreted on U, which describe the internal structure of the object. This is usually denoted as $S = (U, R)$, so the model and target object structure are S_M and S_T, respectively. In the case of isomorphism, a mutually clear assignment can be established between U_M and U_T, and it also preserves the relations, i.e. to every r_{iM} relation interpreted by U_M, a r_{jT} relation interpreted by U_T can be assigned in such a way that exactly those elements are in relation to each other that correspond to each other. Other, less strict morphisms are also discussed in the literature, thus trying to deal with, for example, the previously mentioned problem of symmetric representation. The isomorphism relation interpreted on the set of structures is symmetric (also transitive and reflexive - thus an equivalence relation), which also means that the object or target system to be represented also represents the given model (Frigg 2006, Winther 2016, Suarez 2010).

The next trend is primarily associated with the name of Ronald Giere, and instead of the strict morphism of the structural approach, it emphasizes the similarity of the model and reality (or more precisely, the target system under investigation). In the case of the structural approach, the model and the target system are in a precisely defined relationship, but in the case of similarity, the definition of this relationship is not a goal or even not achievable. For this reason, the structural relationship can also be considered a special case of the similarity approach (Suarez 2010). An important difference, however, is that Giere emphasizes the role and intentionality of the researcher, i.e., the representation between the model and the target system is not created simply by similarity, but by the researcher using this similarity or certain elements of it during modeling for certain for a predetermined purpose (Giere 2004, 2006).

Against the structural and similarity approach, Suarez (2003) provides five independent objections and argues for the inadequacy of these approaches. The first refers to the fact that in scientific practice, a considerable part of the models is not isomorphic and not even similar to the investigated phenomenon. According to the second, the logical relationship between the model and the target system resulting from the above theories is inappropriate: as discussed above, isomorphism is a reflexive, symmetric and transitive relation, and similarity is reflexive and symmetric: in reality, however, neither property is fulfilled in practice. Another problem is that it is difficult to interpret incorrect representation or inaccuracy in the case of isomorphic and similar representations. To understand the former, think of the evening when we identify a person in a photograph with someone other than the person actually depicted in the photograph. In other words, argues Suarez, the agent's activity during the experience of representation is not a consequence of the representational relation. All of this also means that mere similarity or isomorphism in some structure does not constitute a representation, if this were the case, then the mathematical description of the general theory of relativity created by Einstein to describe space-time should actually be linked to the name of Bernhard Riemann, since he was the one who provided the necessary developed mathematical equations (Suarez 2003. p. 234). Another challenge is the issue of inaccuracy:

usually the models we use do not accurately reproduce the properties of the target system: however, assuming isomorphism, this is simply not possible (the inaccuracy objection affects the similarity approach less). A representation can be created even if isomorphism and similarity are not fulfilled, which means that they are not necessary criteria.

According to Callender - Cohen (2006), the problem of demarcation, which they call the (problem of representation), that is, how it is created, what the representation between the model and reality actually consists of, ceases to be a problem when approached from a different perspective. Based on the language philosopher Paul Grice, primary representations are distinguished from derived ones. The former essentially mean mental representations, while the latter include other representations in art and language, which in some way are based on the first, more fundamental mental representations, and derive from them. The key is how the derived representations are created. According to Callender and Cohen, scientific representations also belong to derived representations and are created in a stipulative manner based on the agent's intentions, i.e. a scientific model *M* represents the target system *T* by the user of the model stipulating that *M* represents *T* (Frigg - Nguyen 2016, Callender - Cohen 2006). On the one hand, the difficult question of representation is pushed back into the world of fundamental representations, that is, into the field of mental representations and states, which is the research area of this philosophy of mind. At the same time, the problem of demarcation was thus solved: the representation is created according to the researcher's intentions. The selection between the various resulting models is (also) based on pragmatic aspects, where both the different morphisms and the similarity between *M* and *T* can play a major role, but these are not necessary conditions for the creation of the representation.

The description of representations developed by R. I. G. Hughes and briefly referred to as DDI (Denotation, Demonstration, Interpretation) divides the modeling process into three phases. First of all, the investigated phenomenon must be marked and symbolized in some way (denotation). Hughes rejects similarity/structural conditions in relation to notation and representation, i.e., the created symbol or model does not necessarily have to resemble the investigated phenomenon. In the second phase, we examine the internal dynamics of the model and draw conclusions (demonstration). Typically, this involves performing operations on mathematical objects and analyzing other non-linguistic representation tools. It is important to see that our conclusions are drawn here in relation to the model (ie the equation, diagram, figure, etc.), not directly to the phenomenon under investigation. The latter occurs in the third stage of the DDI, during the interpretation of the results, when the results of the model are interpreted with regard to the investigated phenomenon (Hughes 1997).

Suarez uses Hughes' description of DDI as the starting point for the so-called inferential conception. According to Suarez, in the theory of representation, the need to define the necessary and sufficient conditions must be abandoned and an agreement must be reached on the definition of the more modest, necessary conditions. In his theory, he defines two factors as necessary conditions for scientific representation: representational force and inferential capacities. The power of representation essentially refers to the fact that the things used as representations are created through stipulation, that is, the intentions of an agent are also needed to establish the direction of representation (that is, that *M* represents *T*). With this condition, it solves the problem of symmetrical representation, however, an arbitrary representation is not interesting from a scientific point of view: it must have some additional cognitive content. The second condition, i.e. the explicit formulation of the condition of substitute reasoning, applies to this: the model (*M*) should enable a properly informed individual to draw conclusions about the investigated phenomenon (*T*) (Suarez, 2004).

Gabriele Contessa further develops Suarez's theory and argues that substitute reasoning and representational power, together with the concept of interpretation, provide necessary and sufficient conditions for representation (Contessa 2007). As a criticism of Suarez's description, it does not reveal how substitution reasoning actually takes place. At Contessa, the interpretation is meant to remedy this: in essence, similar to the structural concept, it defines a mathematical relationship between the representation tool (*M*) and the target system (*T*). Briefly, this happens as follows: first, the user identifies the relevant objects, the relations between them, and the relevant functions interpreted on them (the latter two essentially give the system its structure) for both the representation tool and the target system. After that, between them (that is, between the relevant objects, relations and functions of the representation and the target system) a bijective mapping is established by prescribing and stipulating (i.e. the act that the user does *M* to the representation of *T*) the given representation device (*M*) to the target system (*T*) relevant interpretation.

According to Contessa's intentions, such a formal definition of interpretation allows him to create a similar set of formal rules for substitute reasoning. To understand how this happens, let's assume that the researcher or user uses *M* as described above in his interpretation of *T*. Then the user's conclusion that an object ($O_{T,i}$) is included in the target system (*T*) is correct if and only if $O_{T,i}$ corresponds to an object, $O_{M,i}$ in the model, *M* -in. Formally, he makes completely identical restrictions regarding the conclusions regarding relations and functions (Contessa 2007 p. 61-62). According to Frigg-Nguyen (2016), Contessa's theory provides an explanation for the epistemic representation problem but cannot deal with misrepresentation (see above). To this we can add that the representation without a target system is not even conceivable in Contessa's model.

Table 2. Models and philosophy of science

Most important concepts	Main findings	Key publication (Author / year)
Structural relationship, isomorphism	Using the tools of set theory and mathematical logic, the structural concept intends to describe the model and the target system as some formally definable structure. And the representation is created by the fact that an isomorphism can be established between the two structures.	Da Costa - French 2003
Similarity	Instead of the strict morphism of the structural approach, it emphasizes the similarity of the model and reality (or more precisely, the target system under investigation).	Griere 2004

Stipulation, primary and derived representation	Scientific representations also belong to derived representations and are created in a stipulative manner based on the agent's intentions, i.e. a scientific model M represents the target system T by the user of the model stipulating that M represents T	Callende r - Cohen 2006
DDI (Denotaton, Demonstration, Interpretation)	He divides the modeling process into three phases: the investigated phenomenon must be marked and symbolized in some way (denotation); in the second phase, we examine the internal dynamics of the model and draw conclusions (demonstration); in the third stage, the results are interpreted - the results of the model are interpreted with regard to the investigated phenomenon (interpretation)	Hughes 1997
representational power, inferential capacity	It defines two factors as necessary conditions for scientific representation: representational power (things used as representations are created through stipulation) and inferential capacity (the model should enable a properly informed individual to draw conclusions about the phenomenon under investigation)	Suarez 2004
Surrogate reasoning, representational power, interpretation	Surrogate reasoning and representational power, together with the concept of interpretation, provide necessary and sufficient conditions for representation. In essence, the interpretation defines a mathematical relationship between the representation tool and the target system, similar to the structural concept.	Contessa 2007

Source: own construction

3.2 Data models, equation-based and agent-based models

3.2.1 Data models

A significant part of economic publications is of an empirical nature, i.e. a presentation of some kind of observation, data collection, or experiment, with the use of appropriate and accepted statistical-econometric tools.

The data model is used to describe the data recorded during the observations, because the results of the experiments and observations are primarily manifested in the form of some data set, which the scientist must then interpret. Both steps of the process, i.e. data collection and interpretation, raise questions. In the first step, the data is essentially collected through some kind of sampling procedure. James McAllister, for example, believes that an infinite number of patterns can be identified in a set of data obtained during any observation, from which we cannot select the one that would be ontologically connected to the real phenomenon based on the knowledge of the data alone. The researcher is the one who chooses from the patterns, that is, as a result of his preliminary knowledge and theorizing process, he creates the correspondence between the selected pattern and the phenomenon to be explained (McAllister 1997). Consider money market time series as an example. Price, interest, and exchange rate data are typically recorded as a function of time, i.e., at regular intervals (for example, daily) a sample is taken from the market transactions of the given instrument, and then based on the sample, a price is assigned to the time in question. The price can be determined in several ways, for example, it can be the weighted average price of the transactions of the given period, or the closing price, etc. The basis of this latter assignment can be, for example, the weighted average of the period, the opening price, closing price, etc. However, the statistical properties of the time series collected in this way are not always ideal: the distribution of returns is significantly different from normal, heteroscedasticity and autocorrelation can be observed (Kiss 2017). However, financial market data can be recorded not only as a function of time, but also, among other things, in the number of transactions (so, for example, we consider 100 deals as one unit and assign price information to it), or also in the total value of deals. These latter data series have statistical properties that are much easier to handle (López de Prado 2018).

In the course of the research, I used an alternative and traditional (time-based) sampling procedure to create data series from the trading data of the Microsoft share and compared their statistical properties. To test the normality of the distribution, I performed the Jarque-Bera test. The null hypothesis refers to a normal distribution, the alternative hypothesis refers to its absence. I used the ARCH LM test to check homoscedasticity, the null hypothesis in this case is homoscedasticity, and H1 is heteroscedasticity. Finally, I also examined autocorrelation using the Ljung Box Q statistic, here the null hypothesis means the absence of autocorrelation.

Table 3 Financial sampling methods

Sampling method	Sample is taken:	Example
Time based	Regular time interval	Hourly, daily, quarterly
Transation based	Fix number of transactions	100 transaction
Volume based	Number of exchanged instruments	1000 stock
Value based	Cumulative value of exchanged instruments	100000 dollar

Source:own construction.

3.2.2 Equation-based and agent-based models

Equation-based model (EBM): the most important methodological tool of economic theoretical work. In general, it can be said that there are two distinguishable entities in an investigated system: individuals and states. Individuals make up the system, and they are clearly distinguishable both from each other and from the environment (e.g. households, companies, etc.). States or state variables are measurable properties of a system that require interest (e.g. prices, quantity, GDP, etc.). Equation-based modeling focuses on observable states, the basic unit of the model in the equation-based case is the equation itself, which tries to capture the relationship between states and state variables.

Agent-based modeling (ABM): in contrast to EBM, ABM focuses on individuals and the relationships between them. The unit of analysis is the individual, which is represented by an agent. It is used to study complex systems, when the behavior of a large number of heterogeneous individuals and the observable interactions between them shape the macro behavior of the system. In general, it can be difficult to examine with equation-based models, because excessive simplifications have to be used in order to be solvable. In some ways, ABM stands between theoretical and empirical research, because we study the empirical behavior of the system simulated by the agent-based model.

3.3 Child as a positional good – related theory

In the economic sense, positional goods are among those goods whose evaluation by the consumer depends in some way on the environment. The concept originates from Fred Hirsch (Hirsch 1976), its antecedents can be found earlier with Veblen and Galbraith (Vatiero 2008).

According to the basic idea, certain goods are socially scarce, their consumption depends negatively on the number of consumers, and they also cause external effects (Frank 2005). Social scarcity refers to the fact that the scarcity of certain goods does not occur due to physical resource limitations, but due to social factors. For example, the number and number of players on the field in a soccer team is determined by the rules and the coach's decisions, while the number of leadership positions within a company or other organization is a consequence of the organizational structure (Vatiero 2008).

A classic example of a purely positional good is social status: a higher social status can only be interpreted if there are individuals who have a lower status. In addition to purely positional goods, it is true for almost all goods that their evaluation is partially influenced by the environment, the difference in the case of positional goods is that the environment shapes the usefulness perceived by the consumer to a relatively greater extent. The key here is that it is not the absolute value of consumption that determines it, but its relative measure compared to other relevant members of the community (Frank 2006).

In the case of social status, there will be higher and lower positions compared to the rank and status occupied by the individual. Because of the former, it suffers negative utility (due to negative consumption), and because of the latter, positive. The balance of the two gives the individual the (total) utility from the given position. If the status changes, for example it becomes bigger, then it will realize negative utility due to fewer positions, and positive utility compared to more positions. That is, as status increases, total utility increases (marginal benefit is positive), and it will not necessarily be decreasing.[1]

[1] Intra-elite competition is the rivalry between elite aspirants for a limited number of elite positions within society (such as top political leadership positions, board seats in large corporations, etc.). Following Goldstone (1991), Turchin (2016) believes that the

Based on the above thought process, it seems reasonable to assume that having children can be partly motivated by positional considerations, i.e. a part of the usefulness hoped for or experienced by the parent, initially, comes from how the child performs compared to his peers (academic, sports, artistic or other area), and later what social status it achieves. As is also true for other factors affecting fertility, in the case of positional effects, it does not hurt to emphasize that its effect is not independent of the given social context. In archaic societies, the division of labor and specialization are less typical, and the egalitarian nature of those societies gives little space for status struggles between children and adults (Boehm 2001), while in a modern environment, where parents typically have 1-2 children, and inequalities it is better tolerated (Morris 2017), where both the space (supply) and the need (demand) for positional goods are greater. In theory, this could lead to an "arms race" between parents, i.e. the increasingly intensive education and training of children in order to get into high schools and universities with a good reputation and/or to achieve outstanding results in the fields of sports and art (Deresiewicz 2015). In the United States, for example, homework is becoming more common in kindergartens, thus preparing preschool children for the school environment and expectations. More affluent parents finance more and more activities outside of school (sports, art, volunteer activities, etc.) for their children, knowing that they will be able to earn extra points during their university admissions.
All of this also means that it is not quality in itself that is important to the parent, but how it is formed into social status. Due to the above, it may be worthwhile to supplement traditional fertility economic models with positional considerations, the formal explanation of which will be detailed in the next section.

3.4 Social groups, in-group bias

Groups play an essential role in the structure of human societies: humans are inherently social beings; people are predisposed to join groups, derive part of their identity and self-esteem from group – membership (Tajfel and Turner 2004; Baumeister and Leary 1995; Fiske

intensification of this competition can greatly contribute to political instability in any society. This suggests that the marginal utility of social rank may be increasing.

1992). Social exclusion and lack of group membership can lead to psychological and physical symptoms (DeWall - Richman 2011; Eisenberger et al. 2003). As the community became larger and more complex, people perceive and distinguish more and more social categories and groups in order to cope with the computational and cognitive challenges of the social world (Macrea - Bodenhausen 2000). In this way, the individual can be a member of several social groups at the same time, for example family, ethnic, religious community, nation, workplace community, imagined and virtual communities (Brewer 2000; Anderson 2016).

Social groups have an important defining characteristic: norms. Norms are informal rules that are formed through interactions between group members and guide and limit the behavior of members, and also distinguish social groups from each other (Cialdini - Trost 1998; Hogg 2010). People have an internal predisposition to internalize the norms of the groups to which they formally or informally belong (Spitzer et al. 2007; Chudek - Heinrich 2011), they sacrifice resources (mostly time and cognitive capacity, but often concrete material goods as well) in order to maintain the cultural practices, traditions and customs of the group. A social group that does not have members does not exist - the norms and customs that once influenced the behavior of members of historical groups remain recognizable, but since no one is a member of these groups, they play no role in current social structures.

Furthermore, the so-called in-group bias (in-group favoritism) develops in the individuals belonging to the social group, i.e. they will prefer the group members over the individuals belonging to the external group, they will be more likely to cooperate with them (Greenwald - Pettigrew 2014).

Consequently, social groups compete for the members' resources (primarily for their time) in order to maintain the cultural customs, norms, and symbolic features that distinguish them from other social groups. In the model, I examine the effect of this competition for group members and their resources, in-group bias and multi-group membership on cooperation at the population level. To implement this, I developed an agent-based model that uses both network theory and game theory.

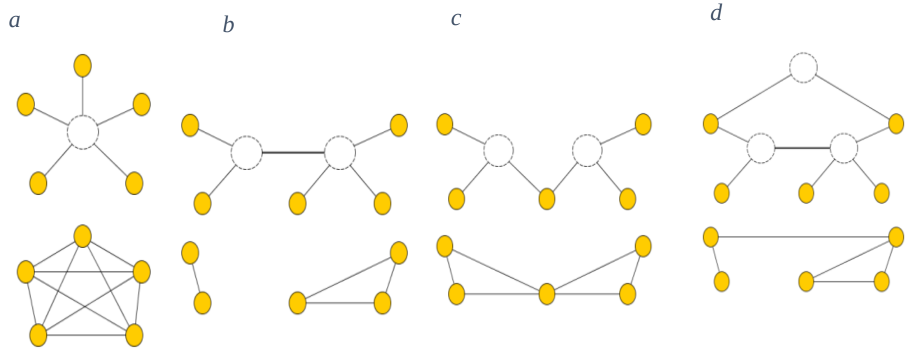


Figure 1. Social structure and cooperation. The graphs in the first row show group membership. The colored and white nodes represent individuals and groups, respectively. An edge between a group and an individual represents group membership. Edge between groups means possible strategic interaction, dual membership is not allowed between connected groups. The lower graphs are the individual projection graphs of the upper graphs, an edge indicates cooperation between two individuals. (a) One group, every individuals is member of the same group, the projection graph is complete, i.e. everyone cooperates with everyone else. (b) Two groups, dual membership is not allowed, the projection graph is disconnected. (c) Two groups, dual-membership is allowed, the projection graph is connected. (d) Three groups, dual-membership is partly allowed, the projection graph is connected.

In the model, we assume that individuals only cooperate with their own group partners. If there is only one social group in the given community, then cooperation is complete in the sense that in-group bias is applied to everyone, no one is excluded due to group membership (see figure 1. a).

However, if there are two or even more groups, a given individual may be a member of only one. In this case, he will not cooperate with members of the other group, and the community will therefore be fragmented (see Figure 1b). The problem can be solved if it becomes possible for individuals to join both groups, thus cooperation with members of both groups can be realized (Figure 1 c). This social structure can be grasped with a bipartite graph, where there are two types of vertices: one represents individuals (the colored vertices in the figure), and the other represents groups (white vertices), the edge between them expresses the individual's group membership. The projection of the graph constructed in this way onto the individuals shows the cooperation and cooperation network. That is, the projection of the graph is derived from the original graph, in such a way that the vertices in the projection show the individuals, and the edge between them means that there is a group of which both are members (this is shown in the second row in Figure 1 visible graphs), thus there is no obstacle to cooperation between them. Examining the projection, the fragmented society can be visualized easily, namely the graph that is not connected (Figure 1 b); there are sub-communities that have absolutely no connections to the rest of the graph and thus to society.

What we want to investigate is how the projective graph evolves over time under such model assumptions. In our model, there should therefore be n individuals and m groups. Individuals have some amount of initial resources, r. A part of the resource is randomly allocated to one of the groups (r/k, k>0), expressing group membership, which requires time and other expenditure on the part of the individual). The same process is repeated until all of the individual's resources are distributed among groups. During the cycle, it may happen that the individual is already a member of a group (he has already "given" a part of his resources to the group) and joins another one. We see this as a strategic situation from the point of view of the two groups: cooperation between the two groups means that the group allows its members to join another group, refusal to cooperate (defection) expresses the prohibition of this. More precisely, in the case of groups, non-cooperation means that within the group there are norms, written and unwritten rules, which implicitly or explicitly limit the individual, the member of the group, from joining the other group. As an example, consider the institution of dual citizenship: the acquisition of the first citizenship is usually simple, almost automatic, whereas the second (and any subsequent ones) are usually subject to conditions: it must be applied for, and the applicant must meet certain criteria, in fact, some countries do not recognize it at all the institution of dual citizenship. The question arises, why do such norms, rules and restrictions exist? When an individual becomes a member of two or more groups, he must share his available resources between the groups. If the group restricts its members' multi-group membership, there is a chance that the individual will allocate the rest of his resources to the given group. In concreto, in the model, when an individual joins two groups that have mutual cooperation, the individual remains a member of both groups. If one group refuses to cooperate, the individual withdraws his resources from the other group (r/k) and randomly reallocates them to another group, so in the case of m groups, the resource in question is returned to the person who refuses to cooperate with a probability of 1/m. If both groups refuse to cooperate, then the individual randomly withdraws the dedicated resource from one group and is again randomly assigned to another group. The exact payoff matrix of the strategic situation is shown in Table 4.

Table 4. Payoff matrix between two groups (row player perspective).

	C	D
C	0	-r/k
D	$1/m \times r/k$	$-½ \times r/k + ½ \times 1/m \times r/k$

The characters *r*, *m* and *k* denote the total resource of an individual, the number of groups in the model, and a constant (*k*>0), respectively. If both groups cooperate, the individual remains member of both groups. Whenever a defector meets a cooperator the individual leaves the latter group and its resource is randomly allocated to a new group. Upon mutual defection, the individual leaves one of the groups randomly, and its resource is reallocated to a new group, randomly.

To ensure some kind of dynamics in the model, a part of each individual's resources is reallocated every 10 periods on average. The evolutionarily stable strategy is DD, i.e. both groups in a strategic position prohibit their group members from dual group membership. In the initial bipartite graph, we can illustrate this by allowing edges between vertices representing groups, which represent the possible strategic interaction between the two groups, i.e. then the groups can develop norms and rules that apply to dual group membership. The absence of an edge between two groups shows that the given groups do not "see" each other, do not get into a strategic situation with each other. In other words, this means that when an individual joins two such groups (between which there is no edge in the graph), he keeps his membership in both groups. A possible reason for this may be that the groups in question are new or not populated enough, and there has not been time to develop and spread such norms. The complete graph topology interpreted between vertices representing groups (the situation where there is an edge between all vertices) means that each group in a given society can establish and maintain norms vis-à-vis each other group. In this case, the evolutionarily stable strategy spreads easily, that is, each group within the society will limit its members regarding multigroup membership. Consequently, the social structure will be fragmented, the projection of the graph, i.e. the cooperation network, will not be connected, each individual will be a member of exactly one group, members of different groups will not cooperate with each other (Figure 1b). The entire graph structure between the groups can be interpreted as a very rigid caste system.

For comparison, in addition to the full graph topology, the effect of the star and scale-independent topology was also examined. As discussed earlier, scale-independent networks are found in many real-world, complex networks, essentially meaning that most vertices have relatively few connections, but there are some with extremely many connections. In the case of star topology, the graph has a central vertex and all other vertices are connected exclusively to it. This would represent a situation where each group could be in a strategic situation with the same specific group (this is the central peak), but apart from this, the multigroup membership of individuals is not restricted. The star topology can be thought of as a social structure where every small and large community defines itself against the same group.

4. Main results

4.1 Data model

Table 5 Basic statistics of stock returns sampled in different ways from intraday data.

	1 min returns	Transaction based returns	Volume based returns	Dollar returns
Mean	-0.0000227	-0.000029	-3.72E-05	-0.0000369
Median	0	-0.0000396	0	0
Max	0.001809	0.002045	2.14E-03	0.002221

Min	-0.001806		-0.002348		-0.002272		-0.001882	
SD	0.000652		0.000676		0.000689		0.000731	
Ferdeség	-0.081094		-0.051097		-0.02386		0.077483	
Csúcsosság	2.853577		3.176807		3.043865		2.826557	
Jarque-Bera stat., p val	0.774	0.679	0.671	0.715	0.0636	0.969	0.726	0.696
ARCH-LM stat., p val	8.216	0.0164**	5.707	0.058*	2.844594	0.241	17.15	<0.01***
Ljung-Box Q stat., p val	1.722	0.423	3.2174	0.2	3.4594	0.177	2.2247	0.329
No. of obs.	389		386		363		322	

Note: The null hypotheses are for the JB test, ARCH-LM test and Ljung-Box are normal distribution, homoscedasticity, no serial correlation, respectively.
Source: own calculations.

It can be seen that the distribution can be considered normal in all cases, the JB test statistics are smaller in the case of the alternative methods, that is, they are closer to normal. Homoscedasticity appears in the data series basically using the transaction and volume-based method, the p-value of the latter is much more convincing (ARCH-LM test), while autocorrelation is not characteristic of the sample taken with either method (Ljung-Box test). A summary of these data can be found in Table 2. Based on the results, it can be said that in the case of intraday data, the statistical properties of the sample constructed on the basis of volume are the most favorable, but the other three do not perform badly either.

Table 6. Basic statistics of stock returns sampled in different ways from 1 year of trading data.

	Daily returns		Volume based returns		Dollar returns	
Mean	0.001999		0.001753		0.001851	
Median	0.001565		0.001935		0.002023	
Maximum	0.143094		0.049882		0.095846	
Minimum	-0.148353		-0.062841		-0.083622	
SD	0.025463		0.018165		0.021075	
Ferdeség	-0.040455		-0.139978		0.126396	
Csúcsosság	12.92568		3.255202		5.878468	
Jarque-Bera stat., p val	1018.1	<0.01	1.476885	0.47786	85.93011	<0.01
ARCH-LM stat., p val	47.80707	<0.01	0.581686	0.7476	18.46171	<0.01
Ljung-Box Q stat., p val	55.402	<0.01	0.9561	0.62	3.8342	0.147
No. of obs.	248		247		247	

Note: The null hypotheses are for the JB test, ARCH-LM test and Ljung-Box are normal distribution, homoscedasticity, no serial correlation, respectively.
Source: own calculations.

Based on the JB test, the data calculated from the daily data is not normally distributed, and it is also characterized by autocorrelation and heteroskedasticity. The statistics of the volume-based exchange rate return, on the other hand, show a very favorable picture from the point of view of statistical modeling: normal distribution, free of autocorrelation and homoscedastic. In other words, with volume-based sampling, it was again possible to generate a data set with better properties and approaching the ideal. In the case of total value-based data, the results are mixed: there is a smaller peak compared to the daily returns, the value of the JB test statistic is an order of magnitude smaller, but even so it cannot be considered a normal distribution. This sampling procedure did not eliminate heteroscedasticity, but it did eliminate autocorrelation. It is also worth noting that the alternative data series have a smaller standard deviation compared to the daily returns.

In the light of the above results, I cannot reject my 2nd hypothesis, that is, the method of sampling (data series generation) has an influence on the stochastic properties of the resulting data series. This also means that there is no "clean data", and this should definitely be taken into account and kept in mind when interpreting the results of empirical research.

4.2. Equation-based model

The problem for the fertility model is the following:

$$\max_{c,l,n,e_c,t_c} \; a_c \log(c) + \; a_l \log(l) + \gamma\, a_n \log(n) + \gamma\, a_p \log(A) + \; \gamma\, a_p\, \alpha \log(e_c) + \; \gamma\, a_p\, \beta \log(t_c)$$

$$\text{S.t.} \qquad w \cdot tc \cdot n \; + \; k \cdot n \; + \; l \; \cdot w \; + \; ec \cdot n \; + \; c \; - \; w - y = 0$$

where c is consumption, n is the number of children, l is free time, t_w is time spent working, t_c is time for children, e_c is money spent on children, and w is wages, parameters $a_n, a_c, a_l \geq 0$, Where k is the fixed expenditure associated with having a child ($k \geq 0$), γ is a parameter representing the total importance attached to having children ($\gamma \geq 0$), p is the status achieved by the offspring

Solving this for n and taking the partial derivative of (8) with respect to w, we arrive at the following expression:

$$n^* = \frac{-\gamma(a_p\, \alpha + \; a_p\, \beta - a_n)w}{(\gamma a_c \; + \; a_l \; + \; a_p)k}.$$

Since $a_c, a_l, a_p, a_n, k, \gamma \; \geq \; 0$, it can be seen that the relative size of the parameters a_p and a_n determines the direction of the relationship between w and n. We can distinguish three cases:

$a_p > \frac{a_n}{\alpha+\beta}$ the increase in income has a decreasing effect on fertility,

$a_p < \frac{a_n}{\alpha+\beta}$ the number of children increases with income,

$a_p = \frac{a_n}{\alpha+\beta}$ there is no direct relationship between income and fertility.

The denominator of the expression on the right-hand side is the sum of the exponents of the production function that determines the quality of the offspring: α+β=1 means a constant return to scale for the quality function of the child, so proportionally increasing the time and material expenditure results in a corresponding improvement in quality.

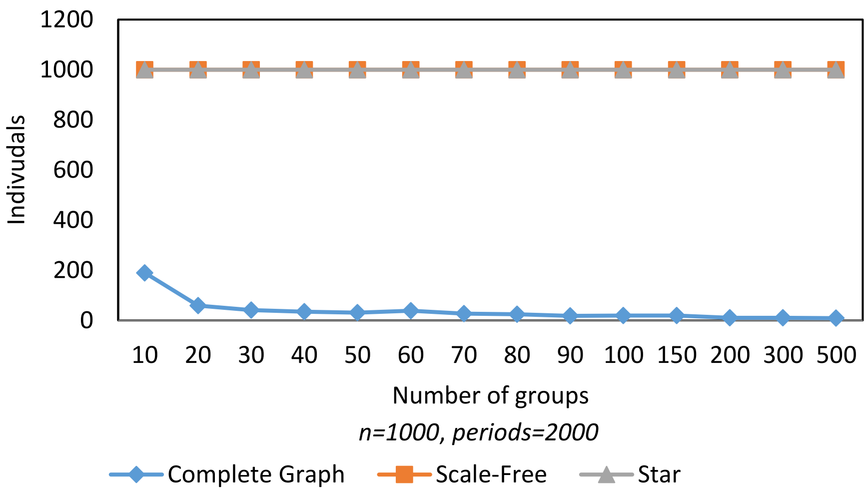
Based on the model, we can therefore say that if households consider the positional aspects of having a child to be sufficiently important (that is, the child's status, relative position, advancement among contemporaries, and soon in society as a whole), and they consider that by increasing the parental effort, the child their socio-economic status can also be improved, then, in addition to higher income, they intend to have fewer children.

On the other hand, if the institutional and cultural environment does not allow social mobility, that is, quality does not necessarily mean a substantially higher status (e.g. in feudal conditions) and/or the parental investment does not result in a significant increase in quality, then higher income means higher brings fertility.

Based on the results, I cannot reject my 3rd hypothesis: a child is partly a positional good, and for this reason, depending on the social environment, income can have a negative effect on fertility.

4.3. Agent-based model

A computer simulation was used to gain insight into the empirical behavior of the model, the results are shown in Figure 2. From the point of view of our investigation, the most important attribute is the size of the largest component within the individuals' projection graph. The largest component is nothing but the largest connected subgraph within the entire network, i.e. the most populous group between which there is cooperation, directly or indirectly (or there is a path between any two vertices of the subgraph). I also examined the size of the largest component with different group numbers, m. In the case of the complete graph topology, the size of the largest connected component became smaller and smaller as m increased, that is, the entire society is made up of small, closed communities with approximately the same population, whose members do not cooperate with each other. However, in the case of the other two configurations, the largest component covers the entire society. It is important to emphasize that the evolutionarily stable strategy also spreads among the groups in these cases, i.e. where there is strategic interaction between the groups, they also act prohibitively against the sharing of group membership, however, due to the fact that not all groups enter into such interaction, the groups they do not completely limit individuals' group choice options. Therefore, situations can easily arise in which two individuals are members of two different groups that prohibit joining the other group, but at the same time both individuals are members of a third group against which there are no such restrictions on the part of either group, which allows cooperation between the two. between.



2. ábra. *A legnagyobb összefüggő komponens mérete az egyéni projekciós gráfban. A grafikon a szimulációs eredményeket mutatja különböző csoportközi topológiával (teljes gráf, skálamentes és csillag topológia) különböző m (csoportszám) értékek mellett. Az n és periódus paraméterek az egyének számát, illetve a szimuláció hosszát adják meg. Forrás: saját szerkesztés.*

Using the presented ABM, I found that the ability of individuals to join more than one social group is key to sustaining cooperation among large populations. Thus, I cannot reject my 4th hypothesis**.**

References

Anderson, B. 2016. Imagined Communities: Reflections on the Origin and Spread of Nationalism. London: Verso.

Baumeister, R. F., M. R. Leary. 1995. The need to belong: Desire for interpersonal attachment as a fundamental human motivation. *Psychol. Bull.* 117:497–529.

Boehm, C. (2001). Hierarchy in the Forest: The Evolution of Egalitarian Behavior. Harvard University Press, New York.

Brewer, M. B. "Reducing prejudice through cross-categorization: Effects of multiple social identities" in *Reducing prejudice and discrimination*, S. Oskamp. Eds. (Lawrence Erlbaum Assoc. Publ., 2000) chap. pp. 165-183.

Callender, C., Cohen, J. (2006). There Is No Special Problem About Scientific Representation. *THEORIA*, 21(1), 67–85.

Colander, D. – Kupers, R. (2014). Complexity And The Art Of Public Policy - Solving Society's Problems From The Bottom Up. Princeton University Press, Oxford.

Contessa, G. (2007). Scientific Representation, Interpretation, and Surrogative reasoning. Philosphy of Science 74 p. 48-68.

Chudek, M., J. Henrich. 2011. Culture–gene coevolution, norm-psychology and the emergence of human prosociality. *Trends in Cognitive Sciences.* (15) p. 218–226.

Da Costa, N. C. A., French, S. (2003). Science and Partial Truth: A Unitary Approach to Models and Scientific Reasoning. Oxford University Press, New York.

DeWall, C. N., B. S. Richman. 2011. Social Exclusion and the desire to reconnect. *Social and Personal Psychology Compass* 5:919-932.

Eisenberger, N. I., M. D. Lieberman, K. D. Williams. (2003). Does rejection hurt? An fMRI study of social exclusion. *Science* 302:290-292.

Fiske, A. P. (1992). The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review* 99**:** p. 689 – 723.

Frank, R. H. (2005). Positional Externalities Cause Large and Preventable Welfare Losses, American Economic Review, 95(2), p. 137-141.

Frigg, R. (2006). Scientific Representation and The Semantic View of Theories. *Theoria* 55 p. 49-65.

Frigg, R., Nguyen, J. (2016). „Scientific Representation" in The Stanford Encyclopedia of Philosophy (Winter 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/scientific-representation/>.

Giere, R. N. (2004). *How Models Are Used to Represent Reality*. Philosphy of Science 71 (5) 742-752

Giere, R. N. (2006). *Scientific Perspectivism*. University Of Chicago Press, Chicago.

Greenwald, A. G., T. F. Pettigrew. 2014. With malice toward none and charity for some – ingroup favoritism enables discrimination. *American Psychology* 7:669-684.

Hughes, R. I.G. (1997). Models and Representation. *Philosophy of Science* 64 p 325-336.

Darley, J. M. (2008). Social comparison motives in ongoing groups” in *Blackwell Handbook of Social Psychology: Group Processes*. M. A. Hogg, R. S. Tindale. Eds. (Blackwell) chap. 14. pp. 281–297.

Hirsch, F. (1978). *Social Limits to Growth*. Routledge.

Hogg, M. A.. “Influence and leadership” in *Handbook of Social Psychology*, S. Fiske, D. T. Gilbert, G. Lindzey. Eds. (Wiley, 2010), vol. 2. chap. 31. pp. 1166-1207.

Kiss Gábor Dávid (2017). Volatilitás, extrém elmozdulások és tőkepiaci fertőzések. Jatepress, Szeged.

Macrea, C. N., G. V. Bodenhausen. (2000). Social cognition: thinking categorically about others. *Annual Review of Psychology*. 51:93-120.

Lopez de Prado, M. (2018). *Advances in Financial Machine Learning*. Wiley, New York

Mcallister, J. (1997). Phenomena and patterns in data sets. *Erkenntnis* 47 217-228.

Macrea, C. N., G. V. Bodenhausen. (2000). Social cognition: thinking categorically about others. *Annual Review of Psychology*. 51:93-120.

Morris, I. (2017). *Foragers, Farmers, and Fossil Fuels: How Human Values Evolve (The University* Center for Human Values Series). Princeton University Press, Princeton.

Pagano, U., (1999), Is Power an Economic Good? Notes on Social Scarcity and the Economics of Positional Goods. In Samuel Bowles, Maurizio Franzini and Ugo Pagano, eds., The Politics and the Economics of Power. London: Routledge. p. 63-85.

Richerson, P – Boyd, R. (2005). *Not By Genes Alone How Culture Transformed Human Evolution*. Chicago University Press, Chicago.

Spitzer, M., U. Fischbacher, B. Herrnberger, G. Grön, E. Fehr. 2007. The neural signature of social norm compliance. *Neuron* 56:185–196.

Suarez, M. (2003). Scientific Representation: against similarity and isomorphism. *International Studies in The Philosphy of Science* 17 (3).

Suarez, M (2010). Scientific Representation. *Philosophy Compass* 5 (1) p. 91-101.

Tajfel, H., J. C. Turner. ”The Social Identity Theory of Intergroup Behavior” in *Political psychology: Key readings*, Sidanius, J. Eds. (Psychol. Press, 2004) chap. 16. pp. 276-293.

Turchin, P. (2016). Ages of Discord: A Structural-Demographic Analysis of American History. Beresta Books

Vatiero, M. (2011). The Institutional Microeconomics of Positional Goods.

Winther, R. G. (2016). *The Structure of Scientific Theories*. The Stanford Encyclopedia of Philosophy, Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/structure-scientific-theories/>.

Publications related to the thesis

Janos Zoltan Varga (2021): Interest rate pass-through using Weighted Average Cost of Liabilities Method. Evidence from Czech Republic, Hungary and Romania. *Acta Oeconomica* 71(4) p. 551-567

Janos Zoltan Varga (2021): Effect of various sampling methods on the statistical properties of financial data and their application in risk management. *Statisztikai Szemle* 99 (3) p. 233-252.

Janos Zoltan Varga (2018): Competition between Social Groups, in-group favoritism and population-level cooperation. *Cliodynamics: The Journal of Quantitative History and Cultural Evolution*, 9 (1).

Gábor Dávid Kiss, Máté Csiki, János Zoltán Varga (2019). Comparing the IMF and the ESM through Bond Market Premia in the Eurozone. *Public Finance Quarterly* 64 (2).

György Kovács, Gábor Dávid Kiss, János Zoltán Varga (2016): Expectations and Monetary Policy - with Special Attention to the Hungarian Experiences (Várakozások és a monetáris politika – különös tekintettel a magyarországi gyakorlatra). *Economic Review (Közgazdasági Szemle)* 63, November pp. 1192-1216. DOI:10.18414/KSZ.2016.11.1192

Janos Zoltan Varga (2016): The Effect of Interbank Liquidity Surplus on Corporate and Interbank Interest Rates. *Public Finance Quarterly* 61(1) pp. 95-115.

Gábor Dávid Kiss, János Zoltán Varga (2016): Extreme fluctuations of financial time series. *Statisztikai Szemle (Statistical Review)* DOI:10.20311/stat2016.02.hu0162

János Zoltán Varga Interest rate pass-through using Weighted Average Cost of Liabilities Method. Evidence from Czech Republic, Hungary and Romania. In: Udvari, B.; Voszka, É. (szerk.) Proceedings of the 3rd Central European PhD Workshop on Economic Policy and Crisis Management. Szeged, Magyarország: Szegedi Tudományegyetem Gazdaságtudományi Kar, pp. 60-77.

Varga János Zoltán (2017). A Niche Construction Theory (NCT) közgazdaságtani alkalmazásának lehetőségei. In: Farkas, Beáta; Pelle, Anita (szerk.) Várakozások és gazdasági interakciók. Szeged, Magyarország : JATEPress, pp. 57-68.

Varga, János Zoltán (2016). Comparison of the Current Monetary System and the Full Reserve System by Agent-based Models. Artificial Economics Conference, Róma, Olaszország, 16 p.