

# Impact of Propagation Information in the Model of Tax Audit

Elena Gubar, Suriya Kumacheva, Ekaterina Zhitkova,  
and Olga Porokhnyavaya

**Abstract** An effective tax system is an important part of economic and social interactions in human society. The key element of the tax system is tax control which provides the main functions of taxation and allows for increasing tax revenue and fees to the state budget. However, total tax audits of a population of taxpayers is economically unreasonable, and even selective tax audits are not always profitable. In this case the propagation of information can be viewed as an “infection of the mind,” and its spread shows an interesting resemblance to that of epidemics. We thus use a modification of the classical Susceptible-Infected-Recovery model to describe the process. We assume that information propagates through the population by pairwise contacts between spreaders and others in the population and Informed agents disseminate information through their network of contacts or social networks. We study a model of spreading information in a large population of taxpayers and describe the dynamics of this process in complex social networks. We formulate an optimal control problem of tax auditing and analyze the behavior of agents in different subgroups depending on received information.

**Keywords** Tax control • Information spreading • SIR model • Epidemic process • Optimal control • Social networks

## 1 Introduction

One of the most important aspects of modeling taxation is tax control and so it has been the subject of continuous interest in much recent research. The standard model of taxation used to describe the behavioral relationship between taxpayers and the tax authority is static. For example, [1, 17–19] formulate mathematical models of tax evasion and auditing. Two of the most famous works [17] and [18] applied

---

E. Gubar (✉) • S. Kumacheva • E. Zhitkova • O. Porokhnyavaya  
Faculty of Applied Mathematics and Control Processes, St. Petersburg State University,  
Universitetskii prospekt 35, Petergof, St. Petersburg 198504, Russia  
e-mail: [e.gubar@spbu.ru](mailto:e.gubar@spbu.ru); [s.kumacheva@spbu.ru](mailto:s.kumacheva@spbu.ru); [e.zhitkova@spbu.ru](mailto:e.zhitkova@spbu.ru);  
[olga.porokhnyavaya@gmail.com](mailto:olga.porokhnyavaya@gmail.com)

a game-theoretical approach to describe the taxation problem for the first time: they have presented the interaction between tax authority and taxpayers based on a hierarchical game “principal-to-agent.” As well as in [1] and [19], optimal strategies are defined as optimal scheme or optimal contract, including tax and penalty rates and the probabilities of auditing.

However as tax collection is a periodic event, this problem can be considered as a dynamic process which occurs, for example, on a yearly basis. Moreover, it has been proved that a total tax audit is quite expensive and so the tax authority must define a method which helps to collect taxes while minimizing the costs of inspections. In particular, spreading information about the positive social benefits of tax collection or possible penalties for tax evasion through social networks and media provides a tool of control of large group of taxpayers.

A major difference between the models which have been studied in the past and our current study is that we combine an approach based on game theory with one accounting for the process of spreading information. Recent studies have been shown that the process of spreading information resembles an epidemic process and hence it is possible to use a modification of the Susceptible-Infected-Recovered (SIR) model to characterize the propagation of information.

As in classical SIR model we consider a large but finite population of taxpayers which is divided into several subgroups subject to the relevance of information. The entire population may be sorted out to Uninformed, Informed, Indifferent, and Resistant. The subgroup Uninformed consists of agents, who do not have information about a future tax auditing campaign. Informed agents received information and disseminate it if it is important to them. Indifferent taxpayers get information, but is not interesting to them and they may not transfer it.

We suppose that information is propagated through the population by pairwise contacts between spreaders and others in the population of taxpayers. The important fact is that if someone has adopted information and it is important, or she believes in it, then the agent is capable of spreading it to others.

In real life, social networks offer a good platform for interactions among agents in a population and many taxpayers have extensive social contacts and can disseminate information through their contacts network. It is a fact that in a social network information spreads rapidly through different channels without many restrictions and it is possible to consider the Internet and social networks as an effective tool for the propagation of information. However people are more likely to believe news from their friends and relatives while network information must be verified and from time to time as it has a lack of credibility. Hence we have to take into account a group of agents who ignore the received news. The scale-free structure of the Internet implies that each agent with an access to the social network has a statistically significant probability of having a very large number of contacts, which can be estimated by the average connectivity of the network.

During the past decades different models for the propagation viruses and information in networks have been developed. One of the first papers, applying epidemic processes to the spreading of the rumors, ideas, and information, is [3]. In [14] the spreading of computer virus over the network is considered as the propagation of epidemic process.

In this paper, we establish a complex control-theoretic model to design tax authority control strategies through the propagation of information and advertising the bonuses of participating in tax collection campaign to mitigate the impact of nonpayment on society. Information transmission can be represented by dynamics on a graph where vertices denote individuals and an edge connecting a pair of vertices indicates interaction between individuals. Due to a large population of people involved in the process of spreading rumors and information, random graph models such as scale-free networks in [7, 13] are convenient to capture the heterogeneous patterns in the large scale complex network.

This paper is organized as follows. In Sect. 2, we review a static game-theoretical model of tax control. In Sect. 3, we formulate and analyze the complex model of propagation information through the social network in a population of taxpayers; we formulate an optimal control problem and present a structure of an optimal program for spreading information by the tax authority, employing Pontryagin's maximum principle. In Sect. 4 we show a modification of algorithm which forms a scale-free network. Finally, we present simulations and conclusions about the model and discuss the impact of parameters to the system so that we can come up with suggestions for possible preventative or control methods.

## 2 Static Model

Based on a game-theoretical model presented in [1] in this section we present a static model of tax control. In the mentioned model the tax authority (high level of the hierarchy) and  $N$  taxpayers (low level of the hierarchy) are players. Each taxpayer has income level equal to  $i_j$ , where  $j = \overline{1, N}$ . At the end of every tax period the  $j$ th taxpayer can declare her income as  $r_j$  which can be less or equal to her true income  $i_j$  ( $r_j \leq i_j$  for each  $j = \overline{1, N}$ ).

After collecting the tax returns the tax authority audits taxpayers with the probability  $\bar{p}$ . The tax auditing supposed to be absolutely effective, that is, it reveals the existing evasion.

Let  $\xi$  be the tax rate,  $\pi$  be the penalty rate (these rates are assumed to be constants). If the evasion is revealed as a result of a tax audit, then the evaded taxpayer should pay unpaid tax and the penalty, which depends on the evasion level:  $(\xi + \pi)(i_j - r_j)$ .

Then the  $j$ th taxpayer's expected payoff is defined from the equation

$$\beta_j = i_j - \xi r_j - \bar{p} (\xi + \pi)(i_j - r_j), \quad (1)$$

where the first summand is always paid by the taxpayer (pre-audit payment), and the second—as the result of the tax auditing—made with probability  $\bar{p}$  (post-audit payment).

Then the tax authority's net income can be defined as

$$J = \sum_{j=1}^N (\xi r_j + \bar{p}(\xi + \pi)(i_j - r_j) - \bar{p} \cdot \bar{c}), \quad (2)$$

where  $\bar{c}$  is the unit cost of one audit. Let's define it as

$$\bar{c} = \frac{B}{\nu}, \quad (3)$$

where  $B$  is the tax authority's budget,  $\nu$  is the share of the audited taxpayers:

$$\bar{p} = \frac{\nu}{N}. \quad (4)$$

Naturally, every players' aim is to maximize their expected payoffs.

For obtaining the further results we should use the next proposition, which was formulated for the model, described above, and proved in [1].

**Proposition 1.** *Let the inequality*

$$(\xi + \pi)i_j \geq \bar{c}, \quad (5)$$

*be fulfilled for the subset  $\{1, N_0\}$  ( $N_0 < N$ ) of  $N$  taxpayers. The optimal tax authority's strategy is  $p^* = \frac{\xi}{\xi + \pi}$  for every  $j = \overline{1, N_0}$ . The  $j$ th taxpayer's optimal strategy is*

$$r_j^*(\bar{p}) = \begin{cases} 0, & \text{if } \bar{p} < p^*, \\ i_j, & \text{if } \bar{p} \geq p^*. \end{cases}$$

*Let the inequality (5) be not fulfilled for every  $j = \overline{N_0 + 1, N}$ . The optimal tax authority's strategy is  $\bar{p} = 0$ . The  $j$ th taxpayer's optimal strategy is  $r_j^*(0) = 0$ .*

The first case of the Proposition 1 is about the optimal strategy of tax authority in terms of what the tax audit is profitable for it (inequality (5) is satisfied). In response, the optimal strategy for lower level players is to decide to pay taxes or not, depending on the probability of being audited, chosen by the top player. This result is similar to the "threshold rule," obtained in [19] for another mathematical model of tax control.

The second case is a pessimistic situation. In this case, the tax authority does not have sufficient funds to carry out the necessary tax audits. Taxpayers are rational, and in these conditions they can afford not to pay anything.

### 3 Dynamic Model of Spreading Information

As we indicated above, taxation is a regular process and can be formulated as a dynamic model that takes into account the dissemination of information as a factor to stimulate tax compliance.

Let's assume that at the first moment of the process the audit probability  $\bar{p} = 0$ . The total population of rational taxpayers evades of taxation in accordance with the second case of the Proposition 1.

In practice there is no information about the relation of parameters in (5) for every  $j = \overline{1, N}$ , therefore, the tax authority does not know whether auditing is profitable or not. Moreover, as the tax authority's budget  $B$  is strongly limited, auditing with probability  $p^*$  is practically impossible.

Therefore, the tax authority has to stimulate unaudited taxpayers to pay tax that corresponds to their true income level. The means of such stimulation is spreading information about future audits to the taxpayers (which can be irrelevant in general). This information makes rational taxpayers think that the audit probability is high enough that paying taxes is less costly than evading them and risking having to pay back-taxes along with the penalties. Within the framework of this model, this information is given by inequality

$$\bar{p} \geq p^*, \quad (6)$$

where  $p^* = \frac{\xi}{\xi + \pi}$  (due to Proposition 1).

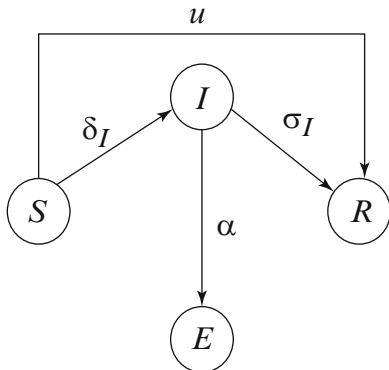
The tax authority spreads this information with the intensity  $u(t)$  (the share of the Informed taxpayers per unit time),  $u \in [0; \bar{u}]$ , where  $\bar{u}$  is the possible maximum value of control function  $u$ .

#### 3.1 Scheme of Spreading Information

Here we consider the process of spreading information over the network of contacts modeled as a scale-free network. Each node of such network represents a taxpayer, who gets the information and propagates it over her social contacts, internet, social networks, etc.

The entire population of taxpayers is divided into four subgroups according to their relation to spreading information (see [6, 8, 13]):

- Uninformed taxpayers  $S$  (we denote the number of agents in this group as  $n_S$ ). They do not have any information about future auditing and, therefore, do not pay taxes ( $r_j = 0$  due to Proposition 1).



**Fig. 1** Scheme of information spreading in a taxpayers population

- Informed taxpayers  $I$  (group  $n_I$ ). These taxpayers receive information and propagate it: at first, they pay the taxes corresponding to their true income levels ( $\xi i_j$ ); second, they spread the information over Uninformed taxpayers.
- Indifferent taxpayers  $E$  (group  $n_E$ ). They get information, but do not propagate it: they do not spread the information over Uninformed taxpayers and pay the taxes and penalties ( $(\xi + \pi) i_j$ ) if and only if they were audited;
- Resistant taxpayers  $R$  (group  $n_R$ ). The taxpayers from this subgroup are those who lost their interest in the information, because they paid taxes and propagated the information or, vice versa, did not propagate it and, so, paid penalties. In any case, the information becomes irrelevant for them.

Let's denote shares of Uninformed, Informed, Indifferent, and Resistant as

$$S^k = \frac{n_S}{N}, I^k = \frac{n_I}{N}, E^k = \frac{n_E}{N}, R^k = \frac{n_R}{N},$$

where  $S^k + E^k + I^k + R^k = 1$  and  $k$  is the degree of each taxpayer's connections at time  $t$ . Initial states are  $I^k(t_0) = I_0^k > 0$ ,  $E^k(t_0) = E_0^k > 0$ ,  $R^k(t_0) = R_0^k > 0$ ,  $S^k(t_0) = 1 - I_0^k - E_0^k - R_0^k$ .

The scheme of information spreading in a population of taxpayers is presented in the following diagram. See Fig. 1.

### 3.2 Constructing the Aggregated System Profit

In the current study, the aggregated system profit consists of two different parts: first, profit which was received by the tax authority from the propagation of information and, second, tax auditing. This step gives us the following conclusions.

The first conclusion is that in the model examined there is a two-component budget

$$B = \int_0^T (b_1(\bar{p}(t)) + b_2(u(t))) dt, \quad (7)$$

where  $b_1(\bar{p}(t))$  is the cost of auditing with probability  $\bar{p}$  and  $b_2(u(t))$  is the cost for activating the spread of information: it is twice differentiable and increasing function in  $u(t)$ , such that  $b_2(0) = 0$ ,  $b_2(u) > 0$ , when  $u(t) > 0$ .

The second conclusion is that the aggregated system profit also consists of two components:

$$J = J_{aud} + J_{inf}, \quad (8)$$

where  $J_{aud}$  is the tax authority's net income, obtained as a result of auditing, and  $J_{inf}$  is the profit, obtained from spreading information.

The first summand is the post-audit payments of Indifferent taxpayers (from the subgroup  $E$ ) without total audit cost:

$$J_{aud} = g_E(E^k(T)) - b_1(\bar{p}(T)), \quad (9)$$

where  $b_1(\bar{p}(T))$  is defined from

$$b_1(\bar{p}(T)) = n_E \bar{p}(T) \bar{c}, \quad (10)$$

and the post-audit payments of the taxpayers from  $E^k$  are

$$g_E(E^k(T)) = (\xi + \pi) \bar{p} \sum_{j=1}^{n_E} i_j. \quad (11)$$

To simplify the following reasoning, we will substitute  $g_E(E^k(t))$  in the next continuous estimation:

$$\widehat{g}_E(E^k(T)) = (\xi + \pi) \bar{p} N E^k(T) \hat{i}, \quad (12)$$

where  $\hat{i}$  is the average taxpayers' income.

The second summand of the aggregated system profit is the profit from the propagation of information:

$$J_{inf} = \int_0^T (f_R(R^k(t)) - f_E(E^k(t)) - b_2(u(t))) dt, \quad (13)$$

where the integrand is a sum of the taxes  $f_R(R^k(t))$  collected from the Resistant taxpayers  $R^k$ , without the taxes unpaid by the Indifferent taxpayers from  $E^k$  and the cost of activating information spreading  $b_2(u(t))$ .

The first summand under the integral in (13) is

$$f_R(R^k(t)) = \xi \left( \sum_{j=1}^{n_R} i_j \right). \quad (14)$$

The second summand with minus is the unpaid taxes

$$f_E(E^k(t)) = \xi \left( \sum_{j=1}^{n_E} i_j \right). \quad (15)$$

We should use the continuous estimations for  $f_R(R^k(t))$  and  $f_E(E^k(t))$ , as it was done for  $g_E(E^k(t))$  in (12):

$$\widehat{f}_R(R^k(t)) = \xi NR^k(t) \hat{i}, \quad (16)$$

$$\widehat{f}_E(E^k(t)) = \xi NE^k(t) \hat{i}, \quad (17)$$

where  $\hat{i}$  is the average taxpayers' income.

$f_R(R^k(t))$  and  $f_E(E^k(t))$  (from (16) and (17) correspondingly) are non-decreasing and differentiable functions, such as  $f_R(0) = 0$ ,  $f_E(0) = 0$ ,  $f_R(R^k(t)) > 0$ ,  $f_E(E^k(t)) > 0$  for  $R^k(t) > 0$ ,  $E^k(t) > 0$ .

The cost for activating information spreading  $b_2(u(t))$  can be defined as

$$b_2(u(t)) = Nu(t)\tilde{c}, \quad (18)$$

where  $\tilde{c}$  is the unit cost of information spreading.

Thus, the aggregated system profit (the tax authority's net income) is

$$J = \int_0^T [f_R(R^k(t)) - f_E(E^k(t)) - b_2(u(t))] dt + g_E(E^k(T)) - b_1(\bar{p}(T)). \quad (19)$$

### 3.3 Constructing the System of Equations

We define a process of spreading information as a system of nonlinear differential equations corresponding to the scheme (Fig. 1). In our study, we use a modification of a classical epidemic model (see [10, 14]):

$$\begin{aligned} \dot{S}^k &= -\delta_I I^k S^k \Theta_I - u S^k; \\ \dot{I}^k &= \delta_I I^k S^k \Theta_I - (\sigma_I + \alpha) I^k; \\ \dot{E}^k &= \alpha I^k; \\ \dot{R}^k &= \sigma_I I^k + u S^k; \end{aligned} \quad (20)$$



where control

$$0 \leq u(t) \leq \bar{u} \leq 1, \text{ for all } t \in [0, T], \quad (21)$$

where  $\bar{u}$  is a boundary value of control;  $\delta_I$  is the rate of spreading of information in subgroup  $I$ ;  $\sigma_I$  is the rates of forgetting of information in subgroup  $I$ ; and  $\alpha$  is a probability that received information is not important for an agent.

$\Theta_I(t)$  represents a probability that any given link points to an Informed or Indifferent taxpayer (see [7, 14]), as

$$\Theta_I(t) = \sum_{k'} \frac{\tau(k')P(k'|k)I_{k'}^I}{k'}, \quad (22)$$

where  $\tau(k)$  denotes the infectivity of a node with degree  $k$  [7, 14]:

1.  $\tau(k) \leq k$ ;
2.  $\tau(k)$  is monotonically increasing;
3.  $\lim_{k \rightarrow \infty} \tau(k) = M > 0$ ;

$P(k'|k)$  shows the probability of a node with degree  $k$  pointing to a node with degree  $k'$ :  $P(k'|k) = \frac{k'P(k')}{\langle k \rangle}$ , where mean value  $\langle k \rangle = \sum_k kP(k)$ .

Within the framework of a model statement (20)–(21), we solve the optimal control problem. We find the optimal intensity of information spreading  $u(t)$ , which gives maximum to the functional (13)

$$J_{inf} = \int_0^T [f_R(R^k(t)) - f_E(E^k(t)) - b_2(u(t))] dt \rightarrow \max. \quad (23)$$

### 3.4 Optimal Control Problem of Propagation Information

We find the optimal propagation strategy  $u$  to the problem described above applying Pontryagin's maximum principle [5, 15]. We define the associated Hamiltonian  $H$  and adjoint functions  $\lambda_S, \lambda_I, \lambda_E, \lambda_R$  as follows:

$$H = -b_2(u) - f_E(E^k) + f_R(R^k) + (\lambda_I - \lambda_S)\delta_I S^k I^k \Theta_I + (\lambda_R - \lambda_S)u S^k + (\lambda_E - \lambda_I)\alpha I^k + (\lambda_R - \lambda_I)\sigma_I I^k. \quad (24)$$

Adjoint system is

$$\begin{aligned} \dot{\lambda}_S(t) &= -\frac{\partial H}{\partial S^k} = (\lambda_S - \lambda_I)\delta_I I^k \Theta_I + (\lambda_S - \lambda_R)u; \\ \dot{\lambda}_I(t) &= -\frac{\partial H}{\partial I^k} = (\lambda_S - \lambda_I)\delta_I S^k \Theta_I + (\lambda_I - \lambda_E)\alpha + (\lambda_I - \lambda_R)\sigma_I; \\ \dot{\lambda}_E(t) &= -\frac{\partial H}{\partial E^k} = f'_E(E^k); \\ \dot{\lambda}_R(t) &= -\frac{\partial H}{\partial R^k} = -f'_R(R^k); \end{aligned} \quad (25)$$

with the transversality conditions given by

$$\lambda_S(T) = 0, \lambda_I(T) = 0, \lambda_E(T) = 0, \lambda_R(T) = 0. \quad (26)$$

According to the Pontryagin's maximum principle [15], there exist continuous and piecewise continuously differentiable co-state functions  $\bar{\lambda} = (\lambda_S, \lambda_I, \lambda_E, \lambda_R)$  that at every point  $t \in [0, T]$ , where  $u$  is continuous, satisfy (25) and (26). In addition, we have

$$u \in \arg \max_{\underline{u} \in [0, \bar{u}]} H(\bar{\lambda}, (S^k, I^k, E^k, R^k), \underline{u}). \quad (27)$$

The derivative of Hamiltonian by  $u$  is

$$\frac{\partial H}{\partial u} = -b'_2(u) + (\lambda_R - \lambda_S)S^k \geq 0. \quad (28)$$

It is easy to see that Hamiltonian  $H$  reaches its maximum if condition (28) is satisfied.

According to the standard approach our main results are formulated in the following proposition and auxiliary lemmas:

**Lemma 1.** *Function  $\phi$  is decreasing over the time interval  $[0, T]$ .*

**Lemma 2.** *For all  $t, 0 < t < T$  the following condition holds  $(\lambda_R - \lambda_S) < 0$ .*

Proofs of Lemmas 1 and 2 follow the same technique as in [4, 11].

**Proposition 2.** *In the problem statement (20), (21) (23) optimal control  $u(t)$  has the following structure:*

- *When  $b_2(\cdot)$  is concave function for (23), then there exist the time moments  $\bar{t}, \bar{t} \in [0, T]$  such as*

$$u(t) = \begin{cases} \bar{u}, & \text{if } \phi > b_2(\bar{u})/\bar{u}, \text{ for } 0 < t < \bar{t}; \\ 0, & \text{if } \phi < b_2(\bar{u})/\bar{u}, \text{ for } \bar{t} < t < T. \end{cases} \quad (29)$$

- *When  $b_2(\cdot)$  is strictly convex function, then there exist the time moments  $t_0, \bar{t} \in [0, T], 0 \leq t_0 \leq \bar{t} \leq T$  such as:*

$$u(t) = \begin{cases} \bar{u}, & \text{on } 0 < t \leq t_0; \\ \text{is continually decreasing function,} & \text{on } t_0 < t \leq \bar{t}; \\ 0, & \text{on } \bar{t} \leq t \leq T; \end{cases} \quad (30)$$

where  $\phi = (\lambda_R - \lambda_S)S^k$  is switching function for control problem (20), (24), (25),  $\bar{u}$  is defined in (21).

Rewrite Hamiltonian in the following form:

$$H = -f_E(E^k) + f_R(R^k) + (\lambda_I - \lambda_S)\delta_I S^k I^k \Theta_I + (\phi u - b_2(u)) + (\lambda_E - \lambda_I)\alpha I^k + (\lambda_R - \lambda_I)\sigma_I I^k. \quad (31)$$

To prove the main statement of the Proposition 2 we consider two cases:

(1) Consider a case when  $b_2(\cdot)$  is concave.

Since function  $b_2$  is concave ( $b_2'' \leq 0$ ), then  $(u\phi - b_2(u))$  is convex function of  $u$  in (31) and for any  $t \in [0, T]$  it reaches its maximum either at  $u(t) = \bar{u}$  or  $u(t) = 0$ . From (31) we have that optimal  $u(t)$  satisfies  $u\phi - b_2(u) \geq \underline{u}\phi - b_2(\bar{u})$ , where  $\underline{u}$  is any admissible control,  $\underline{u} \in [0, \bar{u}]$ . If  $u = \bar{u}$ , then switching function is satisfied  $\phi \geq b_2(\bar{u})/\bar{u}$  and if  $u = 0$ , then  $\phi \leq b_2(\bar{u})/\bar{u}$ .

Lemma 1 suggests that  $\phi$  is decreasing function, then there can be at most one moment  $t \in [0, T]$  at which  $\phi(t) = b_2'(\bar{u})$ , moreover if such moment exists, for example,  $\bar{t}$ , then  $\phi(t) > b_2(\bar{u})/\bar{u}$  on  $0 \leq t < \bar{t}$  and  $\phi(t) < b_2(\bar{u})/\bar{u}$  on  $\bar{t} < t \leq T$ . Then, (29) is satisfied.

(2) Let cost function  $b_2(\cdot)$  be strictly convex.

If function  $b_2$  is strictly convex ( $b_2'' > 0$ ), then minimizer of  $(u\phi - b_2(u))$  is unique. Expression (28) implies that if  $\frac{dH}{du} = -b_2'(u) + \phi = 0$  at optimal  $u$  else  $u \in [0, \bar{u}]$ .

Thus, from continuity of functions  $\phi$  and  $b_2'$  follows that  $u$  is continuous at all  $t \in [0, T]$ . As far as  $b_2$  is strictly convex, then  $b_2'(\bar{u}) > b_2'(0)$ ,  $\bar{u} > 0$ . Lemma 1 requires that there exist time moments  $t_0, \bar{t}$ , such as  $0 < t_0 < \bar{t} < T$ , which are defined from the following conditions:

$$u(t) = \begin{cases} 0, & \text{if } \frac{db_2(0)}{du} \leq \phi; \\ \frac{db_2^{-1}(\phi)}{du}, & \text{if } \frac{db_2(\bar{u})}{du} \leq \phi < \frac{db_2(0)}{du}; \\ \bar{u}, & \text{if } \phi < \frac{db_2(\bar{u})}{du}. \end{cases} \quad (32)$$

## 4 Scale-Free Network

Having considered scale-free network as a tool to structure the population of taxpayers and an engine for effective information spreading we estimated the number of contacts as the average connectivity of the network  $\langle k \rangle$  and suppose that each node has approximately the same number of connections. Usually scale-free network is defined as a random graph whose degree distribution follows a power law and the main characteristic of the network does not depend on its size [7, 14]. The probability that a node of these networks has  $k$  connections follows a scale-free distribution  $P(k) \sim k^{-\gamma}$  with an exponent  $\gamma$  that ranges between 2 and 3.

We studied the SEIR model of spreading information over a scale-free network (SF network), taking into account the impact of scale-free connectivity into the

process of propagations. Based on the algorithm which had been studied in [14] in the present paper we introduce a modification of the algorithm of constructing a SF network. We develop a software product that allows for the observation of the process of dissemination of information and tracking changes in the networks settings. Below we show the key point of the algorithm:

- Initially (in the moment  $t_0$ ), the number of unrelated nodes  $m_0$  is small.
- At any time  $t_i = t_{i-1} + 1$  we add a new node with  $m$  links that point to an existing node  $i$  with  $k_i$  links according to the probability

$$P(i) = \frac{k_i}{\sum_j k_j}. \quad (33)$$

Here  $m_0$  and  $m$  are the parameters, defined by the user, which have some restrictions. The parameter  $m$  characterizes the average number of connections of a single individual  $\langle k \rangle = 2m$ . Suppose that the parameter  $m_0$  must be not more than  $m$  due to the following considerations: if  $m_0 > m$  and the next node is added then, according to (33), the nodes with a zero probabilities of further connection remain in this network. Then the constructed graph will be disconnected and contains nodes with no neighbors. These nodes represent individuals who do not have any contacts in the population and, therefore, are not involved in the epidemic process. These individuals can then be eliminated. After iterating this process we obtain a network with  $N$  nodes with connectivity distribution  $P(k) \sim k^{-\gamma}$ .

The detailed process of forming the SF network can be divided into two main stages:

- Step 1. We build  $m_0$  disconnected nodes. Then, while the number of available nodes is not more than  $m$ , we add node by node, which immediately get communication with others. This approach helps to avoid the unacceptable situation when two nodes are connected by two or more links.
- Step 2. When there are more than  $m$  nodes in the network we can use Eq. (33). Before the size of the network reaches  $N$ , the nodes are added one by one. New node gets the link with one of old nodes according to the calculated probability. At each iteration the denominator in Eq. (33) changes. Nodes that have already connected to the added one are not involved in the process anymore. When the size of the network reaches  $N$ , the algorithm is stopped.

Using a scale-free network to assign the connections between agents in the populations of taxpayers we consider a process of propagating information which resembles an epidemic, then we must define the parameters of the proposed SEIR model:  $\delta_I, \sigma_I$  are the transition coefficients; the initial distribution of the states of nodes;  $\bar{p}$  is the audit probability and  $\bar{c}$  is the audit cost;  $\xi$  and  $\pi$  are the tax and penalty rates correspondingly; distribution of the population by income level.

As has been shown in the previous research, an epidemic process can take place in different ways on a network, even when given the same parameters. This occurs because the initial distribution of the states of nodes introduces an element of

chance—the user defines the number of nodes in a particular state, but does not choose exactly which nodes belong to each group. This also changes the initial value of  $\Theta_I$ , since in one case we can obtain a hub as an Informed, which will increase  $\Theta_I$ , and in the other case we can obtain a node with a small number of connections.

After all the parameters of the epidemic process are defined, the user can run it step by step. At each step, the following actions occur:

- For each Uninformed  $S^k$  node which has a connection with an Informed “neighbor”  $I^k$ , we check if he receives and adopts information in accordance with the specified transition coefficients  $\delta_I$ . If the transmission of information is successful, the Uninformed node changes its status from  $S^k$  to  $I^k$ .
- If a node changed its status from  $S^k \rightarrow I^k$ , then it is necessary to determine if the received information is important to the agent. It means that information is important for agent with probability  $(1 - \alpha)$  and it is indifferent with the probability  $\alpha$ . If the node becomes indifferent, then it belongs to  $E^k$ .
- Each Informed  $I^k$  node loses interest in information in accordance with the transition coefficient  $\sigma_I$  and becomes resistant to information ( $R^k$ ).
- after all transitions we recalculate values  $\Theta_I$  and draw the new network.

Below we depict an example of process of propagation of information on small population of taxpayers. In Figs. 2–4 blue dots correspond to Uninformed taxpayers, red dots—Informed, orange dots—Indifferent, and green dots—Resistant.

## 5 Numerical Simulations

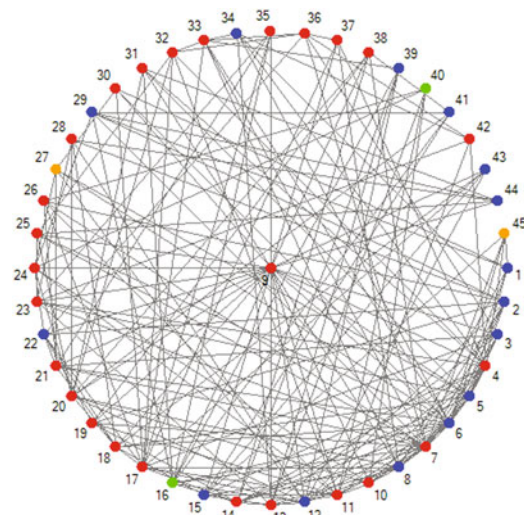
In this section, we present numerical simulations which are used to corroborate the results of the main propositions. We study the model of spreading information with the following parameters: tax and penalty rates are  $\xi = 0.13$  and  $\pi = 0.13$  correspondingly; the value of optimal probability is  $p^* = 0.5$  (according to the fixed values of  $\xi$  and  $\pi$ ); the value of the actual auditing probability is  $\bar{p} = 0.2$ .

We use the distribution of income among the population of Russian Federation in April of 2014 (see [2]) and calculate the average income level as the expected value of the uniform and Pareto distributions [9] (as it was previously done in [12]) to illustrate the simulation results.

We estimate an average monthly income of taxpayers as  $\hat{i} = 30,000$  (rub) (see Table 1). According to the statistical data, costs of audit and information announcements approximately are equal to  $\bar{c} = 7455$  (rub) and  $\tilde{c} = 200$  (rub), respectively. We assume that the duration of time period which is valued to propagate information is  $T = 0.5$  (130 days). In our paper we consider population of size  $N = 1000$  and initial fractions of Uninformed, Informed, Indifferent, and Resistant are  $S^k(0) = 0.9$ ,  $E^k(0) = 0$ ,  $I^k(0) = 0.1$ ,  $R^k(0) = 0$ . We use as a model parameters  $\sigma = \frac{1}{\bar{T}} = 0.0083$ , where  $\bar{T} = 120$  is a period of obsolescence of information, and  $\alpha = 0.1$ . We construct a scale-free network (for  $N = 1000$ ), according to the algorithms, which are presented in [14, 16], using the

**Table 1** The distribution of income among the taxpayers

No	Average income level (rub)	Share of taxpayers
1	1750	0.038
2	4250	0.056
3	6000	0.094
4	8500	0.146
5	12,500	0.202
6	20,000	0.235
7	30,000	0.108
8	70,000 more	0.121



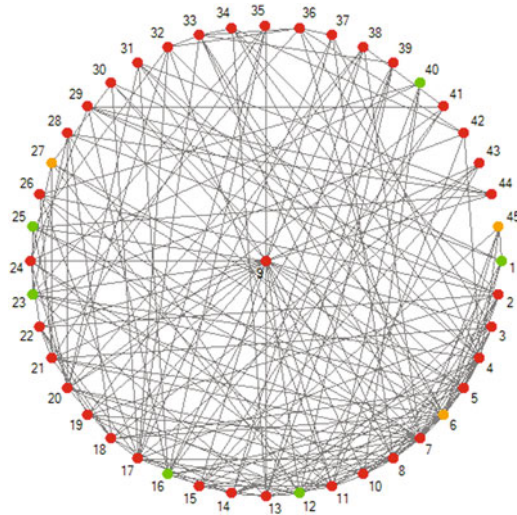
**Fig. 2** An example of the network at different moments  $t$ .  $t = 8$  seconds,  $N = 45$ ,  $S^k(t) = 15$ ,  $E^k(t) = 2$ ,  $I^k(t) = 25$ ,  $R^k(t) = 2$ ,  $\delta_I = 0.4$ ,  $\alpha = 0.1$ ,  $\sigma_I = 0.05$ ,  $\bar{p} = 0.2$

next parameters:  $\langle k \rangle = 6$ ,  $P(k) = \frac{2m^2}{k-3}$ ,  $m = 5$ ,  $m_0 = 4$ ,  $\Theta_I \sim 0.33$ . Examples of the networks are presented in Figs. 2, 3, and 4.

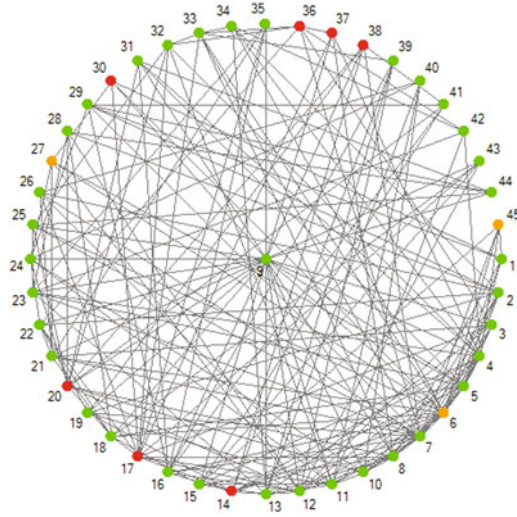
In Figs. 5, 6, and 7 we estimate an impact of transitions rate  $\alpha$  on SEIR system and compare fractions of  $I^k$  in five different cases. A higher value of  $\alpha$  suggests that the fraction  $I^k$  is less and application of control decreases a number of  $I^k$  for the same values of  $\alpha$ . This fact shows that spreading information guarantees the increasing of taxpayers in group  $R^k$  who will pay taxes.

Figures 8 and 9 show the aggregated system profit and demonstrate the influence of parameter  $\delta$  to collected taxes. We observe that total system costs which consist of  $J_{inf}$  and  $J_{aud}$  persistently grow depending on increasing of  $\alpha$  at large value of  $\delta$  (Figs. 10 and 11).

By studying the impact of various parameters to the population of taxpayers where tax authority propagates information about future tax audit we can draw a conclusion that the amount of collected taxes increases if a number of Informed

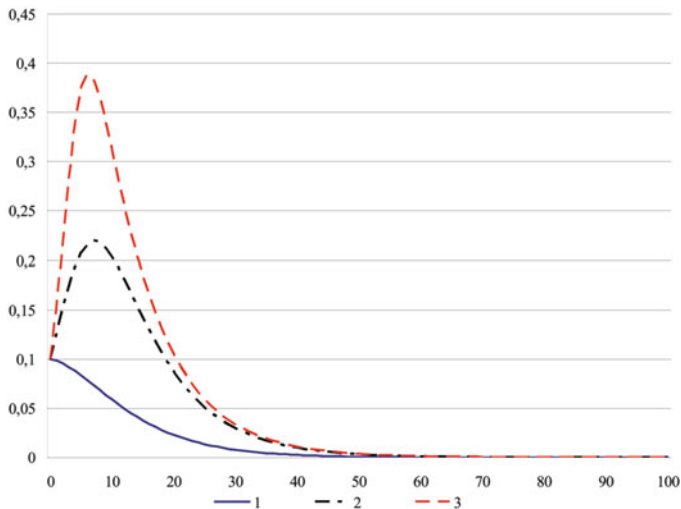


**Fig. 3** An example of the network at different moments  $t$ .  $t = 13$  seconds,  $N = 45$ ,  $S^k(t) = 0$ ,  $E^k(t) = 3$ ,  $I^k(t) = 33$ ,  $R^k(t) = 9$ ,  $\delta_I = 0.4$ ,  $\alpha = 0.1$ ,  $\sigma_I = 0.05$ ,  $\bar{p} = 0.2$

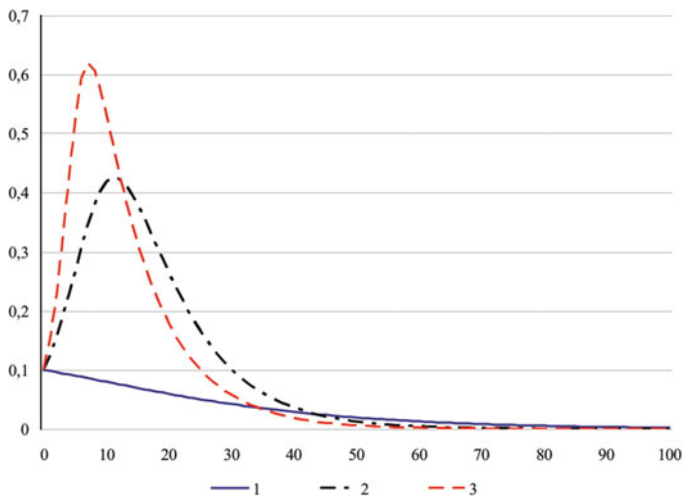


**Fig. 4** An example of the network at different moments  $t$ .  $T = 42$  seconds,  $N = 45$ ,  $S^k(T) = 0$ ,  $E^k(T) = 3$ ,  $I^k(T) = 7$ ,  $R^k(T) = 35$ ,  $\delta_I = 0.4$ ,  $\alpha = 0.1$ ,  $\sigma_I = 0.05$ ,  $\bar{p} = 0.2$ . Results:  $S^k(T) = 0$ ,  $I^k(T) = 0$ ,  $E^k(T) = 3$ ,  $R^k(T) = 42$ , aggregated system profit is  $J = 109,980$  monetary units

taxpayers and spreaders grow. At the same time even if the probability  $\alpha$  is high, the spreading information provokes the augmentation of tax collection with minimum costs. Therefore we are able to say that this approach can be considered as effective and reasonable method to improve taxation system.

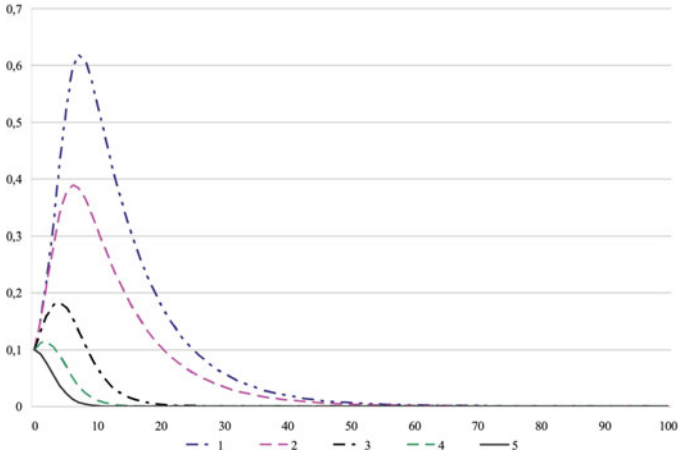


**Fig. 5** The fraction of Informed taxpayers. Controlled case. Initial states:  $S^k(0) = 0.9, E^k(0) = 0, I^k(0) = 0.1, R^k(0) = 0, \sigma_I = 0.0083, \alpha = 0.1.$  (1)  $\delta_I = 0.1, I_{max}^k = 0.1, t_{max} = 0.$  (2)  $\delta_I = 0.4, I_{max}^k = 0.21985, t_{max} = 7.$  (3)  $\delta_I = 0.7, I_{max}^k = 0.38927, t_{max} = 6$

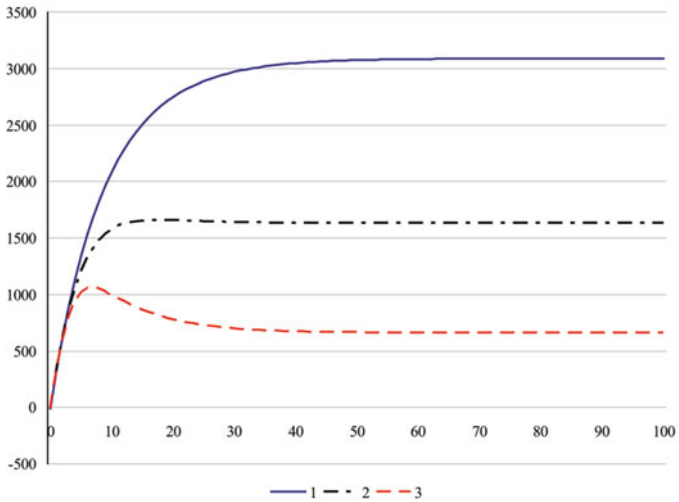


**Fig. 6** The fraction of Informed taxpayers. Uncontrolled case. Initial states:  $S^k(0) = 0.9, E^k(0) = 0, I^k(0) = 0.1, R^k(0) = 0, \sigma_I = 0.0083, \alpha = 0.1.$  (1)  $\delta_I = 0.1, I_{max}^k = 0.1, t_{max} = 0.$  (2)  $\delta_I = 0.4, I_{max}^k = 0.425751, t_{max} = 11.$  (3)  $\delta_I = 0.7, I_{max}^k = 0.619338, t_{max} = 7$

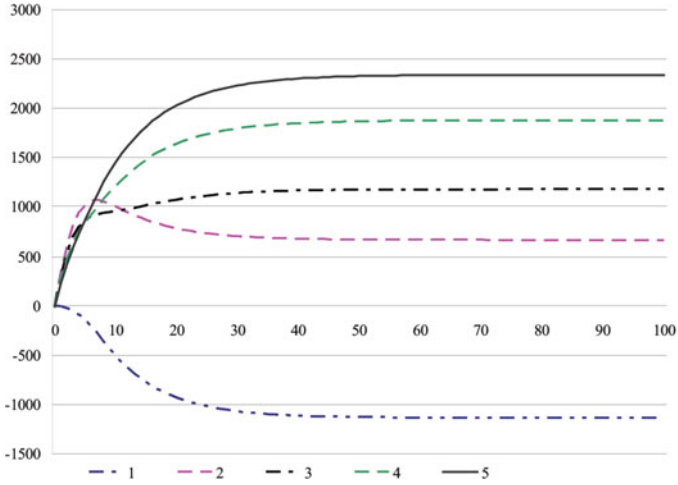




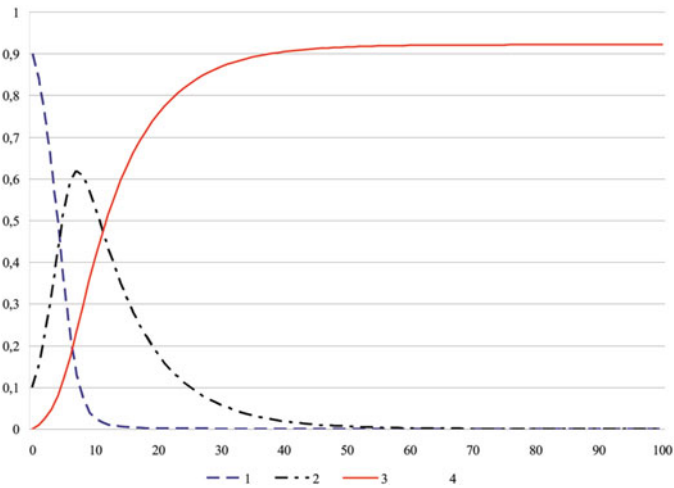
**Fig. 7** Fractions of  $I^k$  in SEIR model, controlled and uncontrolled case.  $S^k(0) = 0.9, E^k(0) = 0, I^k(0) = 0.1, R^k(0) = 0, \sigma_I = 0.0083, \delta_I = 0.7$ . (1) Uncontrolled system.  $\alpha = 0.1, I^k_{max} = 0.6193, t_{max} = 7$ , (2) Controlled system.  $\alpha = 0.1, I^k_{max} = 0.3892, t_{max} = 6$ , (3) Controlled system.  $\alpha = 0.3, I^k_{max} = 0.1831, t_{max} = 4$ , (4) Controlled system.  $\alpha = 0.5, I^k_{max} = 0.1138, t_{max} = 2$ , (5) Controlled system.  $\alpha = 0.3, I^k_{max} = 0.1, t_{max} = 0$



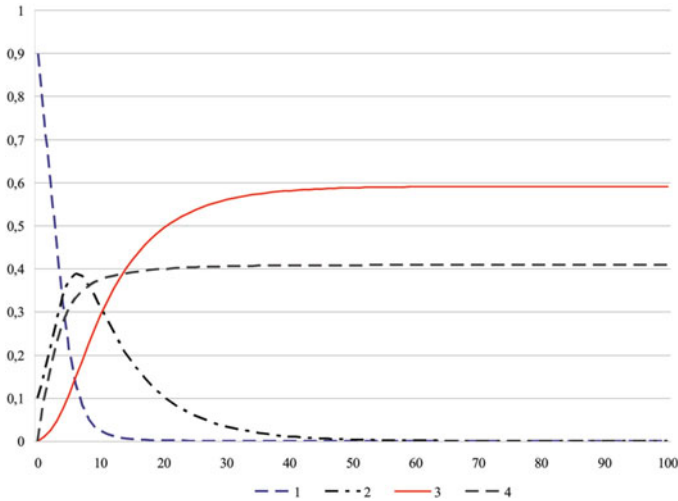
**Fig. 8** Aggregated system profit tax authority throws information into the taxpayers population.  $S^k(0) = 0.9, E^k(0) = 0, I^k(0) = 0.1, R^k(0) = 0, \sigma_I = 0.0083, \alpha = 0.7$ . (1)  $\delta_I = 0.1, J_{inf} = 282,717, J = 282,945$  monetary units. (2)  $\delta_I = 0.4, J_{inf} = 158,526, J = 159,171$  monetary units. (3)  $\delta_I = 0.7, J_{inf} = 71,178, J = 72,101$  monetary units



**Fig. 9** Aggregated system profit tax authority throws information into the taxpayers population.  $S^k(0) = 0.9, E^k(0) = 0, I^k(0) = 0, R^k(0) = 0, \sigma_I = 0.0083, \delta_I = 0.7$ . (1) Uncontrolled case.  $\alpha = 0.1, J = -98,018$  monetary units. (2) Controlled case.  $\alpha = 0.1, J = 701,013$ . (3) Controlled case.  $\alpha = 0.3, J = 111,655$  monetary units. (4) Controlled case.  $\alpha = 0.5, J = 171,343$  monetary units. (5) Controlled case.  $\alpha = 0.7, J = 211,925$  monetary units



**Fig. 10** SEIR model without application of control.  $S^k(0) = 0.9, E^k(0) = 0, I^k(0) = 0.1, R^k(0) = 0, \sigma_I = 0.0083, \delta_I = 0.7, I_{max}^k = 0.6193, t_{max} = 7$



**Fig. 11** SEIR model without application of control.  $S^k(0) = 0.9$ ,  $E^k(0) = 0$ ,  $I^k(0) = 0.1$ ,  $R^k(0) = 0$ ,  $\sigma_I = 0.0083$ ,  $\delta_I = 0.7$ .  $I_{max}^k = 0.3892$ ,  $t_{max} = 4$

## 6 Conclusion

In the present paper we have investigated a complex model in which we combine a game-theoretical approach of tax control with a dynamic model of information propagation over a structured population of taxpayers. We formulated an optimal control problem for a tax auditing policy and analyzed the behavior of agents depending on social contacts and specific cost functions. All theoretical results are supported by numerical simulations with the real statistical data. Connections between taxpayers are modeled as a scale-free network constructed by a specially developed algorithm. As a result of our research we attempted to provide a new method of tax collection which can be more efficient and cost-effective.

## References

1. Bure, V.M., Kumacheva, S.Sh.: A game theory model of tax auditing using statistical information about taxpayers. Vestnik SPbGU, series 10, vol. 4, pp. 16–242 (2010) (in Russian)
2. Federal State Statistics Service: The web-site of the Russian Federation State Statistics Service (2015). Available: <http://www.gks.ru/>
3. Goffman, W., Newill, V.A.: Generalization of epidemic theory: an application to the transmission of ideas. Nature **204**(4955), 225–228 (1964)
4. Gubar, E.A., Zhu, Q.: Optimal control of influenza epidemic model with virus mutations. In: 12th Biannual European Control Conference (ECC 13), pp. 3125–3130. IEEE Control Systems Society, Zurich (2013)

5. Gubar, E.A., Zhitkova, E.M., Kupchinenko, E.M., Petriakova, N.S.: Two modes of vaccination program in controlled SIR model. In: L.A. Petrosyan, N.A. Zenkevich (eds.) Contributions of the 8th International Conference GTM 2012, pp. 84–98 (2015)
6. Gubar, E.A., Kumacheva, S.Sh., Zhitkova, E.M., Porokhnyavaya, O.Yu.: Propagation of information over the network of taxpayers in the model of tax auditing. In: Stability and Control Processes in Memory of V.I. Zubov SCP 2015, IEEE Conference Publications, pp. 244–247 (2015). INSPEC Accession Number: 15637330
7. Fu, X., Small, M., Walker, D.M., Zhang, H.: Epidemic dynamics on scale-free networks with piecewise linear infectivity and immunization. *Phys. Rev. E* **77**, 036113 (2008)
8. Kandhway, K., Kuri, J.: Optimal control of information epidemics modeled as Maki Thompson rumors. *Commun. Nonlinear Sci. Numer. Simul.* **19**(12) (2014)
9. Kendall, M.G., Stuart, A.: *Distribution Theory*. Nauka, Moscow (1966) (in Russian)
10. Kermack, W.O., Mc Kendrick, A.G.: A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Ser. A* **115**(A771), 700–721 (1927)
11. Khouzani, M.H.R., Sarkar, S., Altman, E.: Optimal control of epidemic evolution. In: Proceedings of IEEE INFOCOM (2011)
12. Kumacheva, S.Sh., Gubar, E.A.: Evolutionary model of tax auditing. In: Contributions to Game Theory and Management, vol. 8, pp. 164–175 (2015)
13. Nekovee, A.M., Moreno, Y., Bianconi, G., Marsili, M.: Theory of rumor spreading in complex social networks. *Physica A* **374**, 457–470 (2007)
14. Pastor-Satorras, R., Vespignani, A.: Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**(14), 3200–3203 (2001)
15. Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., Mishchenko, E.F.: *The Mathematical Theory of Optimal Processes*. Interscience, New York (1962)
16. Porokhnyavaya, O.Yu.: One algorithm for constructing a scale-free networks. In: Control Processes and Stability, vol. 2, issue 18, pp. 702–707 (2015)
17. Reinganum, J.R., Wilde, L.L.: Income tax compliance in a principal–agent framework. *J. Publ. Econ.* **26**, 1–18 (1985)
18. Sanchez, I., Sobel, J.: Hierarchical design and enforcement of income tax policies. *J. Publ. Econ.* **50**, 345–369 (1993)
19. Vasin, A.A., Vasina, P.A.: The optimization of tax system in condition of tax evasion: the role of penalty restriction. M., EERC (2002) (in Russian)