TARTU UNIVERSITY

FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

Institute of Computer Science

Chair of Theoretical Computer Science

Master Studies in Computer Science

JAAK SIMM

Method for Creating Empirical Agent-based Models for Economics

M. Sc. Thesis

Supervisor: Merik Meriste

Author: .... "....." May 2004

Supervisor: .... "....." May 2004

Chair: .... "....." 2004

TARTU 2004

1 Table of Contents

1 Introduction

A model represents our understanding about a phenomenon. The model links different aspects together by showing how they influence each other. Economics is a science that creates and analyzes models about complex and dynamic processes, such as economic cycles, inflation, that emerge as a result of human activities. Economic theories are usually based on the understanding about behavior of groups of individual economic actors, such as investors, consumers. These theories try to give insight how the whole economy and its subparts behave.

The critical issue here is how to test the validity of these theories. In other words, how can the researchers evaluate whether and how well the proposed theory actually represents the underlying phenomenon. For that cause economics has adopted range of statistical tools in order to seek for empirical correspondence. Although based on sound mathematical foundations, the standard statistical methods employed in economics mainly operate on aggregated level and do not directly model the actual economic phenomenon or the underlying logic. Therefore, such models do not directly embed agent level economic relationships, such as decision process of single consumer. However, during the recent years economics has seen the advent of multiagent technologies that, contrary to statistical models, start from the agent level and model the processes from bottom up. Although economics has quickly embraced the ideas and a new field called Agent-based Computational Economics (ACE) [1] has emerged, there are no or very few attempts to use these techniques to create methodology for using agent-based models to test aggregate economic phenomenon with real data.[1]

This thesis is an attempt to develop empirical method based on multiagent systems to test economic theories, more accurately macroeconomic theories. The thesis creates a basic framework how to build agent-based models for economics, given the micro structure and quantitative data of the real processes. The goal is to arrive at an agent-based modeling method that allows to test hypothesis about the real data. The employment of agent-based approach might lead to a different view about the modeled phenomenon and, thus, provide several advantages. The incorporation of micro-level economic logic can give better understanding about the examined phenomenon and could result in better models, in terms of reliability and forecasting ability. Additionally, it might be claimed that for some complex models the standard statistical techniques are insufficient and

1 Here the thesis means empirical models that are capable of estimating a series, not just resemble or explain features of a series (e.g. variance, stationarity, particular movements). This is the case for agent stock market models.

complementing them with agent-based models provides deeper insights to the phenomenon.

1.1 Research Problem

The thesis defines its research problem as follows:

How to create agent-based models that are able to simulate real macroeconomic processes defined by given time-series data, with a goal to test the fitness of an economic theory to the real data. Thus, the aim is to develop and evaluate an empirical agent-based modeling method for testing economic theories.

In short, the main objectives can be listed as:

- Define criteria for good modeling methods
- Develop and implement the empirical agent-based method
- Evaluate the method

1.2 Outline

The paper is split into following parts:

1. Conceptual basis of the multiagent systems and economic modeling. Comparison of the statistical and agent-based approach.
2. The main ideas of the modeling technique and covering of specific modeling aspects.
3. Realization details. Object diagram and main implementation issues.
4. Example simulation and hypothesis testing. Modeling an example economic study.
5. Implications of the results.

2 Conceptual Basis

The thesis will now lay down the conceptual framework. The section begins by introducing the principles of economics. Thereafter, the ideas of empirical models are discussed. This leads to the overview of possible approaches to modeling of economic processes, namely statistical and agent-based. The paper gives an overview of the commonly used statistical method and then introduces the concepts and terms of multiagent systems.

2.1 Economics and Macroeconomics

Economics is about how people live their lives, use their time and resources to produce things other people are willing pay for (adopted from [2]). It is a study how limited resources are employed to satisfy the limitless needs of people.

The two main disciplines of economics are microeconomics and macroeconomics. Microeconomics examines in high detail how individual entities make choices. These entities include consumers, employees, firms, investors, politicians. On the other hand, the macroeconomic theories look at the economy by summing together all the actions of these microeconomic entities. Thus, macroeconomics views processes at level of whole economy or at the level of large sectors of the economy. For example, when microeconomics is interested in finding out how much a person is willing to pay for a milk, macroeconomics analyzes how big is the current demand for food products in the whole economy. The main themes of macroeconomics include theories explaining the whole output of the economy (gross domestic product), government expenditures, inflation, investment and consumption, as well as international trade and money flows.

Macroeconomic theories are dealing with aggregated processes that emerge from the micro-level actions of individual economic agents. The relationships between macroeconomic variables are ad hoc in the sense that they only proxy the outcomes of individual choices [2]. This fact makes it natural to apply agent-based modeling approaches economics and as well to macroeconomics, since all economic processes are founded on actions of individual agents.

2.2 Empirical Models of Economics

Empirical research is the key that links economic theories with the real world. It is important to keep the theory and reality from drifting apart and living separate lives. A theoretical model gives an abstract interpretation of an economic phenomenon. Whether the interpretation is useful and actually explains the underlying phenomenon needs to be tested with empirical models.

The two main tasks of the empirical research in economics are outlined as follows: [2]

1. Quantification of the parameters that are present in the theoretical models. These parameters can be interpreted by the underlying model and reflect some economic property. For example, such parameters can be the effect of the change in interest rate to investment and the portion of income people use for consumption.
2. Testing of hypothesis that are derived from theoretical models. The quantifiable hypothesis are tested to see whether the theory is supported by real data. An example of a hypothesis is a claim that people spend more of their income, when inflation increases. Such claims need to be by empirical tools.

Current paper investigates only temporal (dynamic with respect to time) empirical models. That means that static models (independent from the flow of time) are excluded. Most of the interesting questions in economics are related to temporal models, for example modeling growth gross domestic product (GDP) or trade flows over time. The data sources for temporal models are based on time-series. Time-series is an ordered collection of values. These values represent the value of an economic variable at successive periods of time. In macroeconomics the time-series are usually on monthly, quarterly or yearly basis.

There are several aspects that determine the usefulness of an economic modeling technique. When developing a modeling technique high attention has to be paid to those factors. Therefore, the paper identifies the main factors:[1]

- **Reliability**. The results of the modeling process can be trusted. The reliability also includes reproducibility of the results, given the same model and time-series.
- **Speed of modeling.** The creation of a model does not take much time for the researcher.
- **Sufficiently fast simulation.** The computational part is fast enough to get the results from the model in reasonable time.
- **Possibilities of modeling.** The more freedom the for the researcher to model the underlying processes.

These factors will be later used evaluate the agent-based modeling method.

1 These factors were identified with an help of Lenno Uusküla, an economic expert in the Bank of Estonia.

The approach to the empirical modeling developed by the current thesis differs from the standard ideas in the empirical research of economics. Standard techniques employ statistical methods for developing time-series models, whereas this thesis tries to use agent-based models as the basis. The overview and differences of these two approaches will be covered in the subsequent sections.

2.3 Statistical View

The standard techniques used for empirical research in the field of economics are called econometrics. Econometrics uses statistical means to create empirical models for the underlying economic theory. The basic principle of econometrics is statistical regression. The regression explains one time-series by a linear combination of other time series using ordinary least square (OLS) methodology [2]. The time-series being explained is called endogenous and modeled by the following formula:

$$Y = \alpha + \beta_1 * X_1 + \beta_2 * X_2 + \ldots + \beta_n * X_n$$

Y here is the endogenous time-series and $X_1,\ldots,X_n$ are time-series used for explaining, $\beta_1,\ldots,\beta_n$ are respective coefficients for the linear equation and α represents the constant term. Thus, Y, the time-series of an economic process, is modeled as a linear combination of other time-series[1].

The coefficients (α, $\beta_1,\ldots,\beta_n$) are determined by least squared error method using continuous mathematics (parameters are found by equating the derivative of the sum of the squared errors with 0 and solving the equations). Based on sound statistical theory, the regression allows to evaluate the significance of each time-series. If the time-series are significant it gives support for the given economic model. Additionally, the expectations formed from the economic theory about the magnitude and sign of each coefficient allow to check whether the empirical model confirms with the theory. For example, if an economic theory predicts that interest rate should negatively affect investment, then negative coefficient for interest rate in the regression, which is trying to explain investment, gives support for that theory.

Due to the non-stationarity of most economic time-series, the usual ordinary statistic techniques, based on stationarity assumption, are augmented with co-integration analysis [3]. This involves making the time-series stationary by taking first (or second) difference of the initial non-stationary series. Usually co-integrated models are based on auto-regressive statistical models like ARIMA [3]. The specifics of these models are out of the scope of this paper, for more

1 The linear combination as noted here, can be both additive as well as multiplicative.

details see [3].

The econometrics looks at all economic processes as time-series that are then modeled with the statistical tools. The use of simple mathematical form for the model allows to employ well-founded means to quickly calculate the empirical model for the given time-series.

2.4 Agent-based View

The agent-based view takes a different approach to modeling. Instead of creating a simple mathematical model, the underlying model is based on a system comprised of various interacting agents. Therefore, its structure and behavior have potential to resemble the actual economic theory and reality better than simple mathematical models. Especially, when the underlying real economic relationships are complex. The thesis will now lay down the important ideas and concepts about agent-based view. These will be the foundations on which the paper later develops the agent-based modeling method (see section Modeling Method).

2.4.1 Need for Agent-Based Modeling

The central idea of agent-based modeling techniques is to represent some economic process as interactions between several autonomous agents. When applied to economics the agent-based approach is natural, because economics deals with processes that are the result of actions of individual agents like consumers, producers, investors.

Thus, the main reasoning for using agent-based modeling method for empirical modeling is coherence between the structures of the real and agent-based model. This also results in the enlarged possibilities of modeling. The method can open up several opportunities and features that are not present in the conventional statistical modeling:

- Testing of complex relationships between variables. It is possible to insert any complex behavior to the agent decision model and test its fit against real data.
- Checking hypotheses about structural aspects of the economic model, such as market mechanisms.
- Straightforward testing of micro-level economic models with macro data.

From the algorithm theory perspective the agent-based systems are different processes than usual prescriptive algorithms. The agent-based systems are interactive processes which make the process complex and unpredictable. Therefore, the behavior of such processes is called emergent.

Following sections will introduce the conceptual basis of agent-based systems. Firstly, the central concepts of agents and multiagent systems are discussed outlining definitions for the most important terms. Secondly, the theories behind the modeling of agent-based systems are covered, focusing on the modeling of economics.

2.4.2 Agents

The concept of agent-based modeling is founded on the idea of an agent. Shortly put, an agent is a independent process that interacts within its surrounding environment. In general sense, the meaning of agent is easy to grasp, but surprisingly there is no agreement among researchers about the exact definition. This can be contributed to the fact that the concept of agents is the basis of several research fields that head into different directions. Nevertheless, there is consensus that autonomy is the central aspect of an agent. Without covering different possible views the paper adopts the definition by [4] as it fits well to the agent-based modeling framework:

> **Definition.** An **agent** is a system that is situated in some environment, and that is capable of autonomous action in this environment in order to meet its design objectives.

Figure 1 depicts the concept of an agent. An agent receives information from and is affected by its environment. In turn, the agent affects the environment with its actions. The environment of the agent consist of objects and

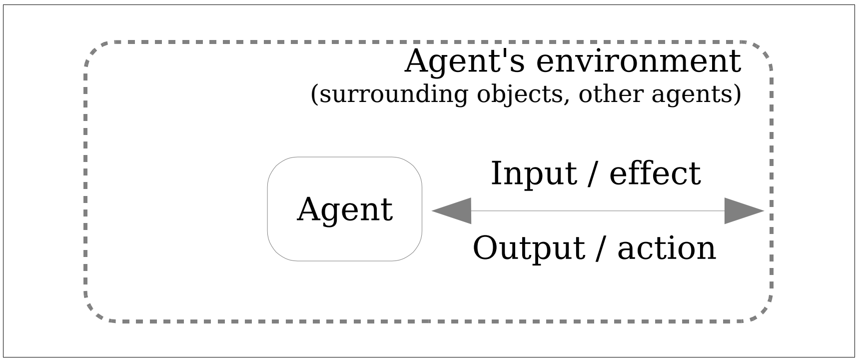


Figure 1. An agent within its environment

entities that the agent is affected and itself can affect. The environment of an agent also includes other agents that directly or indirectly are connected to the agent. It should be noted that the definition of an agent does not require an agent to be intelligent, it can be purely reactive, only acting on the impulses of the environment. This view on agents also applies to real economic agents (like firms and consumers) and software agents.

2.4.3 Multiagent System

Agents on their own, in isolation, do not possess great value. Their significance arises when several agents are grouped together and the emerging behavior of the group is analyzed. Such systems are called multiagent systems or agent-based systems, for which the paper adopts the following definition:

Definition. A **multiagent system** is a system that is comprised of several agents that can interact with each other.

The literature often separates multiagent systems and agent-based systems. The former is regarded as a system with a specific purpose or goal. Current thesis makes no distinction between the two terms. Thus, the definition of a multiagent system does not imply any specific design goal of the system.

Multiagent system is a new paradigm that has had big influence on model building and simulation in the recent years. Its main idea is the interaction of agents that results in emergent behavior of the whole system [5]. Interaction between agents can produce a stable system that shows new global behavior in higher level of abstraction. For example, an analogy in economics is how the agent-level decisions of individual consumers and companies create the macro-level behavior of economic cycles and trends.

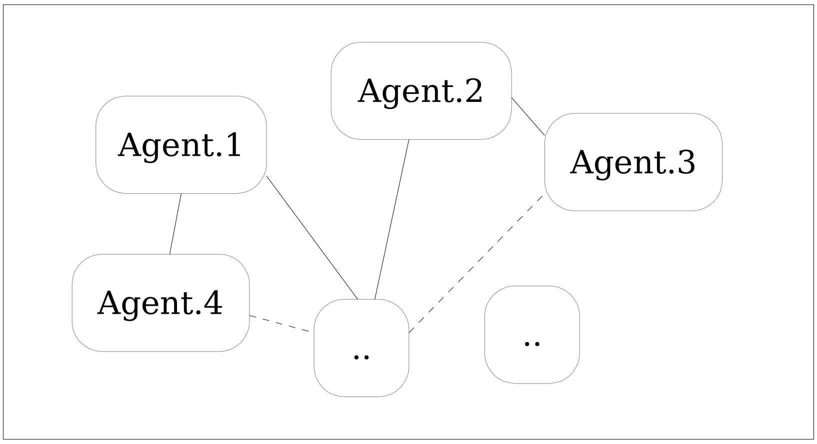


Figure 2 Illustration of a multiagent system.

Figure 2 illustrates a multiagent system. For interaction agents can have two different type of connections:

1. **Permanent connections** are predetermined by the design of the system. These connections are depicted as continuous lines in the figure and show prefixed communication relationships.
2. **Ad-hoc connections** are not predetermined. They are created during the runtime of the system. These are depicted as dashed lines between agents that have started an interaction dynamically during the runtime of the system.

A system with mainly ad-hoc connections has more flexible structure, and can be a test-bed for success of interaction techniques and protocols.

There are three major application areas of the paradigm of multiagent systems:

1. **Empirical sciences**. Empirical sciences analyze real systems and create models for them. The real agents and their interactions are simulated in an artificial computer environment with the goal to replicate the actual behaviors present in the real system. Thus, the agents in the model and their behavior are determined by the actual behavior of the agents in the real system. System theory has shown that numerous real systems despite superficial differences have the same deep structure [5].
2. **Engineering sciences and technology.** Engineering sciences and technology have the goal of designing and developing useful systems that have not existed before. Contrary to empirical sciences the engineering sciences do not seek resemblance with existing systems. The main emphasis is on the function and the performance of the new system. Founding the design of the system on the architectural patterns of multiagent systems has proved to be useful. Two such application areas are robots and complex software systems.
3. **Theory and virtual worlds.** Concept of multiagent systems opens up possibilities to develop new worlds of consisting interacting agents. This is useful for both developing theoretical views and for creating virtual worlds for entertainment.

Current thesis belongs under the application category of empirical sciences, since the goal is to develop techniques for empirical modeling of economics. Due to the good fit between economics and the concept of agent-based systems, the area has grown quickly and separated into a field of its own. The next section overviews the research in agent-based modeling of economics.

2.4.4 Agent-based Modeling of Economics

One of the fields of study under the multiagent paradigm is modeling of economics, also named as agent-based computational economics (ACE). ACE represents economic processes by a decentralized system of autonomous agents that through parallel interaction drive toward their given goals. The central idea of ACE, as noted by [1], is growing economies from the bottom up. Even though the field is new, the research has already separated into many different areas. The ACE can be roughly divided into 8 directions: [1]

1. Learning and embodied mind. Under this area the researches investigate learning and genetic algorithms to represent the learning process of computational agents. The area includes comparisons of different auction trading strategies [6], evaluation of computational financial market trading strategies [7] and studies about the effect of the level of learning in oligopoly games [8].
2. Evolution of behavioral norms. The area investigates how behavioral norms grow and decline as an evolutionary process. The main findings are that simple rules for norms can create dramatic patterns affecting the evolutionary outcome of economic and social processes.
3. Bottom-up modeling of market processes. This is one of the most active areas in ACE research and looks at the modeling of markets on the basis of agents. The main focus is put on simulating financial and commodity markets to resemble the behaviors that have been observed in real markets. This allows to compare different market structures and policies in a simulated environment.
4. Formation of economic networks. The area analyzes how economic agents choose their transaction partners (clients, suppliers, allies etc) and how such transaction networks evolve over time. Earlier researches focused on the efficiency and formation of different trade networks, for example [9]. Recent studies have started investigating specific markets that rely on networks, such as labor and wholesale markets.
5. Modeling of organizations. The area views organizations as complex adaptive multiagent systems and analyzes the effects that organizational structure has on the behavior of the organization[10].
6. Design of computational agents for automated markets. Computational agents can save labor and improve search efficiency in automated markets. Therefore, the performance of computational agents are tested in agent-based models. If a newly designed agent is successful there, it will be employed in real automated markets.
7. Parallel experiments with real and computational agents. Combining both humans and computational agents to experiments can give insights to economic issues. Human-subject behavior can be used guide the determination of learning process for computational agents. At the same time, the behavior of computational agent can be used to formulate hypothesis about the root causes of observed human-subject behavior. The studies include examination of emergence of generally accepted medium of exchange (money) [11] and labor market simulations [12].
8. Building ACE computational laboratories. Computational laboratories provide

useful and simple way to create agent-based systems without the need of strong programming skills. Such frameworks would allow to study and perform controllable and replicable experiments. Numerous systems have been developed and are available through the Internet. Please see [17] for the list of such computational laboratories.

The theme of the current thesis lies in the bottom-up modeling of market processes. The focus is on simulating real macroeconomic processes at micro level. The idea is related to a research by Basu et al. [13], which created an agent-based model for simulating the U.S. economy. The model is called ASPEN and is being developed at the Sandia National Laboratories. ASPEN individually models the actions of economic decision-makers and aggregates the actions to obtain macroeconomic quantities. ASPEN simulates the U.S. economy as a closed economy, which means without international trade and money flows. The macroeconomic model includes households, usual firms, firms producing capital goods, banks, federal reserve (central bank) and a government sector. The model has markets for goods and money. The behavior of agents is based on the behavior of real economic agents. The agents representing firms use genetic algorithms for learning how to operate competitively in the market place. ASPEN was run on massively parallel Paragon computer to obtain required performance. The ASPEN was parameterized to resemble U.S. economy and several economic scenarios (policies and external shocks) were played out with the model. The results met the expectations from the common economic theory.

The current thesis has different aim than ASPEN. Although the idea is to create agent-based macroeconomic models, the goal is to develop framework for modeling specific models for the given economic theory, not for the whole economy. Secondly and more importantly, the thesis wants to create models based on real empirical data.

2.4.5 Interaction Protocols for Economics

Agent-based computational economics, especially those dealing with macroeconomic issues, are based on market mechanisms. Market mechanisms have also been used for coordinating groups of computational agents with minimal direct communication among the agents [14]. Markets describe everything that is of interest of agents in prices. Thus, using market mechanisms for modeling economics is a natural extension. A market is a medium where participants can trade specific standardized goods. Agents bid for goods at various prices, but all exchanges occur at the current market prices. The agents buy or sell through markets to meet their design goals.

There are different types of markets. The markets can be based on discrete and continuous time. Discrete markets handle bids after a certain period, whereas continuous markets allow continuous fulfillment of orders. Secondly, the markets can be dealer- or order-based. In the former the dealer quotes bid and ask prices, at which she is willing to buy and sell. All agents can trade goods through the dealer, leaving the dealer in charge of the balancing the demand and supply by changing the bid and ask prices. Order-based markets rely on automatic matching of orders without a central dealer. Unlike in dealer-based markets where the submitted order always succeeds, in order-based markets the order might wait forever or be denied, if its price or size was inappropriate.

The current paper will employ order-based markets with discrete time. The market clears all orders after a certain period of time. Orders that cannot be committed at the market price are rejected. The detailed reasoning behind the choice is provided in the Modeling Method part. For implementation details see Section on Realization.

Another widely used option in ACE is the contract net protocol [14]. The contract net is an interaction protocol used for cooperative problem solving among agents. It is modeled after the contracting mechanism used by businesses to control the exchange of goods and services. The contract net creates a network of agents, that is a multiagent system. In such network agents can contract more customized orders than through market mechanisms. The contract net protocol is needed when markets do not provide necessary flexibility. The main steps in the protocol occur as follows:

1. A manager agent announces the task or good that needs to be performed or purchased through (possibly selective) broadcast.
2. Receivers evaluate the proposal and if interested submit bids.
3. Then the manager analyzes the responses and chooses the best one. If no suitable exist the manager can choose to abandon the offer.
4. Results are communicated to the contractors and the chosen one performs the order.

The roles of agents are not fixed in advance, each agent can act as a manager, as a contractor, or even as both.

The markets and contract nets are the main building blocks of agent-based models of economics.

2.5 Summary

The two approaches, namely statistical and agent-based, are complementary to each other. Their general relationship to the economic theory and real data is depicted in the Figure 3. First the researcher proposes an economic theory, which she thinks describes the reality well. As the theory needs to be tested with empirical data, the expectations of the theory need to be quantified. The standard way is to employ statistical techniques to create econometric model for the time-series. The possibility examined by the current paper is to use an agent-based model as the link between the economic theory and the actual data. Though, the agent-based modeling method should not be viewed as competing with the standard econometric techniques. It is complementary to them and might provide different insights.

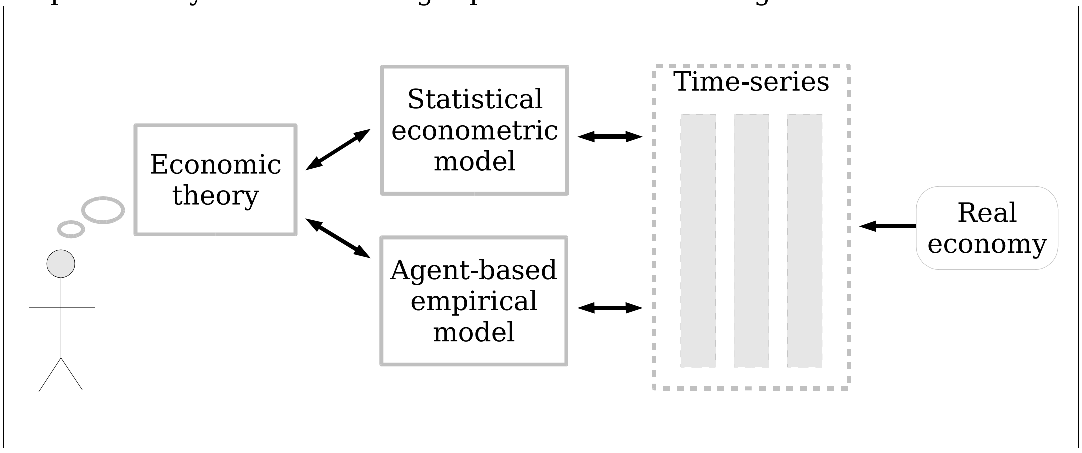


Figure 3 Empirical modeling of economics

An agent-based model provides a different view on the same economic theory. The agent-based approach might be especially useful, when the economic process is too complex to be only represented by simple time-series. Therefore, in the next section the paper will lay down the general ideas and specific details of an agent-based empirical modeling technique.

3 Modeling Method

The goal of the modeling method is to obtain an agent-based model that replicates the behavior of the economic processes characterized by the provided quantitative data. This section will introduce the main ideas and tackle several difficult questions that require addressing before it is effectively possible to build quantitative models for economic theories using agent-based means.

3.1 General Idea

A researcher is interested in modeling a theory that describes economic phenomenon P that is in an environment E. This phenomenon (or process) and its surroundings are based on a multiagent system and partially characterized by a set of time-series. Researcher's goal is to determine whether there is positive or negative support for her model in the given data. In other words, the question is whether the real process, represented by its time-series, can be generated by the suggested computational model using given data. For such models the term empirical agent-based models is adopted, defined as follows:

> **Definition.** An **empirical agent-based model** is a multiagent system that using the provided input time-series models a real process with a goal to quantitatively predict the data (time-series) that characterize the process. As a result an empirical agent-based model must generate output series, which predict series of the given process.

This means that empirical agent-based models can be viewed as functions, in a very broad sense. They receive input in the form of time-series and generate one or several output time-series. Figure 4 depicts the idea graphically. However, the empirical agent-based model is not an ordinary function, because it is based on interactive computation. The interactive computation is usually distinguished

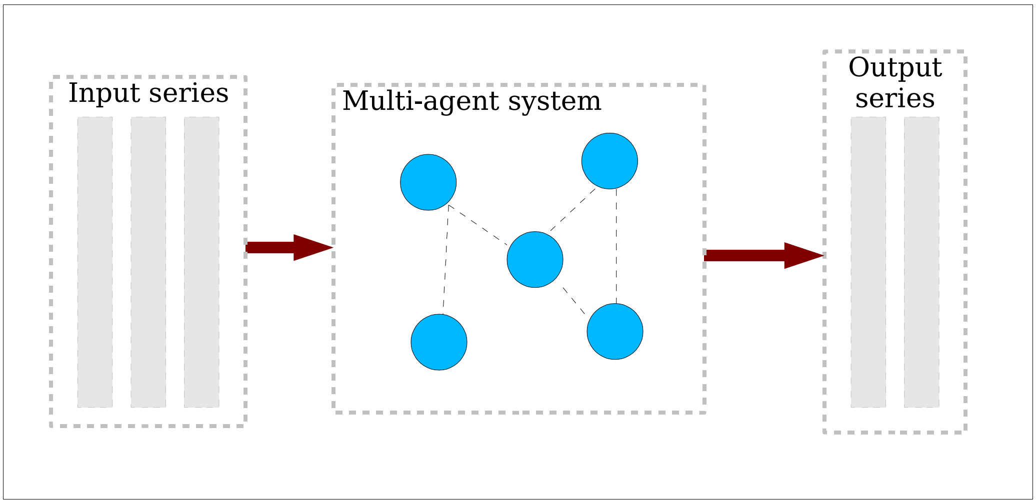


Figure 4 Empirical agent-based model.

from ordinary functions, since they continuously receive additional input from other processes.

As seen from Figure 4, the multiagent system acts as a computational model that transforms the given input time-series to output time-series. This is the central idea of empirical agent-based models, compared to ordinary agent-based models that do not require inputing and outputting time-series. The goal of empirical agent-based models is similar to the goal of statistical time-series models, although the latter use a simple mathematical form (usually additive) instead of a multiagent system for transforming the input time-series to an output time-series. From now on the empirical agent-based models will be also referred as just agent-based models, unless stated differently.

Developing the general idea to a modeling method requires dealing with several problems, which can be classified into 3 groups:

1. **Micro structure of the model**. This covers the types of agents, possible interactions between agents, communication protocols, environment. Since the goal is to arrive at the model that is able to test hypothesis on the basis of actual data, the micro-structure has to incorporate the input data and provide a way to extract output data.
2. **Initialization and parameterization**. This involves the initial setup, as well as the configuration of the agents and environment. Since agent-based models are inherently complex the quantitative setup of the model is crucial.
3. **Testing the success of the model.** It must be possible to test hypothesis about the model. Thus, techniques and methods are required for checking whether the model explains the actual data.

Next sections deal with these 3 areas.

3.2 Micro-structure of the Model

The micro structure of the central multiagent system has to meet several requirements:

1. integration to discrete period-based time
2. use of interaction protocols of economics
3. input data integration
4. output data integration

The first item in the list comes from a property of empirical agent-based models, namely, that the multiagent system will generate time-series resembling the original processes. Time-series represent time in discrete periods. To transform the time-series the multiagent system must be somehow linked to discrete time. There are two design options to solve this problem. First option, the one chosen by this paper, is to make the whole multiagent system work in an environment, where all interactions and decisions are done in turn-based fashion. Thus, the agents act once (or some other number of times) in a period and processes, like market clearing, are performed once in a period. Though it is not very realistic as there is a lot of asynchronous activity in the real economic processes, it is a good approximation if the size of the period is not too large. The main advantage is the direct mapping between the model period and the period in the time-series. Also it simplifies the design of agents. The second option is to use continuous time in the multiagent system and make the results discrete later. This approach is useful when the processes are too dynamic to map them into discrete periods. However, it is difficult to incorporate the data from the time-series into the model.

Second issue is connected to the fact that the underlying field of modeling is economics. This allows to limit the possible ways of communication and standardize on specific methods of interaction, greatly simplifying the creation of agent-based models. The idea is to provide clear model architecture by minimizing the use of “custom” communication. The main interactions between agents should take place through markets or contract nets (see Conceptual Basis section for the overview of these concepts). An example of the architecture is shown in Figure 5.

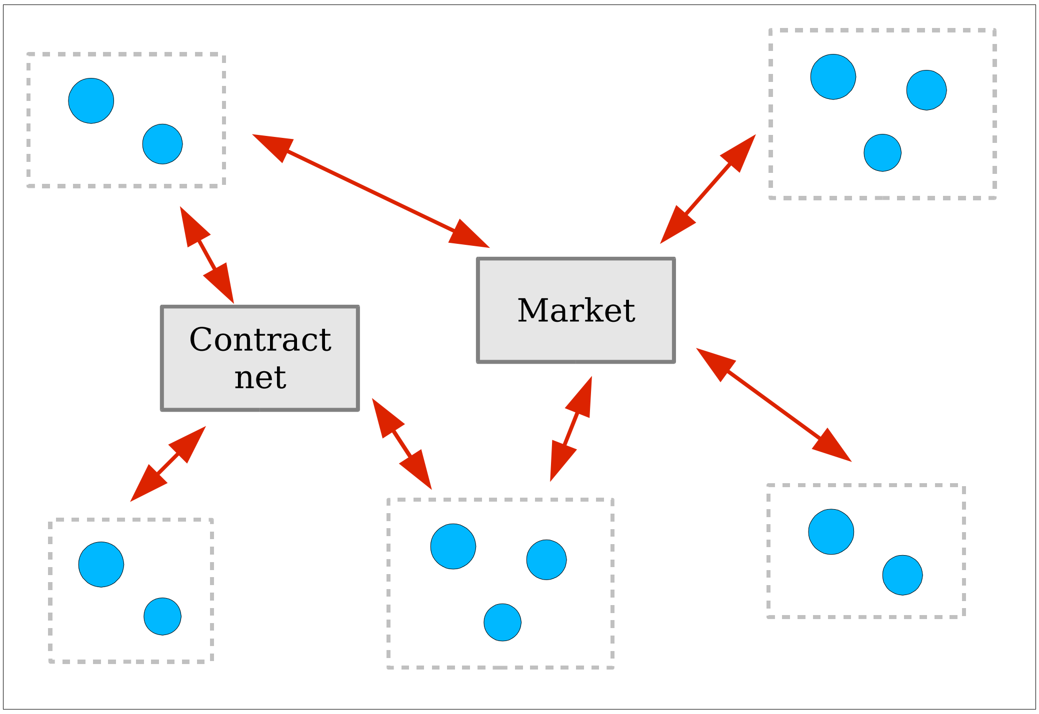


Figure 5 *Interaction protocols in a multiagent system of an agent-based economic model. The circles represents agents that can operate and interact through market(s) and contract net(s).*

As shown in Figure 5, the agents only interact through markets and contract nets. Although many economic models can be built by just using markets and contract nets, sometimes more sophisticated mediums are needed. For example, modeling of bank-runs[1] should take into account the interactions between single agents, because it is the way rumors propagate. The same is true for financial markets, where rumors play an important role. Eventually it is the decision of the modeler to choose which agents can participate in particular markets or contract nets, and what other means of interaction is allowed. Even though this paper will focus on creating economic models using just standard interaction protocols, the subsequent methodology applies to all agent-based models, notwithstanding the choice of interaction protocols.

Until now no comments have been made how exactly a multiagent system receives input and gives output as was previously depicted in Figure 4. The problem is actually trivial, in the case of period-based simulation environment. The integration of input time-series into the model can be done through the use of agents that operate on the basis of input data. Their decision process relies partially or totally on the data from these time-series. Thus, these agents connect

1 A bank-run is a situation when most of the bank's depositors decide to take out their money at the same time, due the fears of collapse of the bank, and hence create serious cash problems for the bank.

the agent-based model to the actual process. These agents are called input agents. An input agent is not restricted to use only one time-series, but as well can use several, if needed.

Similarly, the output of the multiagent model comes through agents. Some agents, called as output agents, perform the output function during their decision process. Their decisions are recorded for that purpose and aggregated to find a single value for a period for each output series. If possible, another reasonable option is to take the clearing quantities or clearing prices of the markets (or contract nets) as the output series. Obviously, this can be done only if the modeled phenomena are prices or quantities of the goods traded. The output and input agents can be mixed, meaning that one agent can simultaneously base its activity on input data and also provide the output.

Usually all necessary micro information for setting up the model is not and cannot be known by the researcher. Therefore, the issue of parameterization has to be considered.

3.3 Parameterization

In an ideal situation all information is available to the researcher. Thus, allowing to create the agent-based model and see how well it operates. But in reality empirical information is scarce and modeling methods need to incorporate some technique to overcome the problem.

In practice the way the exact model is constructed depends on the amount of information available. The most important consideration for agent-based models is whether micro-level data is present. If it is possible to obtain the data on the agent-level, one could with little effort build an agent-based model and input the quantitative data to simulate the real processes. On the other hand, when no such data is available, it might still be useful to construct an agent-based model, given that the micro-structure of the economic process is known.

To solve the data deficiency problem the initially unknown values are parameterized. This means that some of the model's behavior depends on a variable, whose “correct” value is not known. These changeable parts of the model affect the whole outcome. Thus, it the idea is to create a generic model that is then parameterized for the given case. The parameters are not known and need to be found with the help of modeling methodology. Given an agent-based model M two types of parameters can be differentiated:

1. **Structural parameters**. These affect the structure of the model M. For example, the ratio of speculators and long term investors in a stock market model.

2. **Behavioral parameters.** These affect the decision making process of agents in M. For example, consumers' price which they are willing to pay for a milk.

Let the model M have n parameters: $param_1$, $param_2$, ..., $param_n$. Then its parameter space consists of all possible values of the parameters. Also it is practical to limit the parameter space of each $param_i$ by a minimum and maximum value. The parameter space should be derived from the economic theory that the agent-based model is based on.

The main problem is how to find the “correct” parameter values, those which would be the best at explaining the underlying process. It has to be noted that there is no requirement that the number of parameters has to be equal to the number of input time-series. The number of parameters can be equal, larger or smaller from the number of input time-series, depending on the model. Now the paper will look at possible ways of addressing the problem of how to determine the “correct” parameters.

3.3.1 Simple Fixing of Parameters

Fixed parameters and initial setup allows easily to test the model. All parameters of the model are given a value from the parameter space and the model is run. In the case of deterministic models (not containing stochastic processes) just a single run is sufficient for the evaluation. If the model would be non-deterministic, containing some stochastic processes, the simulation has to be run several times. The data supports the proposed parameterized agent-based model if the resulting time-series are good, meaning that they match the actual time series, are unbiased, or have other similar properties (e.g. variance) to actual time series. This methodology is mostly used by in papers about stock markets or general macroeconomic studies. For example Aspen [13] used such method for building an agent-based model of U.S. economy.

The critical problem with this approach is that the parameters are usually not known. Thus, fixing the parameters will be just trying to guess them. Instead, the current paper suggest to intelligently guess them by an automated search for the “best” parameters. Therefore, in the next section the paper covers the ideas behind the dynamic parameterization.

3.3.2 Dynamic Parameterization

For this section the paper assumes that model has only one output time-series, for the multidimensional modeling see the part on Multidimensional models. The goal of parameterization is to automatically find the “best” parameters for the model. For automation an iterative technique has to be used. Statistical regressions use analytical methods, but these are not applicable for agent-based models. This is due to the complex nature of multiagent systems, implying that no analytical form can be computed. Therefore, to find the best values the model is re-simulated with different parameters. The values of the parameters are chosen from the specified parameter space. Simple iterative technique involves iterating through the parameter space with a chosen step size. Although nothing guarantees that the best value was not missed due to large step size, the appropriateness of step size can verified by analyzing the obtained R^2 series (see section on Chaos in the Model for more details).

For each set of parameter values the model needs to be simulated and its performance evaluated. There are several possible criteria for identifying the "goodness" of the given set of parameters, given the output time-series process:

- Least Square Error Fitting (LS)
- Least Absolute Error Fitting
- Match of particular statistics (like variance, stationarity statistics, specific movements).

The paper adopts the first criterion (least mean squared error), because it is also widely used in statistical regressions and thus serves as a good comparison point. Even though least absolute error fitting is considered to be more natural, as it weights all errors equally, it is not used in statistics due its unsuitability for analytical mathematics. Namely, its residuals (errors) cannot be treated as a continuous differentiable quantity [16]. The third criterion is often used for examining the results of stock market simulations.

With the least square error method, the fitness of a model is how much of the variance of the modeled time-series it is able to explain. This method finds parameter values, for the given computational model, that minimize the squared error between the real and estimated values. Squared error is defined as follows:

$$SE = \sum \left((real_i - estimated_i)^2 \right)$$

where SE is squared error, $real_i$ is i^{th} value in the time series of the data being modeled and $estimated_i$ is i^{th} value estimated by the computational model M, for the given parameter values.

In the context of agent-model this is a classical search problem for finding the parameters that produce the smallest squared error value. This is the same as maximizing the R^2 statistic over the parameter space. R^2 is a statistic that is defined as following:

$$R^2 = 1 - SE / Var(real)$$

where SE is squared error, as defined above, and Var(real) denotes the variance of actual time series. Thus, R^2 shows how much of the variance of the actual time series is explained by the computational model. R^2 of 1 means that model perfectly explains the movements of actual time series, whereas value of 0 implies that no explanation is given. It should be noted that R^2 can also receive a negative value, which means that the model explains the data worse than the best constant value (which is the mean of the actual time series) can.

The problem can be considered as a usual search problem allowing to use standard search techniques.

- Blind search. This involves exhaustive trial of all parameter values to find the best ones. It is the simplest approach, but is enormously inefficient, since it does unnecessary model running.
- Dynamic parameter iteration. This is a modification of the blind search. During the iteration process, the parameter step is chosen dynamically instead of using a constant step value. If a small change in parameter values has a constrained (limited) effect on the R^2 (this is the case when R^2 is not behaving chaotically) it is possible to safely enlarge the step size in some cases, without the risk of missing a good parameter value combinations. The cases when it is safe to use larger step sizes are those where the previous R^2 value is small. The difference of R^2, measured after a small change in parameter values, shows how much the parameter values affect R^2. If the effects are small, the step size can be increased. Similarly, if the R^2 is high (i.e. close to maximum values of R^2) then the step size is decreased to achieve better accuracy. This method can avoid useless simulations by decreasing the frequency of simulation in poor areas of the parameter space and increasing the frequency of simulation in good areas of the parameter space.
- Heuristically guided search. A natural heuristic for guiding the maximization is R^2. This technique involves starting from a specific parameter values and moving around in the parameter space. This approach is called hill-climbing and involves moving towards higher values of R^2. Thus, the search avoids areas of the parameter space, which do not have high R^2 values. The heuristically guided search is definitely faster than blind search. The problem is that the search may get locked into a local maximum and, thus, does not find the best parameters. This problem needs to be solved by combining some safety checks for skipped areas or combining heuristically guided search with other techniques.
- Selective deepening and rejection. The technique starts by iterating the parameter space with low frequency (i.e. large step size). Then the areas are analyzed and only promising areas are searched further. Other areas are rejected. This is repeated until necessary accuracy is achieved.

The current thesis implements only the simple blind search, leaving the realization of advanced methods for further research.

3.4 Testing Chaos in the R^2

It is critical to know whether, for the given parameter space and iteration frequency, the R^2 statistic behaves chaotically. For that purposes the statistical noise tests suit well (e.g. white-noise tests for time-series). The R^2 values from the simulation are gathered into several groups, a group for each parameter. Each group describes how the changes of this parameter affect the R^2 statistic. If there is high noise in R^2 for a particular parameter, it implies that R^2 behaves chaotically and the range of that parameter might be either too small or too large. On the other hand,if there is low noise in the R^2 it implies that the iteration frequency is sufficiently small for the parameter to capture its effect on R^2 and thus no troublesome chaos in R^2.

The development and application of specific noise tests is left for further research. See [3] for more details on the noise tests.

3.5 Hypothesis Testing

Parameter testing as done in statistical models is not possible. Regressions use statistics to measure the significance of a variable. The precisely same approach for agent-based models is impossible since it not possible to use analytical form for evaluating the effects of time-series on the resulting R^2 value. The actual outcome emerges from the behavior of agents, which is a complex interactive computation. Therefore, the paper proposes following three methods for checking the validity of the model:

1. Comparison of several decision rules
2. Input time-series testing by using randomized time-series
3. Examination of parameter values

First method is useful if the hypothesis concerns behavior of agents. The decision making process of agents has to be tested by comparing explanatory power of different decision rules. Hypothesis about agents' behavior comes down to whether the rules used by agents create a multiagent system whose behavior can explain the data. Following simple testing method can be used:

The explanatory power of the model is evaluated when the decision logic is removed or replaced by another rule. If the explanatory power will be higher with the original rule it can be considered as significant behavior, since its removal would worsen the model.

Thus, the comparison of different models is done by measuring their respective R^2. The higher the measure the better the model. Though high R^2 does not simultaneously imply that the model is good. If the decision process can be can be divided into separable subparts, it is possible to test the significance of every part separately. If the model is good, all parts of the decision process tested have to be significant. That means that they have to contribute to the explanatory power of the model. Thus, the significance of the behavioral rules of the agents is important factor for evaluating the performance of the model, the same way the significance of the parameters is important contributor to the success of a statistical model. Let the decision process be divided into m parts, the added explanatory power by each part is found as follows:

$$added_i = R^2_{original} - R^2_i$$

where i is number from 1 to m, $R^2_{original}$ is R^2 of the original model and R^2_i is R^2 of the model without the i^{th} decision component. If the added explanatory powers of all components are high, the model is good. On the other hand, the low contribution of some components would cast doubt on the whole model and especially to the suitability of these components.

When the first method allowed to test the significance of decision rules, the second method provides a way to evaluate the significance of the input time-series. Though, it is analytically impossible to separate the effects of different inputs, due to the inseparability of decision processes, it is possible to test whether the input data gives the model more explanatory power. Each input time-series is tested by replacing it by a randomly generated time-series in the original model and finding the explanatory power of such model. The process should be repeated several times and results (R^2 statistics) compared to those of the original model. Similar or higher R^2 statistics of the randomized-input model compared to the original model are evidence of insignificance of the input time-series. Likewise, significantly (in statistical meaning) smaller R^2 values for randomized-input model signals the significance of the time-series and the model as whole. The randomization of a time-series should take into account the properties of the original time-series and generate similar random series. It is useful to replicate following properties:

- Variance
- Mean
- Auto-regressive properties (stationarity, non-stationarity)

Replicating these properties in the randomized time-series provides closer match with the original series. This is important since the parameter space is fixed for the empirical agent-based model. Thus, the parameter space makes assumptions about specific properties (mean, variance, etc) in the time-series. The use of appropriate randomized series allows to compute the significance statistics for each input series. Although the calculation of a precise significance statistic requires huge number of simulations, it is only required to know whether the originally received R^2 is above or below some threshold. Thus, the required number of runs is not very large, especially if the time-series is strongly significant or strongly insignificant.

Although the formulas for the calculation of the statistics are out of the scope of this paper, the general idea is depicted in Figure 6. For each time-series the model is simulated with several randomized series. For each time-series these simulations generate a distribution of R^2 values. An example distribution is depicted in the figure. If the R^2 of the actual model is sufficiently large, compared to the R^2 of randomized simulations, then this given time-series is significant. The point where “sufficiently large” starts is called a critical value and is determined by the significance level. The significance level is a statistical concept, which in the current context means the maximum probability of accidentally accepting an insignificant series as significant. Usually economic researches use 1% or 5% for the significance level. Although the distribution of R^2 is not known (since it depends on the model), it is possible assume normal distribution or also apply non-parametric tests (e.g. Wilcoxon tests), which do not make any assumptions on the distribution. The paper performed few runs with randomized input-series with the example simulation (See Simulation Example section).

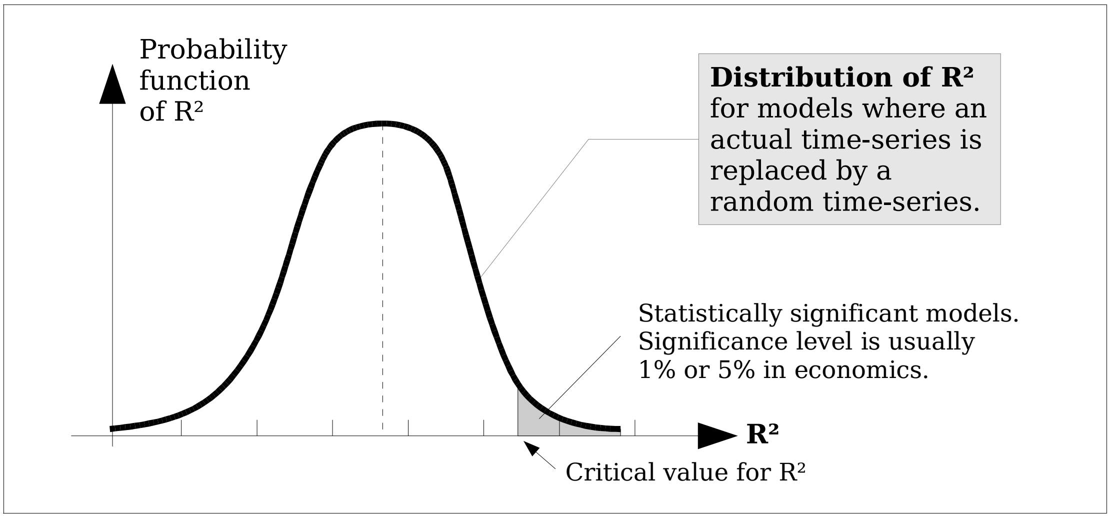


Figure 6 An example of a distribution of R^2 values for randomized time-series models.

Lastly, it is possible to evaluate the fitness of the model by examining the obtained parameter values in context of the theory. If properties of these values (the signs and magnitude of parameters) match the theoretical expectations, the model is supported, otherwise there is doubt about the fitness of the model. Also if the value of the best parameter lies in the edge of its space, then it might be the case that the “best” parameter is actually outside the predefined parametric space. Then it is recommended to re-simulate the model using larger parameter space.

3.6 Complexity of the Algorithm

It is important to know the effect that the number of parameters has on the running time of the search algorithm. The complexity of the algorithm can be expressed as an asymptotic complexity function from the number of parameters and chosen testing frequencies. The testing frequency of a parameter is the number of different values it is assigned. Thus, it is inversely related the step size. The complexity of the simple iterative method with n parameters is just the product of the parameter frequencies:

$f(n) = O(frequency_1 * frequency_2 * ... * frequency_n)$

If all frequencies are given the same constant value, then the asymptotic complexity of the search algorithm becomes n-th power of the given constant and thus, is valued also as:

$f(n) = O(2^{n})$

Therefore, the simple search algorithm is not suitable for models with large number of parameters. Already models with 3 parameters create difficulties, depending on the internal complexity of the model. Thus, use of optimized searches is of high importance.

3.7 Multidimensional Models

Multidimensional modeling means that the model has several output time-series, instead of just one, as was assumed before. In such case a simple R^2 statistic, measuring the explanation power of one predicted time-series, is insufficient. There two main straight-forward options for multidimensional agent-based modeling: sequential and aggregated.

Sequential modeling involves separating the model into sub-components and modeling each sequentially, starting from the component that does not depend on any other. Each component can have only one output series and is modeled using the standard least square technique. The results of already modeled pieces can be included in the modeling of next components. Sequential modeling assumes that such separation is possible and interdependencies do not exist between different components. Though usually not applicable, sequential method has speed benefits, since the number of parameters that need to be found on a single step is decreased.

Simulation through aggregation of the goal heuristic is another method. The whole model is simulated in one step by adopting a single goal statistic that considers all R^2-s of the output time-series. A weighted average can suit well for this purpose. The weights are chosen according to the importance of the time-series. Aggregated method is slower than sequential because all parameters have to be tested simultaneously, but it can be used without restrictive assumptions.

The particular details involved in the realization of multidimensional agent-based modeling are out of the scope of this paper. Thus, multidimensional aspects will not be considered further.

3.8 Summary

For a short summary, the proposed modeling method requires following steps to create an agent-based model:

1. Setup of agents, their decision rules, interaction mechanisms and environment.
2. Choosing input-agents that base their action on the time-series.
3. Output time-series generation mechanism (either through agents or results of market trades).
4. Determination of parameters (either decision parameters or structural).
5. Fixing of parameter ranges, using theoretical considerations. Also econometric models can provide sufficient data.
6. Simulation with the least squared error principle and finding the parameter values for the best R^2.
7. Measuring significance of time-series and testing reliability of the results (testing of chaos).

4 Realization

The realization of the modeling technique was done on Java platform (Sun Java, version 1.4.2). For the software see the accompanied data medium and Appendix A for instructions. The implementation required to make several choices, such as choosing the interaction methods and market model. Nevertheless, the general results of the simulations are not significantly affected by these choices. The three main realization components were the simulation process, the market setup (price clearing method), and the object model of the software implementation. The following sections discuss these components and the choices made. The main principle behind the choices was simplicity and flexibility of the realization.

4.1 Simulation Methodology

The simple iterative search for best parameter values is implemented by the class SimulationRunner, which guides the running of the simulations (see section 4.3 for the Object Model). The SimulationRunner is given the simulation, an object of the class Simulation, (see section 4.3 for the Object Model), which is then "run". The Simulation object holds the agent-based model and provides the actual run() method, which performs a single simulation for the given parameter values. The SimulationRunner iteratively performs several simulation runs with different parameter values and searches for the values resulting in the maximum R^2 value.

For the given parameter values, a single simulation run consist of running all periods of the model. The number of periods is specified by the model. Each period all agents make their decisions and put out selling and buying orders to the markets that are available to them. Then the markets are cleared with successful trades completed and messages about the trades are sent to the traders. The data sets and agents are reset before the next run.

For the Java source code of the simulation loop, see the accompanied data medium for the SimulationRunner.calculateParams() method. This method finds the best parameter values in the preset parameter space. For the implementation of a single run of the model, see method Simulation.run(), which performs the simulation with the given parameters. Both of these classes reside under the agentmodeling namespace.

4.2 Market Setup

As already mentioned, the thesis uses order-based markets with discrete time for the implementation. Discrete time allows easier integration of the time-series data than continuous time solutions. Order-based approach was favored for dealer-based markets due to simpler realization. For the detailed argumentation behind the choices see section Micro-structure of the Model.

The software realization is done in such way that each tradable has its own market. Since the market is based on discrete time the clearing of the market takes place each period. At the beginning of the period traders submit buy and sell orders for the given tradable. When all traders have chosen their actions the clearing price is found by choosing the price that sets the demand equal to the supply. At the end of the period all unmatched orders are removed. For the exact implementation see source code of methods Market.findClearingPrice() and Market.informDeals() located in the accompanying data medium.

4.3 Object Model

The main relationships in the object model of the realization are depicted in Figure 7 (See Appendix B for more detailed UML model).

SimulationRunner is a class that performs the running of instances of Simulation class and provides methods for retrieving the results of the simulation. The output contains the parameter values and R^2 values.

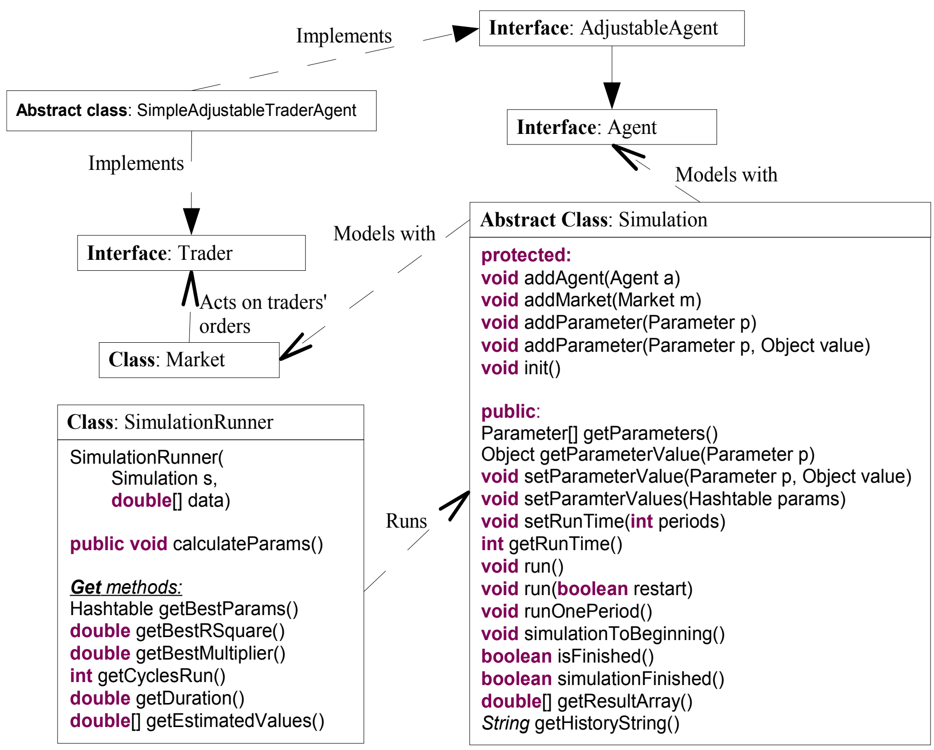


Figure 7 UML model of the realization, main relationships.

Simulation is an abstract class, which is extended by all actual simulation models. The simulations can re-implement its methods where necessary and need to provide a constructor (or multiple constructors) for creating the particular simulation. The constructor of a subclass of Simulation class should initialize the model by setting up agents, markets and parameters. The running of the model is handled by the methods of parent Simulation class, namely its run() method.

To represent agents the simulations use objects, which implement Agent and Trader interfaces. The decision logic of agents is implement in the Agent.performActions(int period) method. The markets agents are allowed to trade on and environment that is observable to agents is completely left to the implementation of agents (or to the particular Simulation subclass).

Objects representing markets do not need to subclass the class Market, since the already class provides all necessary functionality and a practical constructor. Thus, simulations just create instances of Market class and pass them on to agents, who can trade there.

4.4 Optimization

There are several methods for optimizing the search algorithm, such as selective deepening and heuristic search, which were discussed in the Modeling Method section. As mentioned before, the realization does not implement them.

Other optimization possibilities arise from the repeated process of running the same simulation, just with changed parameters. Thus, there are several places where caching can be employed. For example, the current realization employs the caching of decisions of input-driven traders. The caching can be further extended to speed up the process. The simulation running times were decreased by 30% when the input-driven agents stored the input series and automatically used them, instead of doing method requests to the respective Simulation object. Further optimization is possible when a group of agents acts always the same way. These agents can be then replaced by a collective agent that always does the actions the same way.

5 Simulation Example

The thesis paper uses a model from an economic research for evaluation of the modeling method. The idea is to re-simulate the model with the agent-based method and compare the results with statistical regression performed in that research. Using a previous research avoids dealing with the specific methodology of economics and saves the current paper from collecting necessary data and performing parallel statistical regressions. Doing all of these actions would have been too far from the scope of this paper.

5.1 Economic Theory

The main criteria for the choice of the economic study were use of time-series regression, simplicity of the model and availability of used data. Simplicity here means that the concepts used would not be too sophisticated for a person without deep knowledge of economics. The paper was chosen from JStore database (http://www.jstor.org)[1]. The chosen paper is a statistical study by Olson and Curtis about the development of wheat prices in Great Brittan before the World War I [15].

The economic paper studies a major decline in wheat production in Great Britain during the period from 1873 to 1914. The cause for the 50% reduction of wheat production is argued to be the decrease in the world market price of wheat at that time. The importers sold cheaper foreign wheat in Great Britain. The fallen prices discouraged local farmers who then responded by using less land for growing wheat. For that reason the economic research tries to model the amount of land the British farmers used for wheat. In their main model the researchers use two time-series to explain the land under wheat:

1. Seven year average for previous prices of wheat
2. Previous year barley-to-wheat price ratio

The idea is that when farmers plan their year during the spring, they do not know the price they can sell the wheat in autumn. Thus, they will look back at previous years to estimate how much land to use for wheat. The researchers hypothesize that the farmers were looking at general trend in the price of wheat of the previous years. Thus, seven year average for previous prices of wheat shows the general trend. Additionally, farmers are hypothesized to consider barley, the main alternative wheat, because they both required land and no extra equipment was necessary for switching to barley. The higher the price of barley in relation to wheat, the less land the farmers should put under wheat. Thus, the general relationship used by the economic paper in its regression was:

1 About 30-40 papers were looked through dating from 1960s to 1980s.

Acres(t) = a + b_1*previous-wheat-prices(t) + b_2*barley-ratio(t-1)

where a is a free constant, b_1 and b_2 are parameters for the two explaining time-series.

5.2 The Agent Model

Simulating the actions of farmers with an agent-based model is straight forward, since no special interaction mechanisms are required. The only interaction medium for agents is the wheat market. The wheat market is the environment for the following three types of agents that are in the model (see Figure 8):

- Farmers
- Wheat importers
- Wheat consumers

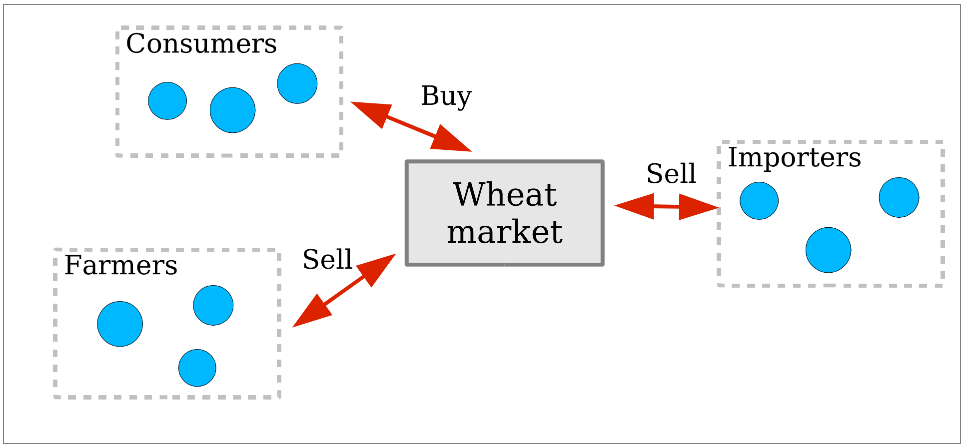


Figure 8 Wheat market model.

All farmers are agents who each period decide who much land they use for wheat growing. They receive two input time-series, namely previous average prices of wheat and last year's barley ratio. At the end of period farmers sell all their production, at the market price, which means that they are price takers. Wheat importers have their world market price, at which they can sell as much as possible. Thus, their price decisions are based on that single input time-series. Wheat consumers, similarly to farmers, are price takers, meaning that they buy their amount with the given market price. Wheat consumers' demand is assumed to be insensitive to price and above the local production. The output time-series is generated by farmers. The acres they chose to use for wheat are aggregated and stored.

The agent-based model also has two parameters. These are the farmers' sensitivity to previous prices and sensitivity to barley-wheat price ratio. Both of them are behavioral parameters, meaning that they affect the decision process of the agents. In that way the agent model exactly replicates the model from the original work. Therefore, this agent-based model does not possess any value from the point of view of the field of economics. The decision rules of the farmer are shown in Figure 9.

```
public void performActions(int period) {
      // c was found by matching the first period acre value.
      acers = c +
            priceSensitivity*previousPrices[period-1] +
            priceRatioSensitivity*priceRatios[period-1];
      acers = Math.max(0,acers);
      market.addSellOrder(this,0,acers);
      // totalAcres is the output time-series
      totalAcres[period-1] += acers;
}
```

Figure 9 Farmer's decision process.

5.3 Simulation Results

The simulation is run using frequency of about 40 for both parameters. The initial parameter space is derived from economic theory: price of wheat should increase acres, and barley-to-wheat price ratio decrease acres. See Appendix C for the details of the simulation.

The total of 1560 simulations were run achieving the maximum R^2 of 0.8997. This means that the model explains about 90% of the variation, which is a good measure. The high R^2 is also accompanied with parameter values that do not lie in the edge of parameter space. The original work received a value of 0.95 for R^2, but these are not directly comparable due to the different modeling techniques. It can be concluded that both, the statistical and agent-based, models exhibit high explanatory power.

The simulations can be also examined in terms of the R^2 matrix, which depicts the R^2 statistic for each evaluated point in parameter space. This is depicted in Figure 10.

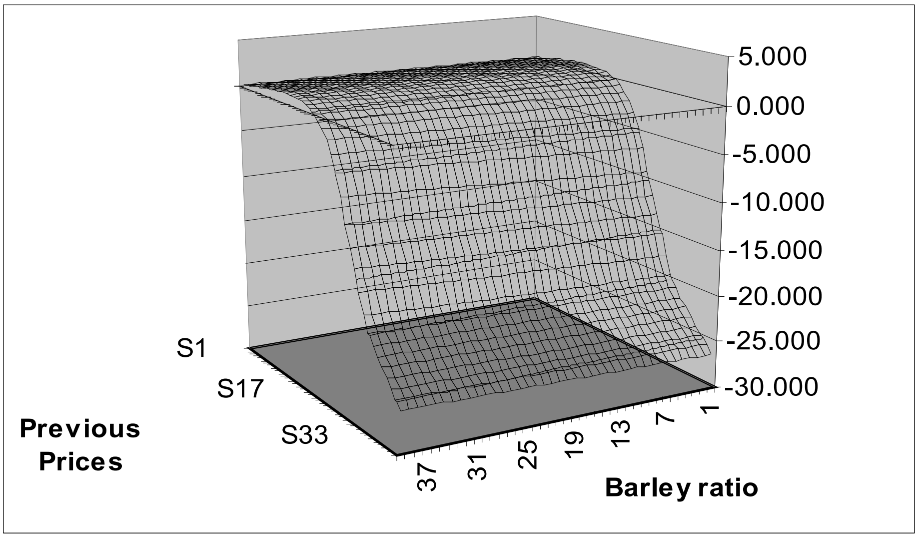


Figure 10 R^2 matrix for the whole parameter space (labels of previous prices and barley-to-wheat ratio do not reflect the correct parameter values)

The Figure 10 clearly shows that for some parameter values the R^2 statistic falls significantly below 0. This shows the potential for optimizations. Also, the effect from the barley-to-wheat ratio is small, compared to the effect of previous average prices. This implies that the barley ratio parameter space does not reach values, which would significantly reduce the R^2. See Appendix D for results with enlarged parameter space. These results also confirm that the original choice of parameter space was correct, since the maximum point is still in initial region.

The negative values of R^2 in Figure 10 skewed the R^2 axis too large. Interesting area is around the positive values of R^2. For that view the negative values were cut and the result is presented in Figure 11.

The Figure 11 shows a fast rise in the values of R^2 in a specific section of the parameter of the previous price sensitivity. That suggests that the parameter space chosen too wide. Most of the simulation is done in the part where the R^2 values are negative. Thus, more efficient methodologies could have saved halve or even more of the simulation steps, just by cutting out the negative part.

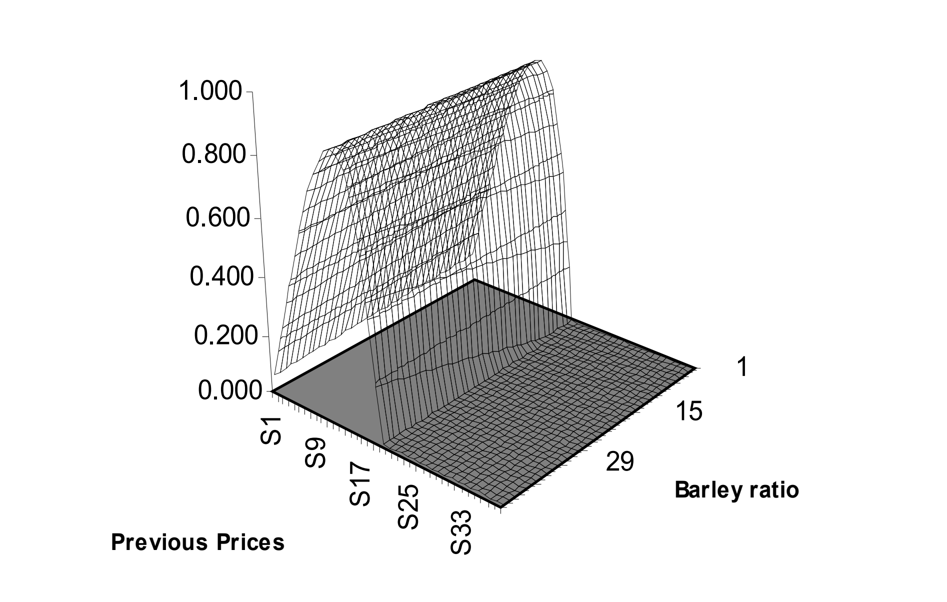


Figure 11 Positive values of the R^2 matrix for the whole parameter space. (The labels of x- and y-axes are just numeration)

As was suggested in the Modeling Method section, one of the ways to test the significance of time-series is to generate random series that have the similar properties of the original series. These random series are then replaced instead of the original series and model is simulated. If the resulting R^2 values are significantly less (in statistical terms) than the R^2 of the original simulation, then the original time-series is significant.

Although the thesis did not do a thorough application of the idea, a sample case was simulated. The time-series of previous prices that was used in the original model was replaced by 5 random series (see the accompanied data medium for the series). These random series were generated using normal distribution for the first differences of the series. The normal distribution had a mean and standard deviation of the first difference of the original series. The resulting values of the R^2 statistics are reported in Table 1. For the values of the randomly generated time-series see the accompanied data medium.

R^2 values
0.85583
0.84195
0.84011
0.8621
0.39381
0.75876 (average)

Table 1. R^2 values from 5 simulations with randomized series.

As the five R^2 values show, the original time-series seems to be significant, since it obtained a R^2 statistic of 0.8997, which is higher than all of the values above. However, these 5 simulations are not sufficient to determine the statistical significance. More sophisticated statistical framework needs to be developed that could precisely outline the methodology for creating the random series and statistical tests for determining the significance. As mentioned, this is, though, out of the scope of this paper.

6 Evaluation

In the Conceptual basis section the paper defined a set of factors that are important for good modeling methods. The developed model is now viewed in the light of these factors.

- **Reliability**. The reliability of an agent-based method is satisfactory given that the parameter space is correctly specified. Improved search techniques, such as selective deepening, will simplify the choice of the parameter space even further, since larger spaces can be used. Only problems can arise when the model performs chaotically, the R^2 statistic is very sensitive to minute changes of parameter values. Such models will suffer from poor reliability of the results. Though, the chaos can be detected and analyzed by examining the behavior of R^2 statistics.
- **Speed of modeling.** Creating an agent-based model is more time-consuming than a statistical model, if done by hand. Though, it is not tremendously longer process, if the good user interfaces are available. The agent-based modeling can be visualized to reduce the gap even more.
- **Sufficiently fast simulation.** The computational simulation is a problematic part for agent-based modeling, especially for models using larger number of parameters. In simple parameter search, adding a parameter to the model increases the running time by a factor of 20-40, depending on the required frequency. Thus, with simple parameter search the modeling is limited to maximum of 2-3 parameters. The implementation of advanced search techniques is crucial for simulating models with 4-5 parameters. Also parallel search techniques using distributed systems can be considered as an option for alleviating the issue.
- **Possibilities of modeling.** The main area where the empirical agent-based modeling shines is the large variety of possibilities. It is possible to create models with complex agents and economic structures. This feature is very attractive and can lead to choosing agent-based models instead of statistical ones.

6.1 Implications

The analysis and the results of the sample simulation indicate that the agent-based empirical modeling method has potential to become a tool for empirical research in economics. The thesis does not propose that agent-based empirical modeling should replace purely statistical methods, but instead complement them. The thesis proposes that the empirical agent-based models could be constructed aside the statistical models to gain further insight to the underlying economic phenomenon. The statistical model can provide information for choosing the parameter space and agent population of the agent-based model.

Although the provided implementation of the modeling method suffers from slowness these problems can be alleviated by taking advantage of more advanced search techniques, such as heuristically guided search.

6.2 Further Research

There are several areas that are left open by the current research. Firstly and most importantly, the method needs to be applied for a real economic research problem. This will show the usefulness of the method in real settings. Second area of further studies involves optimizing the simulation search algorithm. The goal would be to achieve feasibility for models with at least 4-6 parameters.

Also the statistical tools can be further developed to more rigid testing of chaotic behavior of R^2 and significance of time series in the model.

7 Conclusion

The current thesis proposed an agent-based modeling method for empirical economic research. Agent-based models provide a natural possibility to model more complex economic relationships and phenomena than standard econometric methods. Employing agent-based models allows to build the model from bottom up and, thus, resemble the actual ideas of the economic theory. The example simulation shows that use of empirical agent models is feasible. Though, some hurdles remain to be solved, such as improving the speed of the simulation and developing precise statistical means for the testing purposes. The thesis proposes that such agent-based models should be applied together with econometric models for researching complicated economic phenomenon.

Abstract

The M. Sc. thesis develops an empirical agent-based modeling method for economics by building upon the concepts of Agent-based Computational Economics (ACE). The method uses multiagent systems to model economic processes. The novelty of the thesis is the attempt to investigate the possibility of using multiagent systems for empirical modeling of economic systems. The goal of such modeling is to test empirical validity of economic theories.

The modeling method proposes integration of the agent-based models with time-series. This would allow to employ agent-based models in empirical testing and modeling. The data is linked to the model through agents and the performance of the resulting model is evaluated using least squared error method. The least squared error method is implemented by iterating through the parameter space. Paper proposes several iterative techniques for searching the parameter space, including simple exhaustive search, heuristically guided searches and selective deepening.

Finally, the thesis tests the modeling method on an example economic study by creating a parallel agent-based model to the original statistical model. Results show potential for the method, as high explanation level was reached. The paper includes the software of the modeling method developed for the thesis. The implementation was done with Sun Java platform.

Abstract in Estonian

Magistritöö arendab empiirilise agenttehnoloogial baseeruva majanduse modelleerimise meetodi, kasutades Agent-based Computational Economics (ACE) ideid. Meetod kasutab multiagent süsteeme majandusprotsesside modelleerimiseks. Agentmudelid sobivad hästi majanduse modelleerimiseks, sest majandusprotsessid põhinevad üksikute inimagentide tegevusel. Magistritöö uudsus seisneb selles, et ta uurib võimalust kasutada multiagent süsteeme eksisteerivate protsesside modelleerimisel, sidudes mudelid empiirilise infoga välismaailmast. Modelleerimise eesmärgiks on majandusteooriate empiiriline testimine.

Loodud modelleerimise meetod pakub välja võimaluse, kuidas agentpõhised mudelid integreerida aegridadega. See võimaldab agentmudelite kasutamist empiiriliseks testimiseks. Aegread on seotud mudeliga läbi agentide ja saadava mudeli edukust hinnatakse vähim ruutude meetodi abil. Vähim ruutude meetod on realiseeritud iteratiivse otsinguga parameetrite ruumis. Töö pakub välja mitmeid iteratiivseid tehnikaid parameetrite ruumi läbiotsimiseks, näiteks nagu lihtsa täieliku otsing, heuristiliselt juhendatud otsing ja valiv süvendamine.

Töö testib meetodit ühe majandusuuringu peal, luues seal kasutatud statistilise mudeli kõrvale empiirilise agentmudeli. Tulemused näitavad tulevikku agenttehnoloogial baseeruvale meetodile, sest saavutati kõrge seletustase. Töö sisaldab loodud modelleerimise tarkvara, mille abil realiseeriti näidismudel. Rakendus loodi kasutades Sun Java platvormi.

References

[1] L. Tesfatsion. “Agent-Based Computational Economics: Growing Economies from the Bottom Up.”, ISU Economics Working Paper No. 1, December 2001.

[2] Gärtner, Manfred. “Macroeconomics”, Ashford Colour Press, 2003 (pp 1-20, 434-437).

[3] Gujarati. “Basic econometrics”, Boston [etc.] : McGraw-Hill, 2003.

[4] M. Wooldridge and N. R. Jennings. Intelligent agents: Theory and practice. The Knowledge Engineering Review, 10(2): 115-152, 1995.

[5] Ferber. “Multiagent Systems: Introduction to Distributed Artificial Intelligence”, Addison-Wesley, 1999.

[6] Rust, J., Miller, J. H., and Palmer, R. “Characterizing effective trading strategies: Insights from a computerized double auction tournament.” Journal of Economic Dynamics and Control 18, 61-96, 1994.

[7] Klemperer, P. “Auction theory: A guide to the literature”, pp. 3-62 in P. Klemperer, “The Economic Theory of Auctions”, Cheltenham, UK: Edward Elgar, 2000.

[8] Vriend, N. J. “An illustration of the essential difference between individual and social learning, and its consequences for computational analyses.”, Journal of Economic Dynamics and Control 24, 1-19, 2000.

[9] Wilhite, A. “Bilateral trade and `small-world' networks.”, Computational Economics 18, 2001.

[10]Prietula, M. J., Carley, K. M., and Glasser, L. “Simulating organizations: Computational models of institutions and groups.” Cambridge, MA: The MIT Press, 1998.

[11]Duffy, J. “Learning to speculate: Experiments with artificial and real agents.”, Journal of Economic Dynamics and Control 25, 295-319, 2001.

[12]Pingle, M., and Tesfatsion, L. “Non-employment benefits and the evolution of worker-employer cooperation: Experiments with real and computational agents.”, Economic Report 55, Iowa State University, 2001.

[13]Basu N., Pryor R. J., and Quint T. “ASPEN: A Microsimulation Model of the Economy”, Kluwer Academic Publishers, 1998.

[14]Weiss, Gerhard, “Multiagent Systems. A Modern Approach to Distributed Artificial Intelligence.”, The MIT Press, 2001.

[15]M. Olson, C. Harris. Free Trade in "Corn": A Statistical Study of the Prices and Production of Wheat in Great Britain from 1873 to 1914. The Quarterly Journal of Economics © 1959, The MIT Press, pages 145-168.

[16]MathWorld. Least Square Fitting. http://mathworld.wolfram.com/LeastSquaresFitting.html (last visited 01.05.2004)

[17]List managed by Tesfatsion. Listing of ACE computational laboratories: http://www.econ.iastate.edu/tesfatsi/acecode.htm (last visited 01.05.2004)

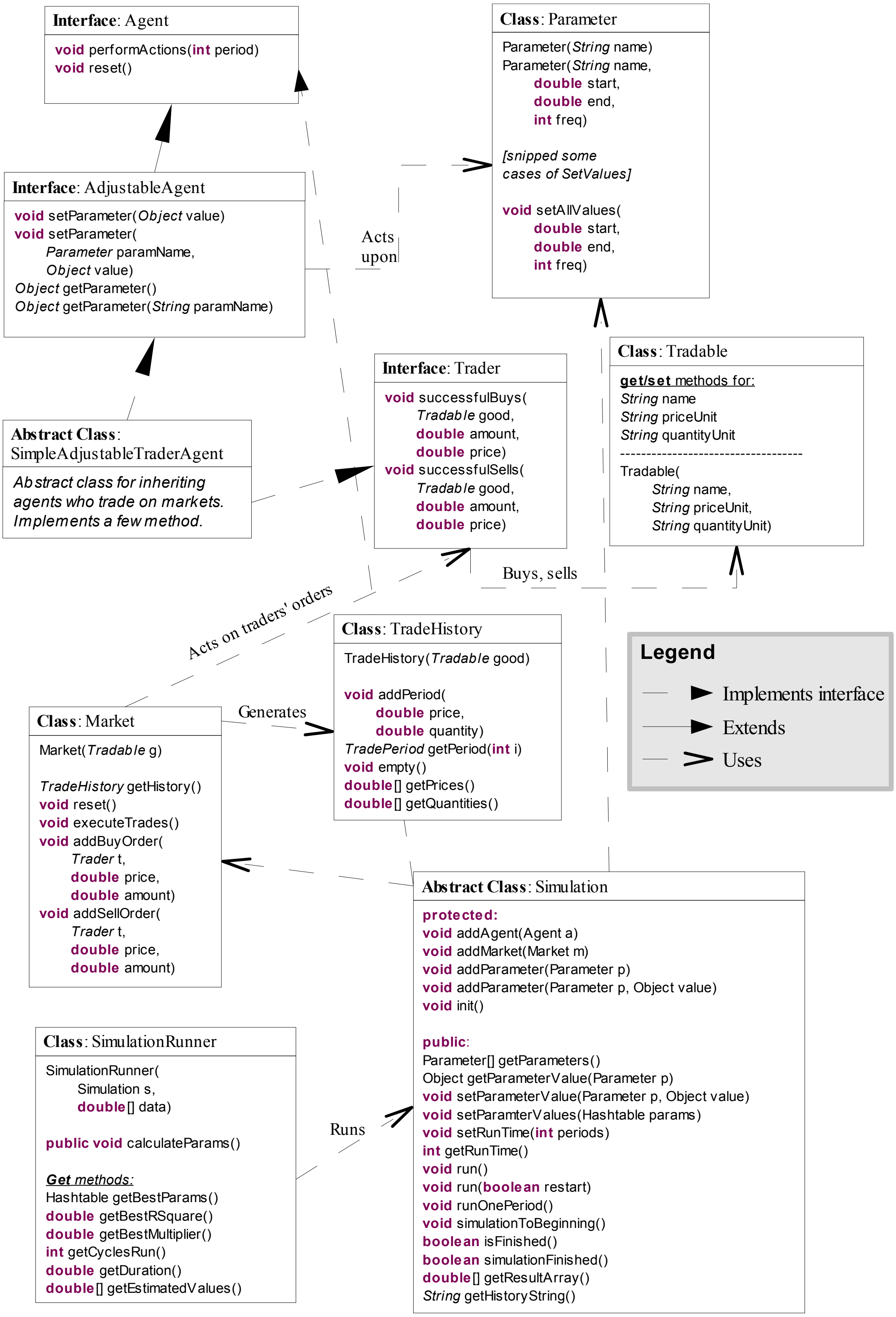
Appendix A

For the instructions of running the software see the accompanied data medium (CD)for file “README.txt”. The contents of the file is also presented here.

```
ABOUT
-----
This is software developed for M.Sc. thesis by Jaak Simm (jaaksimm@ut.ee).
The software implements empirical agent-based modeling in Java platform
(tested with Sun Java 1.4.2).
The implementation was done with Eclipse IDE software,
which is available for free at http://eclipse.org.

The software can be run and compilied without Eclipse, see below.

CONTENTS
--------
The folder hierarchy matches Java class namespace, of course.
The zip file contains source and also compiled version.

agentmodeling/
 - Main classes, implementing common functionality

agentmodeling/agents/
 - Some simple example agents for testing

agentmodeling/simulations/
 - Some simple example simulations

agentmodeling/utils/
 - arraytools for manipulating arrays in Java
 - statistictools for performing statistical analysis (variance, mean, etc)

corn/
 - The example simulation from the thesis
 - data.txt contains the data from the example!
 - results.txt contains some results with different models
   (different time-series, different parameter spaces)
 - random.txt contains randomized series for previous prices
   (used in Model 3 for randomization)


COMPILATION (using javac)
-------------------------
To recompile the source using standard Java compiler (javac):
1) Enter directory "economic-model"
2) javac <java-file-name>
   (e.g. javac corn/Farmer.java)


RUNNING THE EXAMPLE SIMULATION
------------------------------
1) Enter directory "economic-model".
2) To run the example simulation located in directory "corn":
   java corn.WheatSimulation

3) You can specify options for running the simulation:
   java corn.WheatSimulation [-zoomout] [-output_rsquares] [-datafile <filename>] [-model <number>]

   Explanation:
   -zoomout = uses larger parameter space (results were also reported in the thesis)
   -output_rsquares = outputs all calculated R^2 values
   -datafile <filename> = uses the specified data file instead of default ('corn/data.txt')
   -model <number>, where the <number> is:
      1: the default model:
         Acres = const + param1*price(avg) + param2*ratio

      2: Acres = const + param1*price(-1) + param2*ratio

      3: Model simulation with random series (using the series from corn/random.txt).
```

Appendix B

The object-oriented class relationship diagram model represented in UML. The model depicts main classes. For all classes see the included source code.



Appendix C

Results of the example simulation where obtained using following parameter space:

p_1-space: [0 ; 0.000375]

p_2-space: [-0.025 ; -0.0125]

The result of the simulations were:

R^2 = 0.89977

dimension multiplier = 0.9374 (multiplier that is applied to the output series from the agent model to obtain the estimate of the real time-series)

Best parameter values:

p_1 = 1.18421E-4 (farmers' previous price sensitivity)

p_2 =-0.01875 (farmers' sensitivity to barley-wheat ratio)

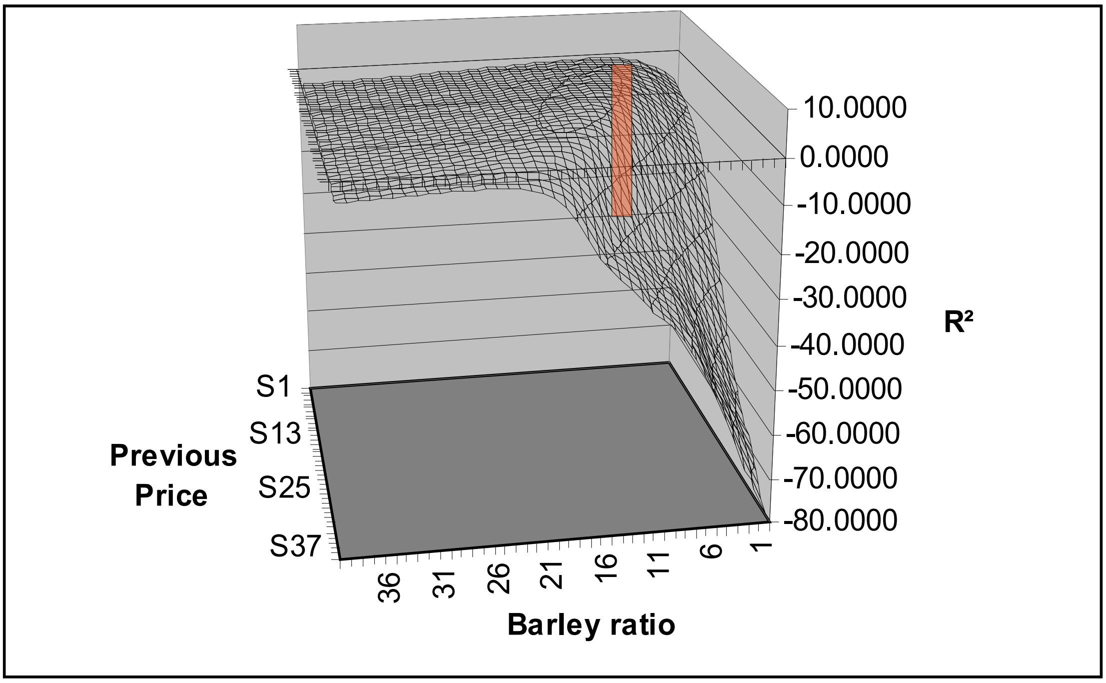
Appendix D

Following figure shows the R^2 statistics matrix from the example simulation for enlarged parameter space:

p_1-space: [-0.000125 ; 0.000375]

p_2-space: [-0.5 ; 1.25]

The highlighted square (red in colored version) in the graph notes the location of the original parameter space as used in the example. All edge values of the parameters are below 0, confirming the correctness of the initial choice of parameter space.



Despite the loss in precision the simulation still performed well, loosing only 0.005 in R^2:

R^2=0.89513 (was 0.89977)

multiplier=0.94732 (was 0.93741)

p_1 = 1.250E-4 (was 1.18421E-4)

p_2 = -0.0125 (was -0.01875)