# Endogenous Shocks in Social Networks: Effects of Students' Exam Retakes on their Friends' Future Performance

**(Preliminary version)**

Maria Marchenko *

March 17, 2016

**Abstract**

Exam retakes, as other types of shocks or the treatment happening to one of the players in the network influence not only their future performance but affect all their network connections. Thus, it is crucial to understand the behavior of the whole network in response to the retake. It is, however, highly endogenous. The logic used in peer effect literature is adopted to develop the dynamic model accounting for the endogeneity of the shock. The model allows predicting the endogenous part of the friends' retake and use the unexpected component to estimate the effect of pure shock on the changes in one's average grade. The identification conditions for the effect are derived and the consistent estimation procedure is proposed. It is applied to the dynamic network data on the students in HSE, Nizhniy Novgorod. The results suggest that on average the retake of the friend may have a negative effect on future performance, however, this effect has a different magnitude for students with and without own retake, as well as for students of different departments.

* JEL classification codes: C21, C49, I21
Please check the latest version at sites.google.com/site/mariavmarchenko/jmp.pdf
Center for Doctoral Studies in Economics, Graduate School of Social Sciences, University of Mannheim, Germany and National Research University Higher School of Economics, Moscow, Russia. Email: maria.marchenko@gess.uni-mannheim.de.
This research has benefited from discussions with Andreas Dzemski, Maria Yudkevich, Alex Teytelboym, Andrea Weber, Enno Mammen, Jan Nimczik, Stephen Kastoryano, seminar participants in Mannheim and participants of RSSIA 2015.
The financial support from the Government of the Russian Federation within the framework of the Basic Research Program at the National Research University Higher School of Economics and within the framework of the implementation of the 5-100 Programme Roadmap of the National Research University Higher School of Economics is acknowledged.

# 1 Introduction

The peer effect, the effect that social connections have on people's behavior and achievements, plays an important role when analyzing educational outcomes. While there are numerous economics papers on peer effects across many fields, from education to juvenile behavior, the effect of shocking events on the friends network is rarely discussed. In particular, in the university framework, students' failures, such as retakes of examinations or dropouts are usually only discussed for the students' results in the same year and not in relation to their friends' future behaviour. However, the shock of a friend's failure influences the future behaviour and outcomes, especially when this failure was not anticipated.

This project contributes to the literature by covering an existing gap in peer effects literature and studying the changes of peers' behaviour and achievement in response to the individual shock. In contrast to some examples in development literature (e.g. Comola and Prina (2014)) considering exogenous individual treatment, I propose the model, allowing the shock to be endogenously formed. Two components of the shock can be disentangled: predicted probability of the shock and unexpected component. The latter is considered to be crucial to the changes of future behavior. I am considering the students' exam failure as the source of the shock and test the model on the sample of students of one cohort at the National Research University - Higher School of Economics, a highly selective university in Russia. The threat of retakes and dropouts may put a lot of pressure on students, and the higher probability of failure may result in lower productivity. Knowing, how these shocks influence the behavior of the students and their friends, can help to understand the whole dynamics of network performance, and maybe help universities to adjust the strategy of setting up the retakes' threshold. Of course, dropouts are likely to influence future behaviour stronger than retakes, since the latter can still be fixed. However, the existing data of the dropouts is not sufficient for proper econometric analysis. I discuss both sources of shock in descriptive analysis but apply the econometric model only to retakes.

The direction of the effect, however, can be twofold. While the unexpected shock may serve as a wake-up call and motivate students to be more dedicated to their studies, the connections can be extremely tight. This can reduce the amount of time spent on one's own studies due to the shared activities with the friend either outside of the university, if the friend left, or helping the friend to prepare for the retake of the exam. The reasons of the retakes during the studies can be different. In the first year, students are more likely to fail due to the lack of the abilities or difficulties with adjustments to the new environment. The fist exams may appear to be too difficult for some of the students, even though they had sufficient abilities to enter the university. Students with lower abilities are either dropping out of the university or adjusting their efforts to improve performance. In the second and higher year, students are more likely to fail due to insufficient efforts. Therefore, the shock during the different time periods may have a different effect on the future performance. This paper discusses only the first year retakes at the moment.

Although I do not study the pure peer effect in this paper, I exploit the general idea of peer effects literature and its methodological fundamentals. Most of the economic literature that analyses peer effects use the framework and the model introduced by Manski (1993). He distinguishes three effects that determine the similar behaviour of peers. *The endogenous effect* explains that the probability of a particular student to drop out of the school or university or to fail an exam will be affected by a number of this student's peers who have already done so. *The exogenous effect* uses mean exogenous characteristics of the peer group, such as parental education, socio-economic status (SES), etc., to determine the probability of the dropout or retake. *The correlated effect* appears due to the similar individual characteristics within a group. The most important task of peer effects analysis is to determine the endogenous effect, which can have important policy implications.

Identification of these three effects in the case of group interactions requires an additional source of exogenous variation, such as exogenous class formation (for example, Carrell et al., 2009 in military institutions framework and De Giorgi et al. (2010) and Androushchak et al. (2013) in university frameworks with randomly assigned groups) or random assignment of dormmates (for example, Sacerdote, 2011). Estimating the endogenous peer effect as an effect of an average group performance obtained some critique, and additional assumptions on the structure or the ranking inside the peer group or even exact links are preferable, but social network data is not always available. Usage of social network data requires other identifying assumptions, which restrict the network. Bramoullé et al. (2009) proved the identification of the peer effect in social networks under rather mild assumptions. Poldin et al. (2015) use the same identification result to study the peer effect in the university framework using HSE dataset.

The identification of the direct effect of shock on the friends' future outcome is, however, more challenging, since the changes of the performance are not driven solely by the effects of the shocks. Exogenous and unobserved characteristics of the student and his peers as well as the changes in the network structure are among the other determinants. Moreover, as was already mentioned, the shock itself is not exogenous, and its significant part is driven by the model itself. The paper proposes an econometric model which deals with both problems and estimates the effect of the shock: a two-step dynamic peer effects model. The first step estimates the probability of the shock adopting the instrumental variable 2SLS approach discussed by Bramoullé et al. (2009) after Lee (2003). The second step uses the residuals from the first stage to estimate the effect of the unexpected component of the changes in students' performance.

To the best of my knowledge, this project is the first to introduce the dynamic peer effect in social networks model with endogenous shock[1]. Moreover, I provide the identification results for this model and propose estimation procedure. The identification and estimation of the first step are the straightforward adjustments of the Bramoullé et al. (2009) approach, and requires the existence of intransitive triads in the network given the assumption of no correlated effects, i.e. friends of some student's friends not connected to him or her. Hence, the friends of friend affect the student not directly, but via the common friend only. If the assumption of no correlated effects is relaxed, the stricter identifying assumption is necessary. The whole network should include pairs of students with the distance between them of length three or bigger. They are not connected directly, and the shortest path from the one to the other has not less than three links. Friends of friends are used to deal with the correlated effect, therefore, the next level of friends is used as an identifying assumption. The identification of the second step is novel and demonstrates the necessity of the network longitudinal variation. Changes of the network allow comparing the influence of "old" and "new" peer group on the outcome. The presence of the new friends and absence of old ones creates variation in the peer group characteristics and this helps to identify social effects and the effect of the shock. However, it is important that the changes of the network are not driven solely by the shock. Moreover, at the moment, I do not model link formation, and therefore, do not distinguish between different types of network changes and treat them all as equal and given.

The variation of the network is a valid assumption for the students' network setting. The links formed in the first year are highly likely to be revised due to the gradual unveiling of the friends' personal characteristics. Some of the links might be broken, however, due to the exam retakes and dropouts of the friends. The student may seek for a more advantageous peer group or he/she no longer spends much time with the friend preparing for the retakes. But even if the friend fails an exam and the link stays stable in the network, the student may tend to connect to the students with higher results, creating new links. The exam retake is endogenous in the model, and only an unexpected

[1]See, for example, a review of the recent econometric literature on networks in Paula (2015)

component of the retake probability is considered as a shock. The influence of this unexpected component on link formation is not the same as possible channels of influence of retakes on link formation, discussed previously, therefore, the actual importance of the shock for link changes might be lower than the one of exam retakes. The model in the paper is discussed without link formation process and, therefore, under the assumption that changes in the network are exogenously given. This setup is a bit restrictive, and relaxation of this assumption will be considered for future research.

The magnitude of the endogenous effects in different periods is considered to be different, since the unexpected shock may affect performance via the changes of the peer groups, and not only directly. The break of the link itself makes the peer group "better", then the improvement of the results can also be caused by the group's refinement.

Dropouts and retakes are important to study from the university's perspective. Dropouts create the sunk costs for the university. For example, costs of the university dropouts in Germany were estimated at the level of $11.5 billion in 2007[2] and in Australia at $1.36 billion[3]. Some of the dropouts are the results of the policies of the university, which can be controlled. In some institutions of higher education, as in the sample used in the analysis, most of the dropouts are directly affected by the retakes. In HSE 3 retakes during the same exam session term will lead to the expulsion of the student. Therefore, understanding the possible mechanisms of retakes' influence on future performance may suggest possible university-level policy improvements in order to reduce sunk costs.

The paper is organized as follows. Section 2 discusses the proposed model, states the identifying assumptions, and proposes the estimation method. Section 3 describes the data used and the institutional environment of the educational system in Russia, as well as results of the descriptive analysis. Section 4 provides the estimation results and evidence of the influence of dropouts and retakes on peers. Section 5 concludes.

# 2 Model

## 2.1 Naïve approach

I propose a two-step model that allows estimating the effect of an unexpected event happening to network connections. Although I do not conduct the pure peer effect estimation, I use the classical peer effect model as a baseline.

A naïve way to write down the dynamic peer effect model without modelling the link

[2] The figures are obtained by the Stifterverband, association of German science and higher education donors. Details can be found on UWN website

[3] According to the report on UWN website

formation:

$$y_i^1 = \alpha_1 + \beta_1 \sum_{j \neq i} G_{ij}^1 y_j^1 + \gamma_1 X_i^1 + \delta_1 \sum_{j \neq i} G_{ij}^1 X_j^1 + \xi_i + \epsilon_i^1, \qquad \mathbb{E}[\epsilon_i^1 | X^1] = 0, \tag{1}$$

$$y_i^2 = \alpha_2 + \beta_2 \sum_{j \neq i} G_{ij}^2 y_j^2 + \gamma_2 X_i^2 + \delta_2 \sum_{j \neq i} G_{ij}^2 X_j^2 + \xi_i + \epsilon_i^2, \qquad \mathbb{E}[\epsilon_i^2 | X^2] = 0, \tag{2}$$

where $y_i^1$ and $y_i^2$ are outcome variables of student $i$ in the first wave and the second wave correspondingly. I will consider the average grade in the main specification of the model. Student's rating or grades for some specific subjects, which last more than 1 term, are used for robustness checks;
$X_i$ is a vector of individual characteristics that should be controlled for, such as gender, city of origin, living conditions, some socioeconomic family characteristics. In the discussed empirical example it also includes the results of the high school examination, universal and obligatory for all the students graduating the high school.
$G_{ij}^1$ and $G_{ij}^2$ are two adjacency matrices for the first and the second waves correspondingly, weighted by the number of links, and their entries have the value of $1/n_i$ if the link from student $i$ to student $j$ exists. Note that this matrices are not necessarily symmetrical, since the social network can be both directed (as in the sample used later) or undirected.
$\xi_i$ - student-level unobserved fixed characteristics, which may influence students' performance and choice of connections.

Those unobserved individual characteristics also reflect the homophily of the individuals, which may influence both link formation and the network outcomes. In the case of group interactions group fixed effects are often introduced to eliminate correlated effects, whereas in the case of interactions in big networks network fixed effects make little sense. Local differences, proposed by Bramoullé et al. (2009), may be used to address the issue of correlated effects. However, the dynamic structure of the data allows solving this issue differently. The dynamic peer model can be then written in terms of differences, and this will eliminate possible unobserved fixed effect component in the error term, consisting of the common for individual's connections unobservable component and individual's own unobserved fixed characteristics.

$$\Delta y_i = \Delta\alpha + \beta_2 \sum_{j \neq i} G_{ij}^2 y_j^2 - \beta_1 \sum_{j \neq i} G_{ij}^1 y_j^1 + \gamma_2 X_i^2 - \gamma_1 X_i^1 + \delta_2 \sum_{j \neq i} G_{ij}^2 X_j^2 - \delta_1 \sum_{j \neq i} G_{ij}^1 X_j^1 + \Delta\epsilon_i$$

*Assumption A.* The outcome variable of a single period can be estimated using the one-period model.

This additional assumption allows avoiding the autoregressive component in the second-period model. Assumption A is valid, because the model, including observed and unobserved fixed effects characteristics as well as endogenous and exogenous peer effects, is sufficient to predict the educational achievements. Therefore, it can be claimed that there is no additional mechanism that can influence the outcome via the previous period's outcome.

The proposed model system (1) and (2), and consequently, the model written in differences, can be further modified in order to catch the desirable effect of shock. In the naïve way, similar to the model of Comola and Prina (2014), the model will now be as follows:

The equation for the first period should remain unchanged:

$$y_i^1 = \alpha_1 + \beta_1 \sum_{j \neq i} G_{ij}^1 y_j^1 + \gamma_1 X_i^1 + \delta_1 \sum_{j \neq i} G_{ij}^1 X_j^1 + \xi_i + \epsilon_i^1,$$

Whereas, the second-period model shall take into account the shock of unexpected retake of the friend. The straightforward way to do it is just to include the binary variable in the vector of controls:

$$y_i^2 = \alpha_2 + \beta_2 \sum_{j \neq i} G_{ij}^2 y_j^2 + \tilde{\delta} D_i + \gamma_2 X_i^2 + \delta_2 \sum_{j \neq i} G_{ij}^2 X_j^2 + \xi_i + \epsilon_i^2$$

where $D_i$ is a dummy for having any friends with a retake in the first period[4].
The system can then be re-written in differences, eliminating the possible individual fixed effect:

$$\Delta y_i = (\alpha_2 - \alpha_1) + \beta_2 \sum_{j \neq i} G_{ij}^2 y_j^2 - \beta_1 \sum_{j \neq i} G_{ij}^1 y_j^1 + \tilde{\gamma} D_i +$$
$$+ \gamma_2 X_i^2 - \gamma_1 X_i^1 + \delta_2 \sum_{j \neq i} G_{ij}^2 X_j^2 - \delta_1 \sum_{j \neq i} G_{ij}^1 X_j^1 + \epsilon_i^2 - \epsilon_i^1$$

However, this type of the equation is only valid if the shock is exogenous, as in the examples of randomized treatment. A big share of the probability of the student's retake can be explained by the observed component of the model, and therefore, the retake itself cannot be considered as unexpected shock. I propose to use the peer effect model of the first period to disentangle predictable and unexpected parts of the probability of

[4]In general the coefficients in the model with the shock are different from the baseline one-period models (1) and (2), but I left the same notations for simplicity

the retake, and use the unpredicted part only to estimate the effect of the shock on the performance.

Comola and Prina (2014) also model the changes of the network as a response to the exogenous treatment. At the moment, I am not modelling the link formation. The variation of the network links is assumed and is a crucial identifying assumption. Importantly, a significant part of the changes in the structure of the friendship networks is caused by the individual characteristics and outcome and not solely by the exam retake. The influence of the retake and of the unpredicted component of the retake on the link formation also should be treated and interpreted differently, since the probability of the exam retake is endogenous. The following assumption, therefore, should be made. *Assumption B.* Changes of the network as a response to unexpected shock are neglected, and all changes of the network itself are treated as exogenous.
This assumption can potentially cause overestimation of the direct effect of the shock, and therefore, should be relaxed in the future research.

## 2.2 Proposed model with no correlated effects

### 2.2.1 The model

Taking into account all above-mentioned argument, I estimate the following model at the first step:

$$P(retake_i) = \alpha + \beta \sum_{j \neq i} G^1_{ij} y^1_j + \gamma X^1_i + \delta \sum_{j \neq i} G^1_{ij} X^1_j + \xi_i + \nu_i, \qquad \mathbb{E}[\nu_i | X^1] = 0 \tag{3}$$

In this specification, the error term consists of two parts: unobserved correlated effect, and conditionally independent noise. Dynamic peer effect model will eliminate the correlated effect component at the second step of the model, leading to the conditional independence of the error term. However, on the first step in general $\mathbb{E}[\xi_i + \nu_i | X^1] \neq 0$. I will discuss two cases: assuming no correlated effects and with correlated effect. The latter will be considered in the later subsections. For the former, (3) will be transformed as follows :

$$P(retake_i) = \alpha + \beta \sum_{j \neq i} G^1_{ij} y^1_j + \gamma X^1_i + \delta \sum_{j \neq i} G^1_{ij} X^1_j + \nu_i, \qquad \mathbb{E}[\nu_i | X^1] = 0 \tag{3a}$$

I then take the residuals of the equation (3a), which is the part of the probability of the friends' retake not predicted by the model. I then construct the shock for student $i$ as the combination of the residuals for the students in the network of $i$. The baseline specification uses the average of the residuals: $UR_i = \sum_{j \neq i} G^1_{ij} \hat{\nu}_j$. However, the other approaches to define $UR_i$ is possible: maximum of friends' unpredicted probability of the exam retakes, residuals for the friends named first, or average weighted according to the order, with which friends are appearing in the answers of the students. The identification results and estimation procedure are not affected by the choice of the approach to defining $UR_i$. Then I am using it as an unexpected shock to plug-in in the following equation:

$$\Delta y_i = (\alpha_2 - \alpha_1) + \beta_2 \sum_{j \neq i} G^2_{ij} y^2_j - \beta_1 \sum_{j \neq i} G^1_{ij} y^1_j + \tilde{\boldsymbol{\delta}} UR_i + \gamma_2 X^2_i - \gamma_1 X^1_i +$$

$$+ \delta_2 \sum_{j \neq i} G^2_{ij} X^2_j - \delta_1 \sum_{j \neq i} G^1_{ij} X^1_j + \Delta \epsilon_i \qquad (4)$$

Since the model in differences eliminates possible individual fixed effect component in error term, I am able to make a stricter assumption on the error term: $\mathbb{E}[\Delta \epsilon_i] = 0$, instead of the conditional expectation. This condition will be used to prove the model identification.

Model in differences, additional to the elimination of individual fixed effect, gives a better interpretation of the studied effect. It estimates the changes of own performance in response to the shock additional to the changes of performance in comparison to the classmates, obtained by the single-period model.

Note that the coefficients for the endogenous peer effect and exogenous characteristics are considered to be different in two periods: $\beta_2$ and $\beta_1$ and $\delta_2$ and $\delta_1$. Students may experience the different magnitude of the effects depending on how advanced they are in their studies, how well they are adjusted to the university environment, etc. Moreover, this also allows to take into account the changes in the network, since the students are experiencing the influence of two different peer groups in two periods.

The own retake of the student is not included explicitly in the model. The unexpected component for the students themselves is close to zero since they can anticipate most of the retakes after writing the exam. Moreover, the outcome of the previous period partially takes care of own retakes. Nonetheless, in the empirical analysis, I will also split the sample and study the effect for those, who were retaking the exams, and for those, who were not, to tackle down possible differences.

### 2.2.2 Identifying assumptions

The identification results for the first step of the model adopt Bramoullé et al. (2009) approach, whereas the result, obtained for the second stage, is, to the best of my knowledge, a novel result for the literature.

**Lemma 1** *Let* $\gamma_1^2 + \delta_1^2 \neq 0$ *and* $\beta_1 \neq 0$[5]*. If matrices* $I$*,* $G^1$*,* $(G^1)^2$ *are linearly independent, coefficients in (3a) are identified.*

The proof of the Lemma is given in Appendix A. This is exactly the condition obtained by (Bramoullé et al., 2009), and can be proven similarly. The identification of the coefficients on the first step, hence, allow using the obtained residuals for the further analysis. The identification is ensured by the existence of intransitive triads in the network, i.e. the existence of a set of three individuals $i, j, k$ such that $i$ is influenced by $j$, $j$ is influenced by $k$, but $i$ is not influenced by $k$. This is a valid assumption for most networks, in particular, for the sample analysed in this paper, which will be discussed in the next section.

**Lemma 2** *In the case of no correlated effects, if the assumptions of* ***Lemma 1*** *hold, if* $\gamma_2^2 + \delta_2^2 \neq 0$ *and* $\beta_2 \neq 0$[6]*, if matrices* $I$*,* $G^2$*,* $(G^2)^2$ *are linearly independent, and if* $G^1 \neq G^2$*, with changes not driven by the shock only, coefficients in (4) are identified.*

Identification of Step 2 relies heavily on the variation in the network structure. However, it is important that some changes in the network are exogenous. This assumption is quite reasonable for the friendship networks. Students are likely to learn more about their classmates with time, and the friendships, created during the first year, are often unstable.

Once there are new links formed in the next period, the variation between new and old connections help to capture the effect of the changes in the average grade. For example, if a student $i$ is no longer connected to student $j$, and therefore, is not affected by student $j$, his performance can be evaluating in the two cases and the comparison of two results will result in the effect of not having friend $j$ , and hence, the social effects are easier to catch. The identifying assumptions also put the restriction on the friendship matrix of the second period, as in the first period: the network should include intransitive triads. The proof of Lemma 2 can also be found in Appendix A, and the validity of identifying assumptions will be discussed in the next Section.

[5]These are the coefficients from the baseline peer effect model (1).

[6]The coefficients from the baseline peer effect model (2)

## 2.3 Model with correlated effects

### 2.3.1 The model

As was already mentioned, the correlated effect appears due to the similar individual characteristics within a group. The correlated effect is unlikely to be present in big networks, however, once the network may suggest existence of smaller groups or subnetworks in it, the correlated effects are more likely to be present. In the empirical application discussed in this paper, most of the connections are formed inside of the same department, and even inside of the same exogenously formed study group. Therefore, the possible correlated effects could not be ignored and can cause an additional identification issue.

To deal with it and eliminate unobserved variables, I propose taking the local differences, i.e. averaging the equation (3) over the friends of $i$ and subtracting this average from (3) and noting that $\xi_i$ are the same for the students in one smaller network, and hence, it will vanish after taking the local differences:

$$P(retake_i) - \sum_{j\neq i} G^1_{ij} P(retake_j) = \beta \sum_{j\neq i} G^1_{ij}[y^1_j - \sum_{k\neq j} G^1_{jk} y^1_k] + \gamma[X^1_i - \sum_{j\neq i} G^1_{ij} X^1_j] +$$

$$+ \delta \sum_{j\neq i} G^1_{ij}[X^1_j - \sum_{j\neq k} G^1_{jk} X^1_k] + \eta_i, \qquad \eta_i = [\nu_i - \sum_{j\neq i} G^1_{ij}\nu_j],\ \mathbb{E}[\eta_i|X^1] = 0 \quad (5)$$

Similarly to the case without correlated effects, I construct the shock for the student $i$, taking the average of their networks residuals: $UR_i = \sum_{j\neq i} G^1_{ij}\hat{\eta}_j$. The second stage is then identical to the case with no correlated effects:

$$\Delta y_i = (\alpha_2 - \alpha_1) + \beta_2 \sum_{j\neq i} G^2_{ij} y^2_j - \beta_1 \sum_{j\neq i} G^1_{ij} y^1_j + \tilde{\boldsymbol{\delta}} UR_i + \gamma_2 X^2_i - \gamma_1 X^1_i +$$

$$+ \delta_2 \sum_{j\neq i} G^2_{ij} X^2_j - \delta_1 \sum_{j\neq i} G^1_{ij} X^1_j + \Delta\epsilon_i \quad (6)$$

Model in differences, additional to the elimination of individual fixed effect, also gets rid off the correlated effects, therefore, no local differences are needed for the second stage equation.

### 2.3.2 Identifying assumptions

The identification results for the first step of the model again adopt Bramoullé et al. (2009) approach, whereas the result, obtained for the second stage, is new.

**Lemma 3** *Let $\gamma_1^2 + \delta_1^2 \neq 0$ and $\beta^1 \neq 0$*[7]*. If matrices $I$, $G^1$, $(G^1)^2$, $(G^1)^3$ are linearly independent, coefficients in (5) are identified.*

The proof is given in Appendix A. This condition again follows the result of (Bramoullé et al., 2009) in the presence of correlated effects, and can be proven in the similar manner. The identification of model with correlated effects is ensured by the existence of distances between two students of length 3 and more, i.e. the existence of a set of at least 4 individuals $i, j, k, m$ such that $i$ is influenced by $j$, $j$ is influenced by $k$, $k$ is influenced by $m$, but $i$ is not influenced by both $m$ and $k$, and $j$ is not influenced by $m$. This is a bit more demanding assumption than in the case of no correlated effects, but still valid for a lot of networks' types, and in particular, for the sampled network, which will be discussed in the next section.

**Lemma 4** *In the case of correlated effects, if the assumptions of **Lemma 3** hold, if $\gamma_2^2 + \delta_2^2 \neq 0$ and $\beta_2 \neq 0$*[8]*, if matrices $I$, $G^2$, $(G^2)^2$, $(G^2)^3$ are linearly independent, and if $G^1 \neq G^2$, with changes not driven by the shock only, coefficients in (6) are identified.*

Identification of Step 2 again heavily relies on the variation in the network structure. Moreover, the restrictions are put on the friendship matrix of the second period, requiring the distances between two students of length 3 and more. The proof of Lemma 4 is presented in Appendix A.

## 2.4 Estimation strategy

### 2.4.1 No correlated effects

I first discuss the model that does not take into account correlation effects: (3) with no $\xi_i$ and (5).

**Step 1.** I partially repeat Bramoullé et al. (2009) for the first step and use the adaptation of Generalized 2SLS strategy proposed by Kelejian and Prucha (1998) and refined by Lee (2003). As the identification result suggests, $((\boldsymbol{G^1})^2\boldsymbol{X}, (\boldsymbol{G^1})^3\boldsymbol{X}, \dots)$ can be used as valid instruments to obtain consistent estimators.

First, recall the peer effect model in reduced form, written in matrix notations, offered in Bramoullé et al. (2009):

$$\boldsymbol{y^1} = \alpha_1 \boldsymbol{i} + \beta_1 \boldsymbol{G^1 y^1} + \gamma_1 \boldsymbol{X^1} + \delta_1 \boldsymbol{G^1 X^1} + \nu^1, \qquad \mathbb{E}[\nu|\boldsymbol{X^1}] = 0,$$

[7] The coefficients from the baseline peer effect model (1)

[8] The coefficients from the baseline peer effect model (2)

which gives

$$\mathbb{E}[\boldsymbol{G^1y^1}|\boldsymbol{X^1}] = (\boldsymbol{I} - \beta_1\boldsymbol{G^1})^{-1}\boldsymbol{G^1}\alpha_1 + (\boldsymbol{I} - \beta_1\boldsymbol{G^1})^{-1}\boldsymbol{G^1}(\gamma_1\boldsymbol{I} + \delta_1\boldsymbol{G^1})\boldsymbol{X^1}$$

Note that the fist step model can be written as follows:

$$\boldsymbol{PR} = \alpha + \beta\boldsymbol{G^1Y^1} + \gamma\boldsymbol{X^1} + \delta\boldsymbol{G^1X^1} + \nu, \qquad \mathbb{E}[\nu|\boldsymbol{X^1}] = 0 \tag{7}$$

I propose the following procedure that gives the consistent estimator of $\theta = (\alpha, \beta, \gamma, \delta)$:

**First,** compute the 2SLS estimator for $\theta^1 = (\alpha_1, \beta_1, \gamma_1, \delta_1)$ of the standard peer effects model, using the following vector of instruments $\boldsymbol{S} = [\boldsymbol{i}, \boldsymbol{X^1}, \boldsymbol{G^1X^1}, \boldsymbol{(G^1)^2X^1}]$, and with the vector of covariates $\boldsymbol{\tilde{X}^1} = [\boldsymbol{i}, \boldsymbol{X^1}, \boldsymbol{G^1X^1}, \boldsymbol{G^1y^1}]$.
$\hat{\theta}^1_{2SLS} = (\boldsymbol{\tilde{X}^1}^T\boldsymbol{P_S}\boldsymbol{\tilde{X}^1})^{-1}\boldsymbol{\tilde{X}^1}^T\boldsymbol{P_S}\boldsymbol{y^1}$, where $\boldsymbol{P_S} = \boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{S})^{-1}\boldsymbol{S}^T$ is a projection matrix.

**Second,** define $\boldsymbol{\hat{Z}} = Z(\hat{\theta}^1_{2SLS}) = [\boldsymbol{i}, \boldsymbol{X^1}, \boldsymbol{G^1X^1}, \mathbb{E}[\boldsymbol{G^1y^1}(\boldsymbol{\hat{\theta}^1_{2SLS}})|\boldsymbol{X^1}]]$,
where $\mathbb{E}[\boldsymbol{G^1y^1}(\boldsymbol{\hat{\theta}^1_{2SLS}})|\boldsymbol{X^1}] = \boldsymbol{G^1}(\boldsymbol{I} - \hat{\beta}_{1,2SLS}\boldsymbol{G^1})^{-1}\hat{\alpha}_{1,2SLS} + \boldsymbol{G^1}(\boldsymbol{I} - \hat{\beta}_{1,2SLS}\boldsymbol{G^1})^{-1}$ $(\hat{\gamma}_{1,2SLS}\boldsymbol{I} + \hat{\delta}_{1,2SLS}\boldsymbol{G^1})\boldsymbol{X^1}$

**Finally,** use $\boldsymbol{\hat{Z}}$ as a vector of instruments to estimate (5). Note that the vector of covariates coincides with the one used at the first step: $\boldsymbol{\tilde{X}^1}$. Then the following consistent estimator is obtained: $\hat{\theta}_{Lee} = (\boldsymbol{\hat{Z}}^T\boldsymbol{\tilde{X}^1})^{-1}\boldsymbol{\hat{Z}}^T\boldsymbol{PR}$.

This procedure is a modification of a procedure proposed in Lee (2003), therefore, the consistency result is closely related to his Theorem 1:

**Lemma 5** *Under regularity conditions defined in Appendix A, the estimator $\hat{\theta}_{Lee}$ is consistent and has the following limiting distribution,*

$$\sqrt{n}(\hat{\theta}_{Lee} - \theta) \xrightarrow{D} \mathcal{N}(0, \Psi), \tag{8}$$

*with* $\Psi = \sigma^2_\nu(lim_{n\to\infty}\frac{1}{n}\boldsymbol{Z}^T\boldsymbol{Z})^{-1}$ *and*
$\boldsymbol{Z} = [\boldsymbol{i}, \boldsymbol{X^1}, \boldsymbol{G^1X^1}, \boldsymbol{G}^1(\boldsymbol{I} - \beta_1\boldsymbol{G^1})^{-1}\alpha_1 + (\boldsymbol{I} - \beta_1\boldsymbol{G^1})^{-1}(\gamma_1\boldsymbol{I} + \delta_1\boldsymbol{G^1})\boldsymbol{X^1}]$

Discussion and detailed proof of the consistency of such estimator are given in Appendix A.

**Step 2.** I am approaching the estimation of the second step also adopting the 2SLS procedure discussed for the first step. First, the model (5) can be rewritten in the following way:

$$\Delta \boldsymbol{y} = (\alpha_2 - \alpha_1)\boldsymbol{i} + \beta_2 \boldsymbol{G}^2 \boldsymbol{y^2} - \beta_1 \boldsymbol{G^1 y^1} + \tilde{\boldsymbol{\delta}} \boldsymbol{UR} + \gamma_2 \boldsymbol{X^2_{TV}} - \gamma_1 \boldsymbol{X^1_{TV}} + \delta_2 \boldsymbol{G^2 X^2} -$$

$$- \delta_1 \boldsymbol{G^1 X^1} + \Delta\epsilon, \qquad \text{with } \boldsymbol{UR} \text{ defined as discussed in Section 2.2} \tag{9}$$

By $X^1_{TV}$, and $X^2_{TV}$ I denote the subset of covariates, which are time-variant to avoid singularity problem of estimation.

Then a vector of covariates is as follows: $\bar{\boldsymbol{X}} = [\boldsymbol{i}, \boldsymbol{X^2_{TV}}, \boldsymbol{X^1_{TV}}, \boldsymbol{G^2 X^2}, \boldsymbol{G^1 X^1}, \boldsymbol{UR}, \boldsymbol{G^1 y^1}, \boldsymbol{G^2 y^2}]$. Following the logic of the first step I use $\boldsymbol{(G^2)^2 X^2}$ as an instrument for $\boldsymbol{G^2 y^2}$. However, $\mathbb{E}[(\boldsymbol{G^1 y^1})^T \Delta\epsilon] \neq 0$, hence the instrument for $\boldsymbol{G^1 y^1}$ is required. I propose to use $\mathbb{E}[\boldsymbol{G^1 y^1 (\hat{\theta}^1_{2SLS})} | \boldsymbol{X^1}]$ as an instrument, as obtained on the step 1. It is obvious that such an instrument is a valid instrument since it is uncorrelated with the second step error term and is clearly correlated with the outcome variable. Then I define $\boldsymbol{M} = [\boldsymbol{i}, \boldsymbol{X^2_{TV}}, \boldsymbol{X^1_{TV}}, \boldsymbol{G^2 X^2}, \boldsymbol{G^1 X^1}, \boldsymbol{UR}, \mathbb{E}[\boldsymbol{G^1 y^1 (\hat{\theta}^1_{2SLS})} | \boldsymbol{X^1}], \boldsymbol{(G^2)^2 X^2}]$ as a vector of instruments.

I modify (7), taking expectations given $\boldsymbol{X^2}$ and recalling $\mathbb{E}[\Delta\epsilon] = 0$:

$$(\boldsymbol{I} - \beta_2 \boldsymbol{G^2})\mathbb{E}[\boldsymbol{y^2}|\boldsymbol{X^2}] = (\alpha_2 - \alpha_1)\boldsymbol{i} + (\boldsymbol{I} - \beta_1 \boldsymbol{G^1})\boldsymbol{y^1} + \tilde{\boldsymbol{\delta}} \boldsymbol{UR} + \gamma_2 \boldsymbol{X^2_{TV}} - \gamma_1 \boldsymbol{X^1_{TV}} + \\ + \delta_2 \boldsymbol{G^2 X^2} - \delta_1 \boldsymbol{G^1 X^1}$$

$$\mathbb{E}[\boldsymbol{y^2}|\boldsymbol{X^2}] = (\boldsymbol{I} - \beta_2 \boldsymbol{G^2})^{-1}[(\alpha_2 - \alpha_1)\boldsymbol{i} + (\boldsymbol{I} - \beta_1 \boldsymbol{G^1})\boldsymbol{y^1} + \tilde{\boldsymbol{\delta}} \boldsymbol{UR} + \gamma_2 \boldsymbol{X^2_{TV}} - \gamma_1 \boldsymbol{X^1_{TV}} + \\ + \delta_2 \boldsymbol{G^2 X^2} - \delta_1 \boldsymbol{G^1 X^1}]$$

Let $\mathbb{E}[\boldsymbol{G^2 y^2}(\phi)|\boldsymbol{X^2}, \boldsymbol{X^1}] = \boldsymbol{G^2}(\boldsymbol{I} - \beta_2 \boldsymbol{G^2})^{-1}[(\alpha_2 - \alpha_1)\boldsymbol{i} + (\boldsymbol{I} - \beta_1 \boldsymbol{G^1})\mathbb{E}[\boldsymbol{y^1}(\theta^1)|\boldsymbol{X^1}] + \tilde{\boldsymbol{\delta}} \boldsymbol{UR} + \gamma_2 \boldsymbol{X^2_{TV}} - \gamma_1 \boldsymbol{X^1_{TV}} + \delta_2 \boldsymbol{G^2 X^2} - \delta_1 \boldsymbol{G^1 X^1}]$, where $\mathbb{E}[\boldsymbol{y^1}(\theta^1)|\boldsymbol{X^1}] = \boldsymbol{G^2}(\boldsymbol{I} - \beta_1 \boldsymbol{G^1})^{-1}\alpha_1 + (\boldsymbol{I} - \beta_1 \boldsymbol{G^1})^{-1}(\gamma_1 \boldsymbol{I} + \delta_1 \boldsymbol{G^1})\boldsymbol{X^1}$.

Then I also define the following vector $\bar{\boldsymbol{Z}} = [\boldsymbol{i}, \boldsymbol{X^2_{TV}}, \boldsymbol{X^1_{TV}}, \boldsymbol{G^2 X^2}, \boldsymbol{G^1 X^1}, \boldsymbol{UR}, \mathbb{E}[\boldsymbol{G^1 y^1}(\theta^1)|\boldsymbol{X^1}], \mathbb{E}[\boldsymbol{G^2 y^2}(\phi)|\boldsymbol{X^2}, \boldsymbol{X^1}]$

I propose the following estimation procedure:

**First,** compute the 2SLS estimator for $\phi = (\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma_1, \gamma_2, \delta_1, \delta_2)$ of the (7), using a vector of instruments $\boldsymbol{M}$ and a vector of covariates $\bar{\boldsymbol{X}}^{\boldsymbol{1}}$, as defined above. $\hat{\phi}^1_{2SLS} = (\bar{\boldsymbol{X}}^T \boldsymbol{P_M} \bar{\boldsymbol{X}})^{-1} \bar{\boldsymbol{X}}^T \boldsymbol{P_M}(\boldsymbol{y^2} - \boldsymbol{y^1})$, where $\boldsymbol{P_M} = \boldsymbol{M}(\boldsymbol{M}^T \boldsymbol{M})^{-1} \boldsymbol{M}^T$ is a projection matrix.

**Second,** define $\hat{\bar{\boldsymbol{Z}}} = \bar{\boldsymbol{Z}}(\hat{\phi}_{2SLS}) = [\boldsymbol{i}, \boldsymbol{X^2_{TV}}, \boldsymbol{X^1_{TV}}, \boldsymbol{G^2 X^2}, \boldsymbol{G^1 X^1}, \boldsymbol{UR}, \mathbb{E}[\boldsymbol{G^1 y^1}(\hat{\theta}^1_{2SLS})|\boldsymbol{X^1}]],$ $\mathbb{E}[\boldsymbol{G^2 y^2}(\hat{\phi}_{2SLS})|\boldsymbol{X^2}, \boldsymbol{X^1}],$

where $\mathbb{E}[\boldsymbol{G^1 y^1}(\hat{\theta}^1_{2SLS})|\boldsymbol{X^1}] = (\boldsymbol{I}-\hat{\beta}_{1,2SLS}\boldsymbol{G^1})^{-1}\hat{\alpha}_{1,2SLS}+(\boldsymbol{I}-\hat{\beta}_{1,2SLS}\boldsymbol{G^1})^{-1}(\hat{\gamma}_{1,2SLS}\boldsymbol{I}+\hat{\delta}_{1,2SLS}\boldsymbol{G^1})\boldsymbol{X^1}$, with $\hat{\theta}^1_{2SLS}$ obtained as the estimation of the first stage on the first step.

and $\mathbb{E}[\boldsymbol{G^2 y^2}(\hat{\phi}_{2SLS})|\boldsymbol{X^2}, \boldsymbol{X^1}] = \boldsymbol{G^2}(\boldsymbol{I} - \hat{\beta}_{2,2SLS}\boldsymbol{G^2})^{-1}[(\hat{\alpha}_{2,2SLS} - \hat{\alpha}_{1,2SLS})\boldsymbol{i} + (\boldsymbol{I} - \hat{\beta}_{1,2SLS}\boldsymbol{G^1})\mathbb{E}[\boldsymbol{y^1}(\hat{\theta}^1_{2SLS})|\boldsymbol{X^1}]+\hat{\hat{\boldsymbol{\delta}}}_{2SLS}\boldsymbol{UR}+\hat{\gamma}_{2,2SLS}\boldsymbol{X^2_{TV}}-\hat{\gamma}_{1,2SLS}\boldsymbol{X^1_{TV}}+\hat{\delta}_{2,2SLS}\boldsymbol{G^2X^2}-\hat{\delta}_{1,2SLS}\boldsymbol{G^1X^1}]$

**Finally,** I use $\hat{\bar{\boldsymbol{Z}}}$ as a new vector of instrument to estimate (7). Then the following consistent estimator is obtained: $\hat{\phi}_{Lee} = (\hat{\bar{\boldsymbol{Z}}}^T\bar{\boldsymbol{X}})^{-1}\hat{\bar{\boldsymbol{Z}}}^T(\boldsymbol{y^2} - \boldsymbol{y^1})$.

The consistency of this estimator is less straightforward, but it holds under the regularity conditions. The proof of the following Lemma is provided in Appendix A.

**Lemma 6** *Under regularity conditions defined in Appendix A, the estimator $\hat{\phi}_{Lee}$ is consistent and has the following limiting distribution,*

$$\sqrt{n}(\hat{\phi}_{Lee} - \phi) \xrightarrow{D} \mathcal{N}(0, \Phi),$$

*with* $\Phi = (\sigma^2_{\epsilon^1} + \sigma^2_{\epsilon^2})(lim_{n\to\infty}\frac{1}{n}\bar{\boldsymbol{Z}}^T\bar{\boldsymbol{Z}})^{-1}$

### 2.4.2 Correlated effects

If the correlated effects are assumed to be present in the model the fist step model can be written as follows in matrix notation:

$$\begin{aligned}&(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{PR} = \beta(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1Y^1} + \gamma(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{X^1} + \delta(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1X^1} + \eta,\\&\eta = (\boldsymbol{I} - \boldsymbol{G^1})\nu,\ \mathbb{E}[\eta|\boldsymbol{X^1}] = 0\end{aligned}$$

I then use the peer effect model in local differences proposed in Bramoullé et al. (2009):

$$\begin{aligned}&(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{y^1} = \beta_1(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1y^1} + \gamma_1(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{X^1} + \delta_1(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1X^1} + (\boldsymbol{I} - \boldsymbol{G^1})\nu^1,\\&\mathbb{E}[\nu^1|\boldsymbol{X^1}] = 0,\end{aligned}$$

which gives

$$\mathbb{E}[(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1y^1}|\boldsymbol{X^1}] = (\boldsymbol{I} - \beta_1\boldsymbol{G^1})^{-1}(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1}(\gamma_1\boldsymbol{I} + \delta_1\boldsymbol{G^1})\boldsymbol{X^1}$$

The proposed estimation procedure, in this case, is close to the first step with no correlated effects. I redo all the steps with the following vectors of instruments and covariates: instruments $\boldsymbol{S} = [(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{X^1}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1X^1}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{(G^1)^2X^1}]$ and covariates

$\tilde{\boldsymbol{X}}^{\mathbf{1}} = [(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{X^1}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1}\boldsymbol{X^1}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1}\boldsymbol{y^1}]$.

Then I find the 2SLS estimator on the first step and use it to get the new vector of instruments: $\hat{\boldsymbol{Z}} = Z(\hat{\theta}^1_{2SLS}) = [(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{X^1}, (\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{G^1}\boldsymbol{X^1}, \mathbb{E}[(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{G^1}\boldsymbol{y^1}(\hat{\boldsymbol{\theta}}^{\mathbf{1}}_{\mathbf{2SLS}})|\boldsymbol{X^1}]]$, where $\mathbb{E}[(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1}\boldsymbol{y^1}(\hat{\boldsymbol{\theta}}^{\mathbf{1}}_{\mathbf{2SLS}})|\boldsymbol{X^1}] = \boldsymbol{G^1}(\boldsymbol{I} - \hat{\beta}_{1,2SLS}\boldsymbol{G^1})^{-1}(\boldsymbol{I} - \boldsymbol{G^1})(\hat{\gamma}_{1,2SLS}\boldsymbol{I}+$ $+\hat{\delta}_{1,2SLS}\boldsymbol{G^1})\boldsymbol{X^1}$.

The consistent estimator can then be obtained as follows: $\hat{\theta}_{Lee} = (\hat{\boldsymbol{Z}}^T\tilde{\boldsymbol{X}}^{\mathbf{1}})^{-1}\hat{\boldsymbol{Z}}^T\boldsymbol{PR}$. Note that the proof of consistency follows directly by combining the result of Lee (2003) and the proof of Lemma 5, which can be found in Appendix A.

**Step 2** also requires some adjustments in this case. Due to the presence of correlated effects, $\mathbb{E}[\boldsymbol{G^1}\boldsymbol{y^1}(\hat{\boldsymbol{\theta}}^{\mathbf{1}}_{\mathbf{2SLS}})|\boldsymbol{X^1}]$ is no longer observable since it includes the unobserved fixed effects correlated with covariates and cannot be used as an instrument. Hence, I need to modify both vectors of covariates and instruments in the following way: $\bar{\boldsymbol{X}} = [(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{X^2_{TV}}, (\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{X^1_{TV}}, (\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{G^2}\boldsymbol{X^2}, (\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{G^1}\boldsymbol{X^1}, (\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{UR}$, $(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{G^1}\boldsymbol{y^1}, (\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{G^2}\boldsymbol{y^2}]$ is a new vector of covariates. I then use $(\boldsymbol{I}-\boldsymbol{G^1})(\boldsymbol{G^2})^{\mathbf{2}}\boldsymbol{X^2}$ as an instrument for $(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^2}\boldsymbol{y^2}$. I propose to use $\mathbb{E}[(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1}\boldsymbol{y^1}(\hat{\boldsymbol{\theta}}^{\mathbf{1}}_{\mathbf{2SLS}})|\boldsymbol{X}]$ as an instrument for $(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1}\boldsymbol{y^1}$. This instrument is clearly a valid instrument since it is uncorrelated with the second step error term and is clearly correlated with the outcome variable. Then I define $\boldsymbol{M} = [(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{X^2_{TV}}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{X^1_{TV}}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^2}\boldsymbol{X^2}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1}\boldsymbol{X^1}, (\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{UR}, \mathbb{E}[(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1}\boldsymbol{y^1}(\hat{\boldsymbol{\theta}}^{\mathbf{1}}_{\mathbf{2SLS}})|\boldsymbol{X}], (\boldsymbol{I} - \boldsymbol{G^1})(\boldsymbol{G^2})^{\mathbf{2}}\boldsymbol{X^2}]$ as a vector of instruments.

Applying the same changes to all relevant vectors, I then fully repeat the estimation procedure of the case of no correlated effects, and obtain the consistent estimator. Consistency of the estimator is achieved by the argument similar to the one in Lemma 6, proof of which and more detailed discussion on estimation procedure can be found in Appendix A.

# 3 Data and Descriptive analysis

## 3.1 The system of higher education in Russia and specifics of the sampled university.

People with completed full vocational education or completed professional education of non-university level are eligible to enter the university[9]. Most of the places in the

[9] It is more accurate to call them *the institutions of tertiary or post-secondary education*, since not all of these institutions in Russia have the status of the university, however, *the university* will be used for simplicity

state universities are financed by the government: around 65%[10], but it differs among institutions. For example, the analysed in this paper university, National State University - Higher School of Economics, Nizhny Novgorod branch, provided 340 state-financed places out of total 431 in 2012[11]. The tuition fee varies from institution to institution, in our example, it varies between 130000 and 165000 Rubles, which equals to 28-36 times the minimum monthly wage or 18-23 times the minimum cost of living in Russia.

The students are accepted to the universities depending on the scores of the obligatory standardized examination, Unified School Examination, conducted at the end of the last school year. Each high school graduate has to take the exam in several subjects: Mathematics and Russian are mandatory to graduate from the school, the other subjects are chosen by the graduates depending on their preferences and the requirements of the universities they are aiming to apply to. For example, economic department of NRU-HSE requires the USE results in Social Studies (a mixture of basic knowledge about different aspects of society: philosophy, sociology, social psychology, law, political science) and Foreign language additional to the mandatory to all graduates Mathematics and Russian. However, regional and national level Olympiads can often be used as the second channel to enter some of the universities. These Olympiads are subject-specific and considered to be more sophisticated than the school exams, so they are designed to attract more talented students. In Higher School of Economics, the winners and prize-takers of these competitions are accepted to the university without exams if the major of the Olympiads corresponds to the university department (Economic Olympiads for economics department, Entrepreneurship Olympiads for management department etc.) or automatically given the highest score for the other subjects. However, those students are still required to take the USE and have the scores not lower than the required minimum (65 out of 100 in 2015, significantly lower than the requirement to be accepted). The share of students entering universities using the Olympiads results is around 5-6% overall in Russia, but it is much higher for the Higher School of Economics, around 40%, because of the selective status of HSE. Therefore, in general, the group of students entering HSE is more or less homogeneous and consists of the high-achievers. Even though Nizhniy Novgorod branch of HSE is less selective than the main Moscow branch, the level of the admitted students is still very high. The list of all accepted students is publicly available in the university itself as well as on the website.

Usually, universities in Russia have an exogenous group formation. The students are randomly split into groups of 20-30 people before the beginning of the studies. These

[10]According to the Monitoring of education markets and organizations (MEMO), NRU HSE. In Russian
[11]The main dataset uses 2012 cohort of students, details are described in the next subsection

groups stay mainly intact for the first three years. Several groups or even all students attend lectures together, whereas each group has separate tutorials. The changes to the group structure may occur if a lot of the students leave the university and the group is too small. Most of the universities have by now adopted the Bologna Process model of 4 years for Bachelor's degree and 1-2 years of Master's degree. In most cases each academic year has 2 terms with exams periods after each, however, HSE has 4 terms per year, with some exams or pass-fail exams after the 1st and the 3rd term and with most exams after the 2nd and the 4th term. The student is not allowed to fail 3 or more exams per half (1+2 or 3+4 term) and the retakes are conducted only after the 2nd and 4th exam periods of each year. All results of all students are publicly available near the students' office in the university and online so that everybody can follow their own performance, compare to the peers, and the tuition students can understand, whether they are eligible for the tuition discount.

## 3.2 Data description

The data is based on two longitudinal studies of the students' network, conducted in the National Research University Higher School of Economics (Nizhny Novgorod branch; state university). The information about the studies is summarized in Table 1.

Table 1: Studies characteristics

| Study | Cohort | Frequency | Departments | Total students |
|---|---|---|---|---|
| I | 2012 | Each year | Economics, Management, Law, Computer Science | 321 |
| II | 2013 | Each 3-4 month | Economics, Management | 205 |

Students were asked to indicate three and two networks correspondingly: friends from the university (same cohort), students from the same cohort, whom they ask for help (in the first study this question is divided into two: help with mathematical subjects and help with humanities). The I study is of the main interest of the paper due to the longer periods between the surveys that are able to capture a more persistent trend of the network dynamic. The II study is used only for the robustness check of the results.

Other data include all exam results, information about retakes and dropouts from the administrative university data, as well as some personal data: gender, high school examination results, type of living (dormitory or not, roommates for those who live in the dormitory), parental education, some indicators of willingness to succeed or efforts (time spent on homework, time spent online on social networks, indicator of having a job parallel to studies).

The typical problems of self-reported data are present in the dataset. There are several observations with partially missing data on the network links. These entries need to be handled with care since they might suggest both the students without friends, indicating the antisocial behavior, or the students that just skipped the questions, while answering the questionnaire. In the *I study* 13 students indicated no friends links, however, two of those provide an information about connections in the help networks, which might demonstrate an antisocial behavior of the students. There is no information on particular friends for 4 more students, who just said they are friends with a lot of students, or even with all students. In the *II study*, there are 9 students without links, however, it is not clear, whether they did not report anybody at all or whether they answered with a sentence, as 4 students from the *I study*, mentioned above.

Sampling is of a slight concern as well. The first survey has 321 observations out of 396 students that entered the 4 departments of the university in 2012, the second has 205 out of 253 students, started in 2013 in 2 departments, that gives approximately 75-80% of the full population of students (**Table 2**). Some of the students could have indicated the link to somebody outside of the sample, which can lead to overestimation of the importance of the observed links. However, the survey was conducted on several occasions, during lecture periods, so those, who did not answer the survey, are likely to attend the university only infrequently, and hence to have less influence on the other students.

Table 2: Comparison of samples and population

| | | Sample | All students | Share |
|---|---|---|---|---|
| I, year 1 | Size | 320 | 432 | 74.07% |
| | Retakes | 157 | 203 | 77.34% |
| | Dropouts | 16 | 40 | 40 % |
| I, year 2 | Size | 296 | 393 | 75.32% |
| | Retakes | 148 | 190 | 77.89% |
| | Dropouts | 24 | 39 | 62.54% |
| II | Size | 205 | 254 | 80.7% |
| | Retakes | 65 | *137* | - |
| | Dropouts | 6 | 21 | 28.57% |

**Table 2** also demonstrates an inability of the dataset to catch all the information about the dropouts (only 40% are present in the first survey) and their small amount in the network. This makes the econometric analysis of dropouts implausible, and forces to study exam retakes instead.

Note that the *I study* restricted the friends' network to 7 names, whereas the *II study* did not put any restriction. This lead to almost 50% of the students in the first period of *I study* reporting exactly 7 friends, whereas only 13,5% of the students in the *II study* indicated the same number of friends. The second wave of the long study also has space for mentioning the maximum of 7 friends, however, this restriction is not mentioned in the question itself. Therefore, 7 friends are the maximum of the 2nd wave of the long study with only 10% indicating exactly 7 friends. The distribution of the number of the friends for both studies is presented in Table 1 and Figure 1 of Appendix B. The average and median number of connections is 6 in both first year of the I sample and in the II sample, whereas it is 4 in the second wave of the I study. It is likely, that in the first wave some of the students had to restrict themselves to exactly 7 names, whereas some felt obliged to include more people than they are actually tightly connected to, which may cause underestimation of the importance of some links and overestimation of the others. Lower average number of friends in the second period may be caused by particularities of the survey construction as well as by the real trends in the network development.

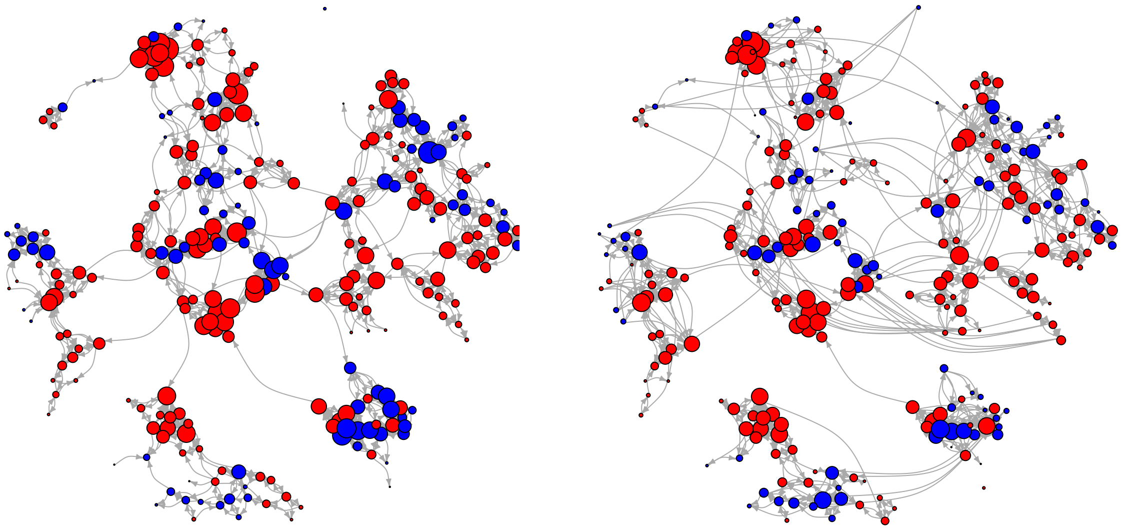
The survey design is different for three networks. The first wave of the I study asks for no more than 7 friends and has 7 lines for the names, which was ignored by approximately 2% of the sample, the second wave of the I study does not put any restriction on the number of friends, although it has 7 lines as well, the short study says explicitly, that a number of friends can be unlimited, but has 15 lines. Therefore, the survey design may influence estimation results from the I survey analysis, hence the analysis of the II study with its unchanged question design can be helpful as a robustness check.

## 3.3 Network characteristics

In this section, I will discuss the validity of the identifying assumptions in a framework of the I survey.

**Network stability.**

**Figure 1** visualizes the whole networks for the first wave (left) and for the second wave (right). Red nodes are females, blue - males, the size of the nodes is proportional to the overall degree of the node. It can be observed from this figure that two networks differ. For example, two clusters in the bottom part of the graph are not connected in the first wave, whereas there are several edges between them in the second wave.



(a) Wave 1. (b) Wave 2.

Figure 1: Networks.

More formal justification of the variability of the network is presented in **Table 3**. Quite a lot of variation can be observed: around 11-12% of the students reported exactly the same set of friends. However, the share of completely new networks varies with gender. Females have only 5% of completely new networks. Hence, females tend to be more persistent in forming and retaining the links.

**Table 4** provides more evidence of the network variation: only around 16% of the links survived after the first period, and around 78% of the links formed in the second period are new.

Table 3: Overlap of network partners

| Network statistics | Full sample | Male | Female |
|---|---|---|---|
| Complete overlap | 11.49 | 11.21 | 11.89 |
| No new links | 24.66 | 22.43 | 26.49 |

| | | | |
|---|---|---|---|
| Partial overlap | 65.20 | 46.73 | 77.30 |
| Complete turnover | 12.16 | 24.30 | 5.41 |
| Observations | 296 | 107 | 185 |

*Note:* Percentages of 1st, 3rd, and 4th rows do not add up to 100%, because there are new observations in the 2nd wave, for which we do not observe the network in the 1st wave

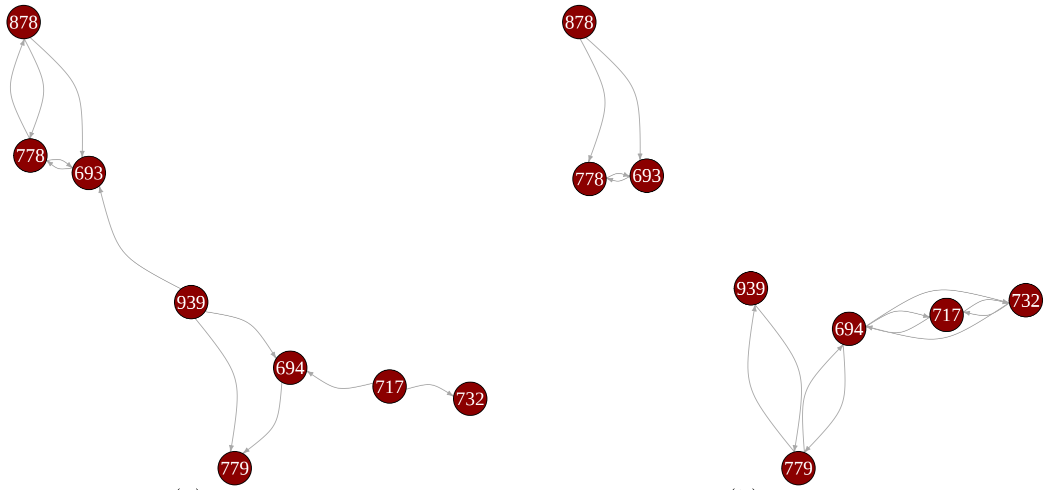
Table 4: Some network characteristics

| Network statistics | Definition | 1 year | 2 year |
|---|---|---|---|
| Average indegree | Average number of ingoing ties | 4.96 (2.73) | 3.93 (2.53) |
| Average outdegree | Average number of outgoing ties | 4.96 (2.01) | 3.93 (2.2) |
| Density | Proportion of existing ties in the network | 0.015 | 0.014 |
| Reciprocity | Proportion of ties which are reciprocated | 0.639 | 0.636 |
| Transitivity | The ratio of the triangles and the connected triples in the graph | 0.454 | 0.443 |
| Share of the links that remained from the 1st wave in total amount of links of the 2nd wave | | - | 22.61% |
| Share of the links that remained from the 1st wave in total amount of links of the 1st wave | | 16.57% | - |

**Transitivity**.

**Table 4** describes several characteristics of the networks in the sample. The transitivity is measured by the shares total amount of connected triangles in the whole graph. So in more than 50% of all possible sets of three students, at least, one link is missing.

**Figure 2** shows the subgraph of the network to demonstrate the existence of intransitive triads in both of the samples. For example, in wave 1 the following triad is intransitive: 717 $\rightarrow$ 694, 694 $\rightarrow$ 779, but 717 $\nrightarrow$ 779. Other examples of intransitive triads are: 939$\rightarrow$ 693$\rightarrow$ 778, 693$\rightarrow$778$\rightarrow$ 878 in the first wave and 939$\rightarrow$ 779$\rightarrow$ 694, 779→ 694→ 717 in the second wave, and some more.



(a) Wave 1. (b) Wave 2.

Figure 2: Subgraph of the network

The characteristics of the networks (see **Table 4**) also clearly suggest that the network is directed and cannot be assumed to be indirected since only around 60% of the links are reciprocal. Also, the networks are sparse with the density of the links around 1.5%.

## 3.4 Descriptive analysis

People often tend to connect based on similarities in their observed and unobserved characteristics. **Table 5** summarizes the findings on the affinity of the peers in the network. Most of the peers are coming from the same group, more than 84% and almost all friends are from the same department. The network can in principle be divided into four smaller networks.

Females are more likely to connect with peers of the same gender, whereas males have more diverse networks. Gender difference also exists in the probability of connecting to the dormitory mates: males are more likely to connect. The share of the friends with the same living conditions is, however, decreasing with time, suggesting that some other characteristics matter more for creating and sustaining the links.

Future plans on average seem not to matter a lot for the link formation: friends with the same plans for the future education are about 50% of the peers. This share could probably be higher, if the students were asked about there plans later in the course of their studies, and not during the first year. However, given the student's willingness to do Master, her peers are as well more oriented on continuing the studies after the Undergraduate level.

Table 5: Characteristics of reported networks links by sex

| Variables | 1st wave | | | 2nd wave | | |
|---|---|---|---|---|---|---|
| | Male | Female | All | Male | Female | All |
| Average size | 4.53 | 5.19 | 4.96 | 3.57* | 4.18* | 3.93* |
| Average size (with out of sample links) | 5.22 | 5.78 | 5.58 | 4.29* | 4.96* | 4.69* |
| Study group/department relation (% of network partners) | | | | | | |
| Same group | 84.17 | 87.23 | 86.76 | 87.21 | 89.89 | 88.78 |
| Same department | 98.54 | 99.21 | 98.99 | 97.54 | 99.39 | 98.95 |
| Individual characteristics of network partners(% of network partners) | | | | | | |
| Same gender | 64.05 | 81.97 | 76.18 | | | |
| Same working status | 62.43 | 70.33 | 67.78 | 50.74 | 60.95 | 56.41 |
| Same education of mother | 61.75 | 66.84 | 65.19 | - | - | - |
| Same education of father | 56.45 | 50.08 | 52.14 | - | - | - |
| Same living conditions | 57.59 | 46.71 | 50.23 | 50.97 | 39.61 | 43.33 |
| Same living conditions (dorm/not) | 84.14 | 76.23 | 78.79 | 74.27 | 70.55 | 72.16 |
| Future plans (% of network partners) | | | | | | |
| Same plans for Master | 54.44 | 57.37 | 56.41 | - | - | - |
| Same plans for Doctorate | 47.18 | 47.32 | 47.27 | - | - | - |
| *Subsample of planning to do Master:* | | | | | | |
| Same plans for Master | 68.34 | 72.42 | 74.46 | - | - | - |

*the network data in the 2nd wave is truncated at 7 friends

More than 1/3 of all links in the first wave are links to the students with retakes (37%). The share of the links to the students with retakes in the first period in the total amount of second wave links is slightly smaller: 33%. It might be caused by the intention of students to improve their peer group and connect to peers with higher outcomes. The average amount of the friends with retakes in the first period is 1.83 while it is lower for the second period: only 1.25. The average amount of peers with exam retakes for the subsample of all students that have at least one peer with retake is higher than the average of the full sample and is equal to 2.5. For the same students in the second wave, the average number of peers who had exam retakes in the first period is now much lower: 1.55. It can be suspected that the decrease in this value may be partially explained by the readjustments of the network towards better connections. Moreover, for the same subsample, the average number of peers with retakes in the second period is even lower: 1.37. Interestingly, some of those, who didn't have any friends with retakes in the first period, connected to new peers that had the retakes in the second period, the average number of such friends is only 0.35 though, but the average number of friends with retakes in the next period is 0.57. So the changes in the network are leading to the improvements as well as worsening of the new peer group. These findings are summarized in the **Table 6**.

Table 6: Distribution of retakes

| | Links wave 1, retakes wave 1 | Links wave 2, retakes wave 1 | Links wave 2, retakes wave 2 |
|---|---|---|---|
| Share of retakes links in all links | 36.99% | 32.99% | 29.15% |
| Average amount of friends with retakes | 1.83 | 1.25 | 1.15 |
| *Subsample with retakes of friends* Average amount of friends with retakes | 2.5 | 1.55 | 1.37 |
| *Subsample no retakes of friends* Average amount of friends with retakes | 0 | 0.35 | 0.57 |

Observe that students in the studied framework tend to connect to peers, having higher average grades than the students themselves, for the full sample as well as for the samples with and without retake friends. Students, who do not connect to peers with retakes, are performing better than those, whose friends are having retakes. However, the improvements in the performance in the future are not significant, with the changes in the performance of the students without peers' retakes being slightly higher.

Table 7: Average grades in samples and subsamples

| | Full sample | With retakes of friends | No retakes of friends |
|---|---|---|---|
| Average grade | 7.04 (0.99) | 6.98 (0.96) | 7.37 (0.98) |
| Average grade of friends | 7.18 (0.65) | 7.03 (0.63) | 7.68 (0.49) |
| Sample size | 320 | 234 | 86 |
| Average grade next period | 7.13 (1.14) | 7.02 (1.15) | 7.44 (1.07) |
| Sample size | 297 | 217 | 80 |

It is not possible to distinguish between the predicted and unexpected components of retakes by simply looking at the data. Therefore, the deeper econometric analysis is needed to make conclusions about the existence and the magnitude of the effect of unpredicted shock.

# 4 Results

## 4.1 Main specification

I use the following variables for the main specification of the model:

**Outcome:** average weighted grade of the student in the corresponding period. The grades are summed up weighted by the amount of the credits assigned to the particular course.

**Retakes:** indicator of at least one retake in the first period.

**Initial ability,** measured as the sum of mandatory Unified State Examinations (mathematics and Russian) plus the sum of cross-products between these USE results and a dummy of winning any relevant Olympiads.

**Controls:** **time-invariant**, such as gender, socio-economic background like a dummy of parental higher education, a dummy of having a single parent before entering the university and dummy for siblings; and a set of dummies for three departments with law department serving as a base.

**Controls:** **time-varying**, such as tuition, which is mostly time-invariant, but some rare students change the type of tuition, working status (dummy for not working versus any type of job) and living conditions (dormitory versus everything else).

Descriptive statistics for these variables is provided in **Table B.3** of Appendix B. It can be observed that the average changes in the time-variant variables are rather modest, as well as the changes in the performance. However, the average grade has higher standard deviation and spread in the second period.

**Table 8** summarizes some of the findings of the estimation of the model without correlated effects. Note that the sample size is smaller than was discussed in the data description, due to the absence of some students in one of the waves. And it is critical to have the information in both waves for each of the students to estimate the effect.

Table 8: Estimation of main specification

| Variable | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Constant | -0.1521 | | -0.1840 | | -0.0482 |
| Unexpected Retake | -0.2638 | -0.2143 | -0.3077• | -0.2064 | -0.3907* |
| Endogenous effect, period 1 | -0.0307 | -0.0425 | -0.0317 | 0.0908* | 0.0614* |
| Endogenous effect, period 2 | 0.0205 | 0.0085 | 0.0218 | 0.0419 | 0.0306 |
| *Time-variant own controls* | | | | | |
| Tuition, w1 | 0.0208 | | | 0.0102 | |
| Tuition, w2 | -0.0912 | | | -0.1518 | |
| Working status, w1 | -0.0664 | -0.0719 | -0.0716 | | |
| Working status, w2 | 0.1381• | 0.1147* | 0.1346• | | |
| Living in dorm, w1 | | | | | 0.1061 |
| Living in dorm, w2 | | | | | 0.1651 |
| *Network's controls* | | | | | |
| Economics, w1 | 0.2417 | 0.1568 | 0.2692 | -0.0732 | |
| Economics, w2 | -0.4681** | -0.4367** | -0.4513** | -0.5893*** | |
| Management, w1 | 0.5409* | | 0.5712* | | |
| Management, w2 | 0.1790 | | 0.1996 | | |
| Working status, w1 | | | | -0.7352* | -0.5420** |
| Working status, w2 | | | | -0.0497 | -0.1903 |
| HE of father, w1 | | 0.4010• | | | |
| HE of father, w2 | | 0.0006 | | | |
| Sample size | 250 | 250 | 250 | 250 | 250 |
| BIC | -216.68 | -225.24 | -226.51 | -225.79 | -196.71 |

*** - p-value < 0.01, ** - p-value < 0.05, * - p-value < 0.1, • - p-value < 0.15

It can be observed that for the full sample the estimator of the effect of the unpredicted component of retakes is negative but in most cases insignificant. The magnitude of the effect in specification (5) suggests that if a friend of some student had a retake during the first year, which this student couldn't predict at all, the difference between the average grade of year 2 and the average grade of year 1 of the student will be on average 0.05-0.39 lower, than in the case the student expected the retake of a friend, depending on the total number of friends. For example, the median student on average improves her grades in the second period in comparison to the first by 0.24, which is 2.4% of the maximum grade. The presence of unpredicted retake of the friend, other things equal, may leave the average grade in the second period at the same level or even decrease it up to 1.5% of maximum grade, changing the direction of the dynamics and, moreover, putting the student on average 5-25 positions lower in the overall students' rating, falling lower with less friends.

Note, that there is a highly significant difference between the economics and other departments for most of the specifications. On average, students of economics department have -0.5 lower difference of grades, which suggest the overall lower grades of the economics department in the second year. This evidence indicates the necessity of using the model with correlated effects or treating the departments separately by splitting the full sample.

Discussing the results for those, who had their own retakes, versus those, who did not is the other possible way to improve the estimation results.

The further analysis is given in the next subsections, where I present the results of estimation in the subsamples, of the model with correlated effects, as well as the estimation with a possibly improved network. However, it is worth pointing out, that the sample size for the main specification is 250 students, which may be not sufficiently big to capture the desired effect, and the results of the estimation in the subsamples should be treated with even more care, since with the lower sample size the asymptotic properties of the proposed estimator may suffer.

## 4.2 Connection to one's own retake

I first report the results for the subsamples of students with and without own retakes. It can be suggested that the students that had their own retake may, in general, be connected to worse peers. Therefore, having friends with retakes might lower the performance even further, whereas the friends' retakes are more likely to have either no effect or even positive influence for the better students.

Table 9: Presence of own retake

| Variable | (1), yes | (2), yes | (1), no | (2), no |
|---|---|---|---|---|
| Constant | 0.2602 | 0.1027 | -0.1001 | -0.3734* |
| Unexpected Retake | -0.2092 | -0.1788 | 0.0246 | 0.0586 |
| Endogenous effect, period 1 | -0.0237 | -0.0539 | 0.0352 | -0.0262 |
| Endogenous effect, period 2 | 0.0756** | 0.0671• | 0.0429 | 0.0577• |
| *Time-variant own controls* | | | | |
| Tuition, w1 | -0.1612 | | 0.0032 | |
| Tuition, w2 | -0.1834 | | -0.2685 | |

| | | | | |
|---|---|---|---|---|
| Working status, w1 | | -0.1175 | | -0.0713 |
| Working status, w2 | | 0.1379 | | 0.1192 |
| | *Network's controls* | | | |
| Economics, w1 | -0.1964 | -0.0474 | 0.0616 | 0.1081 |
| Economics, w2 | -0.9973*** | -0.9131*** | -0.3932 | -0.5344• |
| Management, w1 | | | | 0.3098 |
| Management, w2 | | | | -0.0338 |
| Working status, w1 | | | -0.5632** | |
| Working status, w2 | | | -0.2305 | |
| HE of father, w1 | 0.6306• | 0.3753 | | |
| HE of father, w2 | -0.3627 | -0.4972• | | |
| Dummy siblings, w1 | | 0.7689*** | | |
| Dummy siblings, w2 | | 0.2113 | | |
| Sample size | 83 | 83 | 167 | 167 |
| BIC | -336.10 | -348.14 | -288.34 | -290.82 |

*** - p-value < 0.01, ** - p-value < 0.05, * - p-value < 0.1, • - p-value < 0.15

**Table 9** gives hints that the outcomes are influenced differently by the peers in case of presence of one's own retake and in the case when the student passed all the exams from the first attempt. First of all, unexpected retake has an insignificant and negative effect of higher magnitude in case of own retake than without own retakes. So, when students in the network have retakes together, they will less likely improve in the future. It may be partially explained by the worse peer group, and partially by the fact that fewer friends are able to help to catch up with the courses after retakes. It can also be observed that the endogenous effect changes the sign, from negative to positive, and is more significant for students with own retakes, which may suggest that the students, especially the ones with their own retake tend to seek for the better peers in the future. However, the data does not provide evidence that the willingness to connect to better peers is coming from the discussed shock, therefore, the changes may be considered as a natural learning process.

### 4.3 Effects in different departments

In this subsection, I discuss the results for subsamples of different departments. I present the results for two departments: economics and management. The economics department showed significantly different results in comparison to the others in the main specification, and the management department is quite similar to the economics in the curriculum and direction of study.

Table 10: Departments

| Variable | (1), Econ. | (2), Econ. | (3), Man. | (4), Man. |
|---|---|---|---|---|
| Constant | -0.5228* | -0.2265 | 0.5614* | 0.7032** |
| Unexpected Retake | -0.4375 | -0.4794 | 0.4043 | 0.3943 |
| Endogenous effect, period 1 | -0.0426 | -0.0150 | 0.2884*** | 0.3927*** |
| Endogenous effect, period 2 | 0.0334 | 0.0544 | 0.1635** | 0.2164** |
| *Time-variant own controls* | | | | |
| Tuition, w1 | | 0.2538 | | -0.4889 |
| Tuition, w2 | | -0.0479 | | -0.7778* |
| Working status, w1 | -0.0888 | | 0.0054 | |
| Working status, w2 | 0.2119 | | 0.0880 | |
| *Network's controls* | | | | |
| Ability, w1 | | | -0.0056* | -0.0062** |
| Ability, w2 | | | -0.0049** | -0.0048** |
| Gender, w1 | | 1.0102* | | |
| Gender, w2 | | 0.2889 | | |
| Working status, w1 | | -0.6228 | | -0.6353* |
| Working status, w2 | | -0.5707 | | -0.4209 |
| HE of mother, w1 | | | -0.4308 | -0.7535* |
| HE of mother, w2 | | | -0.3144 | -0.5867* |
| Dormitory, w1 | -1.5248*** | -0.8094• | | |
| Dormitory, w1 | -0.9558** | -0.8145* | | |
| Dummy siblings, w1 | 0.4005 | | | |
| Dummy siblings, w2 | -0.3945 | | | |
| Sample size | 82 | 82 | 68 | 68 |
| BIC | -305.45 | -300.61 | -456.68 | -471.57 |

*** - p-value $< 0.01$, ** - p-value $< 0.05$, * - p-value $< 0.1$, • - p-value $< 0.15$

As can be seen from **Table 10**, the discussed effect is surprisingly different for two departments. While specification (1) for the economic department have the negative effect of the unexpected retake, the same effect in the specifications for management is positive. However, estimators are not significant. Both subsamples have a small number of observations, which can cause the low significance of the effect of the interest, and the results should be treated with caution. It is possible to eliminate the differences between the departments and estimate the full sample, by exploring the model with correlated effects.

### 4.4 Estimation in presence of correlated effects

In this Subsection, I would like to discuss the results of the estimation proposed in Section 2.4.2. Simple estimation in the presence of correlated effects might lead to the biased results. Next table presents the summary of results, judging from which I can then compare the two specifications: with and without correlated effects.

Table 11: Estimation of specification with correlated effects

| Variable | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Unexpected Retake | -0.4144• | -0.3899• | -0.3817• | -0.3817• | -0.4616* |
| Endogenous effect, period 1 | -0.0361 | -0.0526 | -0.0378 | -0.0379 | 0.0186 |
| Endogenous effect, period 2 | 0.0143 | 0.0016 | 0.0544 | 0.0544 | 0.0461 |
| *Time-variant own controls* | | | | | |
| Tuition, w1 | | 0.0834 | 0.0411 | | |
| Tuition, w2 | | -0.1011 | -0.1292 | | |
| Working status, w1 | -0.0382 | 0.0266 | | 0.0411 | 0.0590 |
| Working status, w2 | 0.1077 | 0.1355• | | -0.1292 | 0.0991 |
| Living conditions, w1 | -0.1323 | | | | |
| Living conditions, w2 | 0.2102 | | | | |
| *Network's controls* | | | | | |
| HE of mother, w1 | -0.6547 | -0.5532 | -0.5785 | -0.5785 | -0.6717 |
| HE of mother, w2 | -0.2011 | -0.1541 | -0.3396 | -0.3396 | -0.3875 |
| HE of father, w1 | 0.5325 | 0.5789 | 0.4663 | 0.4663 | |
| HE of father, w2 | -0.0167 | 0.0378 | -0.0831 | 0.3817 | |
| Sample size | 250 | 250 | 250 | 250 | 250 |
| BIC | -183.89 | -185.56 | -195.56 | -192.25 | -197.99 |

*** - p-value $< 0.01$, ** - p-value $< 0.05$, * - p-value $< 0.1$, • - p-value $< 0.15$

Controlling for correlated effects leads to more significant and persistent value of negative effect of unexpected retakes than in the main specification. The magnitude of the effect in specification (5) suggests that if a friend has a retake during the first year, which the student couldn't predict at all, this will make the difference between the average grade of year two and the average grade of year one for this student on average 0.46 lower, if the student has only 1 friend, and approximately 0.065 lower, if the student has 7 friends. The maximum of the grades is 10 so that the person lose almost 5% of the maximum grade when the network includes friends with retakes.

## 4.5 Additional analysis

### 4.5.1 Improving network

As it was mentioned before, students were asked to name up to 7 friends from their cohort, although some named more than 7. However, it is reasonable to assume that all named friends are not equal for the person. I introduce two possible ways to account for better friends so that the quality of the network can be improved.

First, I assume that the friends named among the first are more important than the others, since they were remembered earlier, and the best friends can't be named last. I reduced the network, only taking up to three named first students. I conducted analysis for both models with and without accounting for correlated effects. The suggested improvement of the network didn't, however, increased the significance of the results[12]. The effect of an unpredicted component of friends' exam retake is not significantly different from zero. Therefore, it might be reasonable to conclude that the unexpected negative or positive performance of the whole network of friends is more important for the future performance of students than the performance of only best friends.

Second, I observe that about 60% of the network is reciprocal, so I conduct similar analysis limiting the network to only reciprocal connections. This again does not bring any improvement in terms of the significance of the studied effect. It seems that the students' performance is shaped not only by their mutual friends, but although by those, who don't consider them as friends, but are considered as friends by the students. These students may be viewed as a sort of role models, and therefore, are important to be taken into account.

Thus, the initial full network is able to capture the effect of unexpected shock better than the versions of the network, considered initially as possible improvements.

### 4.5.2 Important classes

The further analysis divides the subjects, studied by the students in the sample, into two parts: more important and less important. All subjects have the corresponding amount of ECTS credits, from 0 to 8 with average around 2.5. For the analysis, I set the threshold of 4 ECTS points. However, some subjects have several exams, for example, Mathematical Analysis, and the weight of some of the exam in the series can be lower

[12]The detailed results are presented in Appendix B

than 4, but, at least, one exam has ECTS higher than 4. In these cases, I am including all the exams of the series in the sample of important exams. This restricts the set of the students with retakes to 2/3 of the initial set.

**Table 12** provides the results of the analysis in the new setting for the model without correlated effects.

Table 12: Estimation with retakes for classes with ECTS 4 and higher

| Variable | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Constant | | -0.2176 | -0.1670 | | -0.2005 |
| Unexpected Retake | -0.4912** | -0.5484*** | -0.5158** | -0.4907** | -0.5564** |
| Endogenous effect, period 1 | 0.1072* | -0.0211 | -0.0160 | 0.1076* | -0.0158 |
| Endogenous effect, period 2 | 0.0378 | 0.0279 | 0.0284 | 0.0401 | 0.0307 |
| *Time-variant own controls* | | | | | |
| Tuition, w1 | 0.0417 | | 0.0430 | 0.0530 | |
| Tuition, w2 | -0.0861 | | -0.0575 | -0.0830 | |
| Working status, w1 | | -0.0568 | -0.0570 | | -0.0616 |
| Working status, w2 | | 0.1488* | 0.1494* | | 0.1469* |
| Living conditions, w1 | -0.0222 | -0.2673 | | | |
| Living conditions, w2 | 0.0312 | -0.1808 | | | |
| *Network's controls* | | | | | |
| Economics, w1 | -0.1032 | 0.1652 | 0.1257 | -0.1129 | 0.1387 |
| Economics, w2 | -0.6177*** | -0.5279** | -0.5466** | -0.6240*** | -0.5411** |
| Management, w1 | | 0.4322 | -0.4049 | | 0.4159 |
| Management, w2 | | 0.1209 | 0.1045 | | 0.1127 |
| Working status, w1 | -0.8120* | | | -0.8186** | |
| Working status, w2 | -0.0074 | | | -0.0102 | |
| Sample size | 250 | 250 | 250 | 250 | 250 |
| BIC | -215.86 | -220.76 | -220.81 | -226.98 | -230.99 |

*** - p-value < 0.01, ** - p-value < 0.05, * - p-value < 0.1, • - p-value < 0.15

It can be observed that a lot of the results resemble the results for the model with all retakes, however, the effect of the unpredicted retake is more significant when only important classes are taken into consideration. The sign of the estimator remains negative but it gains much more significance, suggesting the different effect that different classes may have on the future performance of the network. The results also suggest the higher magnitude than in the initial model. Now, the friend's unexpected retake of the important class may make the difference between average grades in two periods bigger and reduce the average grade of the second year additionally by up to 0.5, which equals to 5% of the maximum grade.

This result is expectable. For example, the new set of retakes does not include the class of Discrete Mathematics in the Economics department but includes Mathematical Analysis. These two classes differ not only in the amount of ECTS but also in the length and importance for the further classes. Mathematical Analysis is studied throughout the whole length of the first year, whereas Discrete Mathematics only for one term. Moreover, the former introduces a lot of methods used later in the core classes of the higher years, such as Micro or Macro, while the latter might be considered to contribute less in future studies. The full list of classes, which were retaken at least once and the subset of more important classes are presented in Appendix B.

The significance of a dummy of the Economics department suggests that the model with correlated effects may be used, as in the model with the full set of retakes. Surprisingly, the estimator of the effect of the unexpected retake in the model with correlated effects loses the significance once I restrict the set of the retakes.

# 5 Conclusion

The paper discusses the spread of the unpredicted shock across the network of friends in the university environment using the newly introduced dynamic peer effect model in the presence of endogenous shock.

Exam retakes play an important role in determining the future of the student. However, it was shown that the unpredicted component of the retake may influence not only the students with a retake but also the whole network of friends. In most of the cases the effect is not very significant, but still should not be ignored. When the threshold of failing the exam is too high, some students, viewed by their friends as high-achievers, are likely to fail. This anticipation mistake leads to the decrease of the average grades of the whole friendship network.

The ideas explored in this paper can be further extended to the analysis of the networks in other settings, not only for educational outcomes. The method is applicable, when the endogenous shocks might have the longitudinal effect on the network outcomes, such as, for example, a treatment that for some reasons cannot be randomized, or conversational networks in developing communities, etc.

I have presented the results for identification of such models, that allow disentangling the effect of unpredicted shock on the future performance. The findings of the paper suggest that it is sufficient to assume time-variability of networks together with the existence of intransitive triads (or distances of length three, depending on the correlated effects assumption) in each of the states of the network for the similar models. Intransitive triads are guaranteed by the presence of two students only connected via the third common friend but not directly. The characteristics of friends of the friends don't influence directly the outcome, and, therefore, can be used as an instrumental variable for the friends' outcome. Such instruments can, therefore, deal with endogeneity issue. The group of new friends, different from the group of old friends, let the model capture the changes, happening due to the shock.

The procedure developed in the paper is shown to yield consistent estimators of the individual characteristics, endogenous peer effect and effect of unpredicted shock.

All theoretical findings are tested on the dataset of university students, connected via the friendship network. Most of the empirical evidence suggest that the unpredicted exam retakes of the friends will have a negative effect on the changes of the performance of students. This effect is more prominent for students with own retakes and for students in the Economics department. The higher significance of the estimators in the model with correlated effects gives evidence of the presence of unobserved homophily that influences link formation. Change of sign of endogenous effect for students with own retakes shows the importance of further exploration of the problem and improvement of the model by inclusion of the link formation mechanism.

## References


Ammermueller, A. and J.-S. Pischke (2009). "Peer effects in European primary schools: Evidence from the progress in international reading literacy study". In: *Journal of Labor Economics* 27.3, pp. 315–348.

Androushchak, Gregory, Oleg Poldin, and Maria Yudkevich (2013). "Role of peers in student academic achievement in exogenously formed university groups". In: *Educational Studies* 39.5, pp. 568–581.

Angrist, J.D. and K. Lang (2004). "Does school integration generate peer effects? Evidence from Boston's Metco Program". In: *American Economic Review*, pp. 1613–1634.

Aral, S., L. Muchnik, and A. Sundararajan (2009). "Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks". In: *Proceedings of the National Academy of Sciences* 106.51, pp. 21544–21549.

*AUSTRALIA: High drop-out rates cost $1.4 billion* (2010). URL: http://www.universityworldnews.com/article.php?story=20101022203542738.

Battin-Pearson, S., M.D. Newcomb, R.D. Abbott, K.G. Hill, R.F. Catalano, and J.D. Hawkins (2000). "Predictors of early high school dropout: A test of five theories." In: *Journal of educational psychology* 92.3, pp. 568–582.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009). "Identification of peer effects through social networks". In: *Journal of econometrics* 150.1, pp. 41–55.

Buchmann, Claudia and Ben Dalton (2002). "Interpersonal influences and educational aspirations in 12 countries: The importance of institutional context". In: *Sociology of education* 75, pp. 99–122.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009). "Peer effects and social networks in education". In: *The Review of Economic Studies* 76.4, pp. 1239–1267.

Carrell, S.E., R.L. Fullerton, and J.E. West (2009). "Does your cohort matter? Measuring peer effects in college achievement". In: *Journal of Labor Economics* 27.3, pp. 439–464.

Chandrasekhar, A. and R. Lewis (2011). "Econometrics of sampled networks". In: *Unpublished manuscript, MIT.[422].*

Coleman, J.S. (1990). *Foundations of social theory.* Harvard university press.

Comola, M. and S. Prina (2014). "Do interventions change the network? A dynamic peer effect model accounting for network changes". In: *Working paper.*

De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli (2010). "Identification of social interactions through partially overlapping peer groups". In: *American Economic Journal: Applied Economics*, pp. 241–275.

Elhorst, J. Paul (2010). "Dynamic panels with endogenous interaction effects when T is small". In: *Regional Science and Urban Economics* 40.5, pp. 272 –282.

French, D.C. and J. Conrad (2001). "School dropout as predicted by peer rejection and antisocial behavior". In: *Journal of Research on adolescence* 11.3, pp. 225–244.

Gardner, Michael (2007). *GERMANY: Heavy cost of student drop-outs.* URL: http://www.universityworldnews.com/article.php?story=20071025102357719.

Garnier, H.E., J.A. Stein, and J.K. Jacobs (1997). "The process of dropping out of high school: A 19-year perspective". In: *American Educational Research Journal* 34.2, pp. 395–419.

Gaviria, A. and S. Raphael (2001). "School-based peer effects and juvenile behavior". In: *Review of Economics and Statistics* 83.2, pp. 257–268.

Goldsmith-Pinkham, P. and Guido W Imbens (2013). “Social networks and the identification of peer effects”. In: *Journal of Business & Economic Statistics* 31.3, pp. 253–264.

Hanushek, E.A., J.F. Kain, J.M. Markman, and S.G. Rivkin (2003). “Does peer ability affect student achievement?” In: *Journal of applied econometrics* 18.5, pp. 527–544.

Hoxby, C. (2000). *Peer effects in the classroom: Learning from gender and race variation.* Tech. rep. National Bureau of Economic Research.

*Indicators of economics of education,* URL: `https://memo.hse.ru/ind`.

Jackson, Matthew O. (2010). “An overview of social networks and economic applications”. In: *The handbook of social economics* 1, pp. 511–585.

Kelejian, Harry H and Ingmar R Prucha (1998). “A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances”. In: *The Journal of Real Estate Finance and Economics* 17.1, pp. 99–121.

Lee, Lung-fei (2003). “Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances”. In: *Econometric Reviews* 22.4, pp. 307–335.

Lomi, A., Tom A.B. Snijders, Christian E.G. Steglich, and V.J. Torl (2011). “Why are some more peer than others? Evidence from a longitudinal study of social networks and individual academic performance”. In: *Social Science Research* 40.6, pp. 1506 – 1520.

Manski, C.F. (1993). “Identification of endogenous social effects: The reflection problem”. In: *The review of economic studies* 60.3, pp. 531–542.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). “Birds of a feather: Homophily in social networks”. In: *Annual review of sociology*, pp. 415–444.

*Metco program official webpage.* URL: `http://www.doe.mass.edu/metco/`.

Newcomb, M.D. (1996a). “Adolescence: Pathologizing a normal process.” In: *The counseling Psychologist* 24, pp. 482–490.

– (1996b). “Pseudomaturity among adolescents: Construct validation, sex differences, and associations in adulthood.” In: *Journal of Drug Issues* 26, pp. 477–504.

Paula, Áureo de (2015). “Econometrics of Network Models”. In: *cemmap Working paper* CWP52/15.

Poldin, Oleg, Diliara Valeeva, and Maria Yudkevich (2015). “Which Peers Matter: How Social Ties Affect Peer-group Effects”. In: *Research in Higher Education*, pp. 1–21.

Rumberger, R.W. (1983). “Dropping out of high school: The influence of race, sex, and family background”. In: *American Educational Research Journal.*

Sacerdote, B. (2011). "Peer effects in education: How might they work, how big are they and how much do we know thus far?" In: *Handbook of the Economics of Education* 3, pp. 249–277.

Shugal, N. (2010). "Potoki obuchayushchikhsya v rossiyskoy sisteme obrazovaniya [Student flows in Russian education system]". In: *Voprosy obrazovaniya / Educational Studies* 4, pp. 122–148.

Smith, J.P. and R.A. Naylor (2001). "Dropping Out of University: A Statistical Analysis of the Probability of Withdrawal for UK University Students". In: *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 164.2, pp. 389–405.

*Socio-Economic situation in Russia* (2014). URL: http://www.gks.ru/bgd/regl/b14_01/Main.htm.

Stinebrickner, T. and R. Stinebrickner (2013). *Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model.* Tech. rep. National Bureau of Economic Research.

Tinto, V. (1975). "Dropout from higher education: A theoretical synthesis of recent research". In: *Review of educational research*, pp. 89–125.

Triandis, Harry C (1989). "The self and social behavior in differing cultural contexts." In: *Psychological review* 96.3, pp. 506–520.

# Appendix

## A. Main proofs

**Regularity conditions** (adaptation of Lee (2003)):

**Assumption 1.** The matrices $(I - \beta^1 G^1)$ and $(I - \beta_2 G^2)$ are nonsingular

**Assumption 2.** The row and column sums of the matrices $G^1$, $G^2$, $(I - \beta^1 G^1)^{-1}$ and $(I - \beta_2 G^2)^{-1}$ are uniformly bounded in absolute value.

**Assumption 3.** The elements of the matrices $X^1$ and $X^2$ are uniformly bounded in absolute value

**Assumption 4.** The error terms $\{\nu_i : 1 \leq i \leq n\}$ are identically distributed. Furthermore, they are distributed (jointly) independently with $\mathbb{E}[\nu_i \boldsymbol{X_i^1}] = 0$ and $\mathbb{E}[\nu_i^2] = \sigma_\nu < \infty$. Additionally, they are assumed to possess finite fourth moments. The error terms $\{\Delta\epsilon_i : 1 \leq i \leq n\}$ are identically distributed. Furthermore, they are distributed (jointly) independently with $\mathbb{E}[\Delta\epsilon_i] = 0$ and $\mathbb{E}[\Delta\epsilon_i^2] = \sigma_{\epsilon^1} + \sigma_{\epsilon^2} < \infty$. Additionally, they are assumed to possess finite fourth moments

**Assumption 5.** The limit $\boldsymbol{J} = \lim_{n\to\infty} \frac{1}{n}\boldsymbol{Z}^T\boldsymbol{Z}$ exists and is nonsingular.

**Assumption 6.** The limit $\bar{\boldsymbol{J}} = \lim_{n\to\infty} \frac{1}{n}\bar{\boldsymbol{Z}}^T\bar{\boldsymbol{Z}}$ exists and is nonsingular.

**Assumption 7. Step 1.** The initial estimator $\beta^1_{2SLS}$ of $\beta_1$ is $n^a$-consistent for some $a > 0$. The initial estimators $\alpha^1_{2SLS}$, $\gamma^1_{2SLS}$ and $\delta^1_{2SLS}$ are consistent estimators of $\alpha^1$, $\gamma^1$ and $\delta^1$, respectively. **Step 2.** The initial estimators $\beta_{1,2SLS}$ and $\beta_{2,2SLS}$ of $\beta_1$ and $\beta_2$ are $n^b$-consistent for some $b > 0$. The initial estimators $\alpha_{1,2SLS}$, $\alpha_{2,2SLS}$, $\gamma_{1,2SLS}$, $\gamma_{2,2SLS}$, $\delta_{1,2SLS}$ and $\delta_{2,2SLS}$ are consistent estimators of $\alpha_1$, $\alpha_2$, $\gamma_1$, $\gamma_2$, $\delta_1$ and $\delta_2$, respectively.

### Proof of Lemma 1.

The structural form equation:

$$P(retake_i) = \alpha^1 + \beta^1 \sum_{j\neq i} G^1_{ij} y^1_j + \gamma^1 X^1_i + \delta^1 \sum_{j\neq i} G^1_{ij} X^1_j + \nu_i, \qquad \mathbb{E}[\nu_i|X] = 0$$

can be rewritten in the reduced form in the following manner:

$$\boldsymbol{PR} = \alpha^1 \boldsymbol{i} + \beta^1 \boldsymbol{G^1 y^1} + \gamma^1 \boldsymbol{X^1} + \delta^1 \boldsymbol{G^1 X^1} + \nu, \qquad \mathbb{E}[\nu|\boldsymbol{X^1}] = 0$$

$$\boldsymbol{PR} = \alpha^1 \boldsymbol{i} + \beta^1 \boldsymbol{G^1 y^1} + (\gamma^1 \boldsymbol{I} + \delta^1 \boldsymbol{G^1})\boldsymbol{X^1} + \nu, \qquad \mathbb{E}[\nu|\boldsymbol{X^1}] = 0$$

Taking conditional expectations:

$$\mathbb{E}[\boldsymbol{PR}|\boldsymbol{X^1}] = \alpha^1 \boldsymbol{i} + \beta^1 \boldsymbol{G^1}\mathbb{E}[\boldsymbol{y^1}|\boldsymbol{X^1}] + (\gamma^1 \boldsymbol{I} + \delta^1 \boldsymbol{G^1})\boldsymbol{X^1}$$

Note that $\boldsymbol{y}$ can be expressed in terms of peer effect model as the one used for the probability of retakes:

$$y_i^1 = \alpha_0 + \beta_0 \sum_{j\neq i} G_{ij}^1 y_j^1 + \gamma_0 X_i^1 + \delta_0 \sum_{j\neq i} G_{ij}^1 X_j^1 + \xi_i, \qquad \mathbb{E}[\xi_i|X] = 0$$

with reduced form:

$$\boldsymbol{y^1} = \alpha_0 \boldsymbol{i} + \beta_0 \boldsymbol{G^1 y^1} + (\gamma_0 \boldsymbol{I} + \delta_0 \boldsymbol{G^1})\boldsymbol{X^1} + \xi, \qquad \mathbb{E}[\xi|\boldsymbol{X}] = 0$$

Then following steps of Bramoullé et al. (2009):

$$\boldsymbol{y^1} = \alpha_0(\boldsymbol{I} - \beta_0\boldsymbol{G^1})^{-1} + (\boldsymbol{I} - \beta_0\boldsymbol{G^1})^{-1}(\gamma_0\boldsymbol{I} + \delta_0\boldsymbol{G^1})\boldsymbol{X^1} + (\boldsymbol{I} - \beta_0\boldsymbol{G^1})^{-1}\xi, \qquad \mathbb{E}[\xi|\boldsymbol{X}] = 0$$

Using $(\boldsymbol{I} - \beta_0\boldsymbol{G^1})^{-1} = \sum_{k=0}^{\infty} \beta_0^k (\boldsymbol{G^1})^k$:

$$\boldsymbol{y^1} = \alpha_0(\boldsymbol{I} - \beta_0\boldsymbol{G^1})^{-1} + \gamma_0\boldsymbol{X^1} + (\gamma_0\beta_0 + \delta_0)\sum_{k=0}^{\infty} \beta_0^k (\boldsymbol{G^1})^{k+1}\boldsymbol{X^1} + \sum_{k=0}^{\infty} \beta_0^k (\boldsymbol{G^1})^k \xi$$

And the expected mean friends' groups' performance conditional on $\boldsymbol{X^1}$ can be written as:

$$\mathbb{E}[\boldsymbol{G^1 y^1}|\boldsymbol{X^1}] = \alpha_0(\boldsymbol{I} - \beta_0\boldsymbol{G^1})^{-1} + \gamma_0\boldsymbol{G^1 X^1} + (\gamma_0\beta_0 + \delta_0)\sum_{k=0}^{\infty} \beta_0^k (\boldsymbol{G^1})^{k+2}\boldsymbol{X^1}$$

As was proven in Bramoullé et al. (2009), if $\gamma_0\beta_0 + \delta_0 \neq 0$ and $I, G^1$ and $(G^1)^2$ are linearly independent, the social effects are identified. So this expression can be plugged-in into the reduced form of the equation for the probability of retake.

$$\begin{aligned}
&\mathbb{E}[\boldsymbol{PR}|\boldsymbol{X^1}] = \alpha^1 \boldsymbol{i} + \beta^1(\alpha_0(\boldsymbol{I} - \beta_0\boldsymbol{G^1})^{-1} + \gamma_0\boldsymbol{G^1 X^1} + (\gamma_0\beta_0 + \delta_0)\sum_{k=0}^{\infty} \beta_0^k (\boldsymbol{G^1})^{k+2}\boldsymbol{X^1}) + \\
&+(\gamma^1\boldsymbol{I} + \delta^1\boldsymbol{G^1})\boldsymbol{X^1} = \\
&= (\alpha^1\boldsymbol{I} + \beta^1\alpha_0(\boldsymbol{I} - \beta_0\boldsymbol{G^1})^{-1})) + \beta^1(\gamma_0\beta_0 + \delta_0)\sum_{k=0}^{\infty} \beta_0^k (\boldsymbol{G^1})^{k+2}\boldsymbol{X^1} + (\gamma^1\boldsymbol{I} + (\beta^1\gamma_0 + \delta^1)\boldsymbol{G^1})\boldsymbol{X^1}
\end{aligned}$$

or

$$\mathbb{E}[\boldsymbol{PR}|\boldsymbol{X^1}] = \alpha^1\boldsymbol{I} + \beta^1(\alpha_0(\boldsymbol{I} - \beta_0\boldsymbol{G^1})^{-1} + \beta^1(\boldsymbol{I} - \beta_0\boldsymbol{G^1})^{-1}(\gamma_0\boldsymbol{I} + \delta_0\boldsymbol{G^1})\boldsymbol{G^1 X^1} + (\gamma^1\boldsymbol{I} + \delta^1\boldsymbol{G^1})\boldsymbol{X^1}$$

Now consider two sets of structural parameters $(\alpha^1, \beta^1, \gamma^1, \delta^1)$ and $(\tilde{\alpha}^1, \tilde{\beta}^1, \tilde{\gamma}^1, \tilde{\delta}^1)$ leading to the same reduced form. It means that:

$$\alpha^1\boldsymbol{I} + \beta^1\alpha_0(\boldsymbol{I} - \beta_0\boldsymbol{G^1})^{-1} = \tilde{\alpha}^1\boldsymbol{I} + \tilde{\beta}^1\alpha_0(\boldsymbol{I} - \beta_0\boldsymbol{G^1})^{-1}$$

$$\alpha^1\boldsymbol{I} - \alpha^1\beta_0\boldsymbol{G^1} + \beta^1\alpha_0\boldsymbol{I} = \tilde{\alpha}^1\boldsymbol{I} - \tilde{\alpha}^1\beta_0\boldsymbol{G^1} + \tilde{\beta}^1\alpha_0\boldsymbol{I}$$

$$(\alpha^1 - \tilde{\alpha}^1)\boldsymbol{I} + (\beta^1\alpha_0 - \tilde{\beta}^1\alpha_0)\boldsymbol{I} - (\alpha^1\beta_0 - \tilde{\alpha}^1\beta_0)\boldsymbol{G^1} = 0$$
$$(\alpha^1 - \tilde{\alpha}^1 + (\beta^1 - \tilde{\beta}^1)\alpha_0)\boldsymbol{I} = (\alpha^1 - \tilde{\alpha}^1)\beta_0\boldsymbol{G^1}$$

and:

$$\beta^1(\boldsymbol{I}-\beta_0\boldsymbol{G^1})^{-1}(\gamma_0\boldsymbol{I}+\delta_0\boldsymbol{G^1})\boldsymbol{G^1}+(\gamma^1\boldsymbol{I}+\delta^1\boldsymbol{G^1}) = \tilde{\beta}^1(\boldsymbol{I}-\beta_0\boldsymbol{G^1})^{-1}(\gamma_0\boldsymbol{I}+\delta_0\boldsymbol{G^1})\boldsymbol{G^1}+(\tilde{\gamma}^1\boldsymbol{I}+\tilde{\delta}^1\boldsymbol{G^1})$$

$$\beta^1(\gamma_0\boldsymbol{I}+\delta_0\boldsymbol{G^1})\boldsymbol{G^1}+(\boldsymbol{I}-\beta_0\boldsymbol{G^1})(\gamma^1\boldsymbol{I}+\delta^1\boldsymbol{G^1}) = \tilde{\beta}^1(\gamma_0\boldsymbol{I}+\delta_0\boldsymbol{G^1})\boldsymbol{G^1}+(\boldsymbol{I}-\beta_0\boldsymbol{G^1})(\tilde{\gamma}^1\boldsymbol{I}+\tilde{\delta}^1\boldsymbol{G^1})$$

$$\beta^1\gamma_0\boldsymbol{G^1} + \beta^1\delta_0(\boldsymbol{G^1})^2 + (\gamma^1\boldsymbol{I} - (\beta_0\gamma^1 - \delta^1)\boldsymbol{G^1}) - \beta_0\delta^1\boldsymbol{(G^1)}^2) = \tilde{\beta}^1\gamma_0\boldsymbol{G^1} + \tilde{\beta}^1\delta_0(\boldsymbol{G^1}^2) +$$
$$+(\tilde{\gamma}^1\boldsymbol{I} - (\beta_0\tilde{\gamma}^1 - \tilde{\delta}^1)\boldsymbol{G^1}) - \beta_0\tilde{\delta}^1\boldsymbol{(G^1)}^2)$$

$$\gamma^1\boldsymbol{I}+(\beta^1\gamma_0+\beta^1\delta_0-\beta_0\gamma^1+\delta^1)\boldsymbol{G^1}-\beta_0\delta^1\boldsymbol{(G^1)}^2 = \tilde{\gamma}^1\boldsymbol{I}+(\tilde{\beta}^1\gamma_0+\tilde{\beta}^1\delta_0-\beta_0\tilde{\gamma}^1+\tilde{\delta}^1)\boldsymbol{G^1}-\beta_0\tilde{\delta}^1\boldsymbol{(G^1)}^2$$

$$(\gamma^1-\tilde{\gamma}^1)\boldsymbol{I}+((\beta^1-\tilde{\beta}^1)\gamma_0+(\beta^1-\tilde{\beta}^1)\delta_0-\beta_0(\gamma^1-\tilde{\gamma}^1)+\delta^1-\tilde{\delta}^1)\boldsymbol{G^1}+\beta_0(\tilde{\delta}^1-\delta^1)\boldsymbol{(G^1)}^2 = 0$$

Now let $I, G^1$ and $(G^1)^2$ be linearly independent. Then the above equality holds only if all three coefficients are 0:

$$\begin{aligned}
&\gamma^1 - \tilde{\gamma}^1 = 0\\
&(\beta^1 - \tilde{\beta}^1)\gamma_0 + (\beta^1 - \tilde{\beta}^1)\delta_0 - \beta_0(\gamma^1 - \tilde{\gamma}^1) + \delta^1 - \tilde{\delta}^1 = 0\\
&\beta_0(\tilde{\delta}^1 - \delta^1) = 0
\end{aligned}$$

If $\beta_0 \neq 0$ and $\gamma_0^2 + \delta_0^2 \neq 0$, two sets of coefficients $(\alpha^1, \beta^1, \gamma^1, \delta^1)$ and $(\tilde{\alpha}^1, \tilde{\beta}^1, \tilde{\gamma}^1, \tilde{\delta}^1)$ are equivalent. Note that the restrictions on the coefficients of the peer effect model suggest that the model has an endogenous peer effect and the performance depends on own set of observed characteristics, or on peers observed characteristics, or on both. These requirements are natural for the peer effect model and therefore, the identification result is achieved. ■

**Proof of Lemma 2. (Identification, Step 2, no correlated effects)**

Recall the second step equation:

$$\Delta y_i = (\alpha_2 - \alpha_1) + \beta_2\sum_{j\neq i} G^2_{ij}y^2_j - \beta_1\sum_{j\neq i} G^1_{ij}y^1_j + \tilde{\boldsymbol{\delta}}UR_i + \gamma_2 X^2_i - \gamma_1 X^1_i +$$
$$+\delta_2\sum_{j\neq i} G^2_{ij}X^2_j - \delta_1\sum_{j\neq i} G^1_{ij}X^1_j + \Delta\epsilon_i$$

It can be rewritten in the reduced form as following:

$$\Delta\boldsymbol{y} = (\alpha_2 - \alpha_1)\boldsymbol{i} + \beta_2\boldsymbol{G}^2\boldsymbol{y^2} - \beta_1\boldsymbol{G^1}\boldsymbol{y^1} + \tilde{\boldsymbol{\delta}}\boldsymbol{UR} + \gamma_2\boldsymbol{X^2_{TV}} - \gamma_1\boldsymbol{X^1_{TV}} + \delta_2\boldsymbol{G^2}\boldsymbol{X^2} -$$
$$-\delta_1\boldsymbol{G^1}\boldsymbol{X^1} + \Delta\epsilon, \qquad \text{with } \boldsymbol{UR} \text{ defined as discussed in Section 2 and } \mathbb{E}[\Delta\epsilon] = 0$$

This can be further modified in the following manner:

$$\mathbb{E}[\Delta \boldsymbol{y}|\boldsymbol{X^2}] = (\alpha_2 - \alpha_1)\boldsymbol{i} + \beta_2 \boldsymbol{G^2}\mathbb{E}[\boldsymbol{y^2}|\boldsymbol{X^2}] - \beta_1 \boldsymbol{G^1}\mathbb{E}[\boldsymbol{y^1}|\boldsymbol{X^2}] + \tilde{\boldsymbol{\delta}}\mathbb{E}[\boldsymbol{UR}|\boldsymbol{X^2}] +$$

$$+\gamma_2 \boldsymbol{X^2_{TV}} - \gamma_1 \boldsymbol{X^1_{TV}} + \delta_2 \boldsymbol{G^2}\boldsymbol{X^2} - \delta_1 \boldsymbol{G^1}\boldsymbol{X^1}$$

with

$$\mathbb{E}[\boldsymbol{y^1}|\boldsymbol{X^2}] = (\boldsymbol{I} - \beta_{0,1}\boldsymbol{G^1})^{-1}\alpha_{0,1} + (\boldsymbol{I} - \beta_{0,1}\boldsymbol{G^1})^{-1}(\gamma_{0,1}\boldsymbol{I} + \delta_{0,1}\boldsymbol{G^1})\mathbb{E}[\boldsymbol{X^1}|\boldsymbol{X^2}] =$$

$$= (\boldsymbol{I} - \beta_{0,1}\boldsymbol{G^1})^{-1}\alpha_{0,1} + (\boldsymbol{I} - \beta_{0,1}\boldsymbol{G^1})^{-1}(\gamma_{0,1}\boldsymbol{I} + \delta_{0,1}\boldsymbol{G^1})\boldsymbol{X^1},$$

since $\boldsymbol{X^1}$ is already known by the time $\boldsymbol{X^2}$ is revealed, therefore, the latter cannot add any new information.

Also:

$$\mathbb{E}[\boldsymbol{y^2}|\boldsymbol{X^2}] = (\boldsymbol{I} - \beta_{0,2}\boldsymbol{G^2})^{-1}\alpha_{0,2} + (\boldsymbol{I} - \beta_{0,2}\boldsymbol{G^2})^{-1}(\gamma_{0,2}\boldsymbol{I} + \delta_{0,2}\boldsymbol{G^2})\boldsymbol{X^2}$$

Note that $\boldsymbol{UR}$ is also defined at the first period, hence, the new information in $\boldsymbol{X^2}$ will not anything new for the expected value of the $\boldsymbol{UR}$, hence $\mathbb{E}[\boldsymbol{UR}|\boldsymbol{X^2}] = \boldsymbol{UR}$.

Also notice than in principle coefficients in the model in differences $\alpha_1, \beta_1, \gamma_1, \delta_1, \alpha_2, \beta_2, \gamma_2, \delta_2$ can be different from the corresponding coefficients in the single period peer effect models $\alpha_{0,1}, \beta_{0,1}, \gamma_{0,1}, \delta_{0,1}, \alpha_{0,2}, \beta_{0,2}, \gamma_{0,2}, \delta_{0,2}$. This can be due to the unaccounted in single period model fixed effects that can be eliminated in the model in differences and due to the presence of the shock in the model, which can take some of the effect, that would be otherwise attributed towards endogenous or exogenous effect.

Then, letting $\alpha = \alpha_2 - \alpha_1$

$$\mathbb{E}[\Delta \boldsymbol{y}|\boldsymbol{X^2}] = \alpha\boldsymbol{i} + \beta_2\boldsymbol{G^2}(\boldsymbol{I} - \beta_{0,2}\boldsymbol{G^2})^{-1}(\alpha_{0,2} + (\gamma_{0,2}\boldsymbol{I} + \delta_{0,2}\boldsymbol{G^2})\boldsymbol{X^2}) - \beta_1\boldsymbol{G^1}(\boldsymbol{I} - \beta_{0,1}\boldsymbol{G^1})^{-1}(\alpha_{0,1} +$$

$$(\gamma_{0,1}\boldsymbol{I} + \delta_{0,1}\boldsymbol{G^1})\boldsymbol{X^1}) + \tilde{\boldsymbol{\delta}}\boldsymbol{UR} + \gamma_2\boldsymbol{X^2_{TV}} - \gamma_1\boldsymbol{X^1_{TV}} + \delta_2\boldsymbol{G^2}\boldsymbol{X^2} - \delta_1\boldsymbol{G^1}\boldsymbol{X^1}$$

**First,** if $\boldsymbol{G^1} = \boldsymbol{G^1}$, then $\delta_2$ and $\delta_1$ are identified only partially, for time-variant variables of $\boldsymbol{X^1}$ and $\boldsymbol{X^2}$ respectively. This assumption can be relaxed, if we let the coefficients of the single period coincide with the coefficients of the coefficients of the model in differences. Then, however, the following assumption need to be made $\tilde{\delta} = 0$, meaning that the shock has no effect on the outcome, which is not true in the setting of the model of the paper. Hence, $\boldsymbol{G^1} = \boldsymbol{G^1}$ is one of the identifying assumptions for the second step model.

**Next,** I follow similar steps to the proof of Lemma 1. Consider two sets of the parameters leading to the same reduced form, $(\alpha_1, \beta_1, \gamma_1, \delta_1, \alpha_2, \beta_2, \gamma_2, \delta_2, \tilde{\delta})$ and $(\tilde{\alpha_1}, \tilde{\beta_1}, \tilde{\gamma_1}, \tilde{\delta_1}, \tilde{\alpha_2}, \tilde{\beta_2}, \tilde{\gamma_2}, \tilde{\delta_2}, \tilde{\tilde{\delta}})$. I do not include the single-period parameters, since their identification is achieved separately, if $\boldsymbol{I}, \boldsymbol{G^1}, (\boldsymbol{G^1})^2$ are linearly independent and if $\boldsymbol{I}, \boldsymbol{G^2}, (\boldsymbol{G^2})^2$ are also linearly independent. Then:

$$\alpha\boldsymbol{I} + \beta_2\boldsymbol{G^2}(\boldsymbol{I} - \beta_{0,2}\boldsymbol{G^2})^{-1}\alpha_{0,2} - \beta_1\boldsymbol{G^1}(\boldsymbol{I} - \beta_{0,1}\boldsymbol{G^1})^{-1}\alpha_{0,1} =$$

$$= \tilde{\alpha}\boldsymbol{I} + \tilde{\beta}_2\boldsymbol{G^2}(\boldsymbol{I} - \beta_{0,2}\boldsymbol{G^2})^{-1}\alpha_{0,2} - \tilde{\beta}_1\boldsymbol{G^1}(\boldsymbol{I} - \beta_{0,1}\boldsymbol{G^1})^{-1}\alpha_{0,1}$$

$$\beta_2\boldsymbol{G^2}(\boldsymbol{I} - \beta_{0,2}\boldsymbol{G^2})^{-1}(\gamma_{0,2}\boldsymbol{I} + \delta_{0,2}\boldsymbol{G^2}) + (\gamma_2\boldsymbol{I} + \delta_2\boldsymbol{G^2}) =$$

$$= \tilde{\beta}_2\boldsymbol{G^2}(\boldsymbol{I} - \beta_{0,2}\boldsymbol{G^2})^{-1}(\gamma_{0,2}\boldsymbol{I} + \delta_{0,2}\boldsymbol{G^2}) + (\tilde{\gamma}_2\boldsymbol{I} + \tilde{\delta}_2\boldsymbol{G^2})$$

$$\beta_1\boldsymbol{G^1}(\boldsymbol{I} - \beta_{0,1}\boldsymbol{G^1})^{-1}(\gamma_{0,1}\boldsymbol{I} + \delta_{0,1}\boldsymbol{G^1}) + (\gamma_1\boldsymbol{I} + \delta_1\boldsymbol{G^1}) =$$

$$= \tilde{\beta}_1\boldsymbol{G^1}(\boldsymbol{I} - \beta_{0,1}\boldsymbol{G^1})^{-1}(\gamma_{0,1}\boldsymbol{I} + \delta_{0,1}\boldsymbol{G^1}) + (\tilde{\gamma}_1\boldsymbol{I} + \tilde{\delta}_1\boldsymbol{G^1})$$

$$\tilde{\delta} = \tilde{\tilde{\delta}}$$

Note, that I added time invariant own exogenous variables to the vectors $\boldsymbol{X^1_{TV}}$ and $\boldsymbol{X^2_{TV}}$. Since they are not in the model, zeros are assumed on the additional elements of $\gamma_1$ and $\gamma_2$.

The third equation can be further simplified as following:

$$\gamma_1\boldsymbol{I} + (\delta_1 - \gamma_1\beta_{0,1} - \beta_1\gamma_{0,1})\boldsymbol{G^1} + (\beta_1\delta_{0,1} - \delta_1\beta_{0,1})\boldsymbol{G^1})^2 =$$

$$= \tilde{\gamma}_1\boldsymbol{I} + (\tilde{\delta}_1 - \tilde{\gamma}_1\beta_{0,1} - \tilde{\beta}_1\gamma_{0,1})\boldsymbol{G^1} + (\tilde{\beta}_1\delta_{0,1} - \tilde{\delta}_1\beta_{0,1})\boldsymbol{G^1})^2$$

Then, if $\boldsymbol{I}, \boldsymbol{G^1}, (\boldsymbol{G^1})^2$ are linearly independent, the coefficients in front of these three matrices are 0:

$$\begin{aligned}
&\gamma_1 - \tilde{\gamma}_1 = 0 \\
&\delta_1 - \gamma_1\beta_{0,1} - \beta_1\gamma_{0,1} = \tilde{\delta}_1 - \tilde{\gamma}_1\beta_{0,1} - \tilde{\beta}_1\gamma_{0,1}, \text{ or} \\
&(\delta_1 - \tilde{\delta}_1) - (\gamma_1 - \tilde{\gamma}_1)\beta_{0,1} + (\beta_1 - \tilde{\beta}_1)\gamma_{0,1} = 0 \\
&\beta_1\delta_{0,1} - \delta_1\beta_{0,1} = \tilde{\beta}_1\delta_{0,1} - \tilde{\delta}_1\beta_{0,1}, \text{ or} \\
&(\beta_1 - \tilde{\beta}_1)\delta_{0,1} - (\delta_1 - \tilde{\delta}_1)\beta_{0,1} = 0
\end{aligned}$$

Now, if $\beta_{0,1} \neq 0$ and $\gamma_{0,1}^2 + \delta_{0,1}^2 \neq 0$, the two sets of the coefficients, $(\gamma_1, \delta_1, \beta_1)$ and $(\tilde{\gamma}_1, \tilde{\delta}_1, \tilde{\beta}_1)$, coincide.

Similar argument is valid for the coefficient in front of $\boldsymbol{X^2}$, hence $(\gamma_2, \delta_2, \beta_2)$ and $(\tilde{\gamma}_2, \tilde{\delta}_2, \tilde{\beta}_2)$, also coincide, when $\boldsymbol{I}, \boldsymbol{G^2}, (\boldsymbol{G^2})^2$ are linearly independent and $\beta_{0,2} \neq 0$ and $\gamma_{0,2}^2 + \delta_{0,2}^2 \neq 0$.

The other two equalities lead then automatically to $\alpha = \tilde{\alpha}$ and $\tilde{\delta} = \tilde{\tilde{\delta}}$ without any additional assumptions. Hence, the identification is achieved under the conditions of linear independence of $\boldsymbol{I}, \boldsymbol{G^1}, (\boldsymbol{G^1})^2$ and $\boldsymbol{I}, \boldsymbol{G^2}, (\boldsymbol{G^2})^2$ and $\boldsymbol{G^1} \neq \boldsymbol{G^2}$ and mentioned assumptions on the coefficients. ■

**Proof of Lemma 3.**

The structural form equation:

$$P(retake_i) - \sum_{j\neq i} G^1_{ij} P(retake_j) = \beta \sum_{j\neq i} G^1_{ij}[y^1_j - \sum_{k\neq j} G^1_{jk} y^1_k] + \gamma[X^1_i - \sum_{j\neq i} G^1_{ij} X^1_j] +$$
$$+\delta \sum_{j\neq i} G^1_{ij}[X^1_j - \sum_{k\neq j} G^1_{jk} X^1_k] + [\eta_i - \sum_{j\neq i} G^1_{ij}\eta_j], \qquad \mathbb{E}[\eta_i|X^1] = 0$$

can be rewritten in the reduced form in the following manner:

$$(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{PR} = \beta(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{G^1}\boldsymbol{y^1}+\gamma(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{X^1}+\delta(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{G^1}\boldsymbol{X^1}+(\boldsymbol{I}-\boldsymbol{G^1})\eta,\ \mathbb{E}[\eta|\boldsymbol{X^1}] = 0$$

$$(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{PR} = \beta(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{G^1}\boldsymbol{y^1} + (\gamma\boldsymbol{I}+\delta\boldsymbol{G^1})(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{X^1} + (\boldsymbol{I}-\boldsymbol{G^1})\eta, \qquad \mathbb{E}[\eta|\boldsymbol{X^1}] = 0$$

Taking conditional expectations:

$$\mathbb{E}[(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{PR}|\boldsymbol{X^1}] = \beta(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{G^1}\mathbb{E}[\boldsymbol{y^1}|\boldsymbol{X^1}] + (\gamma\boldsymbol{I}+\delta\boldsymbol{G^1})(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{X^1}$$

Note that $\boldsymbol{y}$ can be expressed in terms of peer effect model as the one used for the probability of retakes:

$$y^1_i - \sum_{j\neq i} G^1_{ij} y^1_j = \beta_0 \sum_{j\neq i} G^1_{ij}[y^1_j - \sum_{k\neq j} G^1_{jk} y^1_k] + \gamma_0[X^1_i - \sum_{j\neq i} G^1_{ij} X^1_j] + \delta_0 \sum_{j\neq i} G^1_{ij}[X^1_j -$$
$$- \sum_{k\neq j} G^1_{jk} X^1_k] + [\xi_i - \sum_{j\neq i} G^1_{ij}\xi_j], \quad \mathbb{E}[\xi_i|X^1] = 0$$

with reduced form:

$$(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{y^1} = \beta_0(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{G^1}\boldsymbol{y^1} + (\gamma_0\boldsymbol{I}+\delta_0\boldsymbol{G^1})(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{X^1} + (\boldsymbol{I}-\boldsymbol{G^1})\xi, \quad \mathbb{E}[\xi|\boldsymbol{X^1}] = 0$$

Then following steps of Bramoullé et al. (2009):

$$(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{y^1} = (\boldsymbol{I}-\beta_0\boldsymbol{G^1})^{-1}(\gamma_0\boldsymbol{I}+\delta_0\boldsymbol{G^1})(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{X^1}+(\boldsymbol{I}-\beta_0\boldsymbol{G^1})^{-1}(\boldsymbol{I}-\boldsymbol{G^1})\xi, \quad \mathbb{E}[\xi|\boldsymbol{X}] = 0$$

And:

$$\mathbb{E}[(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{G^1}\boldsymbol{y^1}|\boldsymbol{X^1}] = (\boldsymbol{I}-\beta_0\boldsymbol{G^1})^{-1}(\gamma_0\boldsymbol{I}+\delta_0\boldsymbol{G^1})(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{G^1}\boldsymbol{X^1}$$

As was proven in Bramoullé et al. (2009), if $\gamma_0\beta_0 + \delta_0 \neq 0$ and $I, G^1, (G^1)^2$ and $(G^1)^3$ are linearly independent, the social effects are identified. So this expression can be plugged-in into the reduced form of the equation for the probability of retake.

$$\mathbb{E}[(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{PR}|\boldsymbol{X^1}] = \beta((\boldsymbol{I}-\beta_0\boldsymbol{G^1})^{-1}(\gamma_0\boldsymbol{I}+\delta_0\boldsymbol{G^1})(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{G^1})\boldsymbol{X^1}+(\gamma\boldsymbol{I}+\delta\boldsymbol{G^1})(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{X^1}$$

Now consider two sets of structural parameters $(\beta, \gamma, \delta)$ and $(\tilde{\beta}, \tilde{\gamma}, \tilde{\delta})$ leading to the same reduced form. It means that:

$$\beta(\boldsymbol{I}-\beta_0\boldsymbol{G^1})^{-1}(\gamma_0\boldsymbol{I}+\delta_0\boldsymbol{G^1})(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{G^1}+(\gamma\boldsymbol{I}+\delta\boldsymbol{G^1})(\boldsymbol{I}-\boldsymbol{G^1})=$$
$$=\tilde{\beta}(\boldsymbol{I}-\beta_0\boldsymbol{G^1})^{-1}(\gamma_0\boldsymbol{I}+\delta_0\boldsymbol{G^1})(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{G^1}+(\tilde{\gamma}\boldsymbol{I}+\tilde{\delta}\boldsymbol{G^1})(\boldsymbol{I}-\boldsymbol{G^1})$$

$$\beta(\gamma_0\boldsymbol{I}+\delta_0\boldsymbol{G^1})(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{G^1}+(\boldsymbol{I}-\beta_0\boldsymbol{G^1})(\gamma\boldsymbol{I}+\delta\boldsymbol{G^1})(\boldsymbol{I}-\boldsymbol{G^1})=$$
$$=\tilde{\beta}(\gamma_0\boldsymbol{I}+\delta_0\boldsymbol{G^1})(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{G^1}+(\boldsymbol{I}-\beta_0\boldsymbol{G^1})(\tilde{\gamma}\boldsymbol{I}+\tilde{\delta}\boldsymbol{G^1})(\boldsymbol{I}-\boldsymbol{G^1})$$

$$\beta\gamma_0\boldsymbol{G^1}+(\beta\delta_0-\beta\gamma_0)(\boldsymbol{G^1})^2-\beta\delta_0(\boldsymbol{G^1})^3+(\gamma\boldsymbol{I}-(\beta_0\gamma-\delta+\gamma)\boldsymbol{G^1})-$$
$$-(\beta_0\delta-\gamma\beta_0+\delta)(\boldsymbol{G^1})^2+\beta_0\delta(\boldsymbol{G^1})^3)=$$
$$=\tilde{\beta}\gamma_0\boldsymbol{G^1}+(\tilde{\beta}\delta_0-\tilde{\beta}\gamma_0)(\boldsymbol{G^1})^2-\tilde{\beta}\delta_0(\boldsymbol{G^1})^3+(\tilde{\gamma}\boldsymbol{I}-(\beta_0\tilde{\gamma}-\tilde{\delta}+\tilde{\gamma})\boldsymbol{G^1})-$$
$$-(\beta_0\tilde{\delta}-\tilde{\gamma}\beta_0+\tilde{\delta})(\boldsymbol{G^1})^2+\beta_0\tilde{\delta}(\boldsymbol{G^1})^3)$$

*I stopped here*

$$\gamma\boldsymbol{I}+(\beta\gamma_0-\beta_0\gamma+\delta-\gamma)\boldsymbol{G^1}+(\beta\delta_0-\beta\gamma_0-\beta_0\delta+\beta_0\gamma-\delta)\boldsymbol{(G^1)}^2+(\beta_0\delta-\beta\delta_0)\boldsymbol{(G^1)}^3=$$
$$=\tilde{\gamma}\boldsymbol{I}+(\tilde{\beta}\gamma_0-\beta_0\tilde{\gamma}+\tilde{\delta}-\tilde{\gamma})\boldsymbol{G^1}+(\tilde{\beta}\delta_0-\tilde{\beta}\gamma_0-\beta_0\tilde{\delta}+\beta_0\tilde{\gamma}-\tilde{\delta})\boldsymbol{(G^1)}^2+(\beta_0\tilde{\delta}-\tilde{\beta}\delta_0)\boldsymbol{(G^1)}^3$$

$$(\gamma^1-\tilde{\gamma}^1)\boldsymbol{I}+((\beta-\tilde{\beta})\gamma_0-(\gamma-\tilde{\gamma})\beta_0+(\delta-\tilde{\delta})-(\gamma-\tilde{\gamma}))\boldsymbol{G^1}+((\beta-\tilde{\beta})\delta_0-(\beta-\tilde{\beta})\gamma_0-$$
$$-(\delta-\tilde{\delta})\beta_0+(\gamma-\tilde{\gamma})\beta_0-(\delta-\tilde{\delta}))\boldsymbol{(G^1)}^2+((\delta-\tilde{\delta})\beta_0-(\beta-\tilde{\beta})\delta_0)\boldsymbol{(G^1)}^3=0$$

Now let $I, G^1$, $(G^1)^2$ and $(G^1)^3$ be linearly independent. Then the above equality holds only if all three coefficients are 0:

$$\begin{aligned}
&\gamma-\tilde{\gamma}=0\\
&(\beta-\tilde{\beta})\gamma_0-(\gamma-\tilde{\gamma})\beta_0+(\delta-\tilde{\delta})-(\gamma-\tilde{\gamma})=0\\
&(\beta-\tilde{\beta})\delta_0-(\beta-\tilde{\beta})\gamma_0-(\delta-\tilde{\delta})\beta_0+(\gamma-\tilde{\gamma})\beta_0-(\delta-\tilde{\delta})=0\\
&(\delta-\tilde{\delta})\beta_0-(\beta-\tilde{\beta})\delta_0=0
\end{aligned}$$

If $\beta_0\neq 0$ and $\gamma_0^2+\delta_0^2\neq 0$, two sets of coefficients $(\beta,\gamma,\delta)$ and $(\tilde{\beta},\tilde{\gamma},\tilde{\delta})$ are equivalent. Note that the restrictions on the coefficients of the peer effect model suggest that the model has an endogenous peer effect and the performance depends on own set of observed characteristics, or on peers observed characteristics, or on both. These requirements are natural for the peer effect model and therefore, the identification result is achieved. ■

**Proof of Lemma 4. (Identification, Step 2, correlated effects)**

The proof for Lemma 4 follows directly by applying similar arguments to the proofs of Lemma 2 and Lemma 3. Then, the identification is achieved under the conditions of linear independence of $\boldsymbol{I}, \boldsymbol{G^1}, (\boldsymbol{G^1})^2, (\boldsymbol{G^1})^3$ and $\boldsymbol{I}, \boldsymbol{G^2}, (\boldsymbol{G^2})^2, (\boldsymbol{G^2})^3$ and $\boldsymbol{G^1} \neq \boldsymbol{G^2}$ and the following assumptions on the coefficients: $\beta_{0,1} \neq 0$, $\gamma_{0,1}^2 + \delta_{0,1}^2 \neq 0$, $\beta_{0,2} \neq 0$ and $\gamma_{0,2}^2 + \delta_{0,2}^2 \neq 0$. ■

**Proof of Lemma 5. Consistency of $\hat{\theta}_{Lee}$ of Step 1**

$$\sqrt{n}(\hat{\theta}_{Lee} - \theta) = (\frac{1}{n}\hat{\boldsymbol{Z}}^T \tilde{\boldsymbol{X}}^{\boldsymbol{1}})^{-1} \frac{1}{\sqrt{n}}\hat{\boldsymbol{Z}}^T \boldsymbol{PR} - \sqrt{n}\theta = (\frac{1}{n}\hat{\boldsymbol{Z}}^T \tilde{\boldsymbol{X}}^{\boldsymbol{1}})^{-1}(\frac{1}{\sqrt{n}}\hat{\boldsymbol{Z}}^T \boldsymbol{PR} - \frac{1}{\sqrt{n}}\hat{\boldsymbol{Z}}^T \tilde{\boldsymbol{X}}^{\boldsymbol{1}}\theta)$$

Then we can rewrite the last term:

$$\frac{1}{\sqrt{n}}\hat{\boldsymbol{Z}}^T \boldsymbol{PR} - \frac{1}{\sqrt{n}}\hat{\boldsymbol{Z}}^T \tilde{\boldsymbol{X}}^{\boldsymbol{1}}\theta = \frac{1}{\sqrt{n}}\hat{\boldsymbol{Z}}^T (\boldsymbol{PR} - \tilde{\boldsymbol{X}}^{\boldsymbol{1}}\theta) = \frac{1}{\sqrt{n}}\hat{\boldsymbol{Z}}^T (\alpha \boldsymbol{i} + \beta \boldsymbol{G}^1 \boldsymbol{y}^1 + (\gamma \boldsymbol{I} + \delta \boldsymbol{G^1})\boldsymbol{X^1} + \nu -$$

$$-(\alpha \boldsymbol{i} + (\gamma \boldsymbol{I} + \delta \boldsymbol{G^1})\boldsymbol{X^1} + \beta \boldsymbol{G}^1 \boldsymbol{y}^1)) = \frac{1}{\sqrt{n}}\hat{\boldsymbol{Z}}^T \nu$$

Hence,

$$\sqrt{n}(\hat{\theta}_{Lee} - \theta) = (\frac{1}{n}\hat{\boldsymbol{Z}}^T \tilde{\boldsymbol{X}}^{\boldsymbol{1}})^{-1} \frac{1}{\sqrt{n}}\hat{\boldsymbol{Z}}^T \nu$$

Then the following two statements can be shown under the assumed regularity conditions and by direct application of Lemmas A.7, A.8 and A.9 in Lee (2003):

$$plim \frac{1}{n}\hat{\boldsymbol{Z}}^T \tilde{\boldsymbol{X}}^{\boldsymbol{1}} = plim \frac{1}{n}\boldsymbol{Z}^T \boldsymbol{Z} = \boldsymbol{J}$$

$$\frac{1}{\sqrt{n}}\hat{\boldsymbol{Z}}^T \nu \xrightarrow{D} \mathcal{N}(0, \sigma_\nu^2 \boldsymbol{J})$$

which will yield the desired result. ■

**Proof of Lemma 6. Consistency of $\hat{\phi}_{Lee}$ of Step 2**

$$\sqrt{n}(\hat{\phi}_{Lee} - \phi) = (\frac{1}{n}\hat{\bar{\boldsymbol{Z}}}^T \bar{\boldsymbol{X}})^{-1} \frac{1}{\sqrt{n}}\hat{\bar{\boldsymbol{Z}}}^T (\boldsymbol{y^2} - \boldsymbol{y^1}) - \sqrt{n}\phi = (\frac{1}{n}\hat{\bar{\boldsymbol{Z}}}^T \bar{\boldsymbol{X}})^{-1}(\frac{1}{\sqrt{n}}\hat{\bar{\boldsymbol{Z}}}^T (\boldsymbol{y^2} - \boldsymbol{y^1}) - \frac{1}{\sqrt{n}}\hat{\bar{\boldsymbol{Z}}}^T \bar{\boldsymbol{X}}\phi)$$

Then we can rewrite the last term:

$$\frac{1}{\sqrt{n}}\hat{\bar{\boldsymbol{Z}}}^T (\boldsymbol{y^2} - \boldsymbol{y^1}) - \frac{1}{\sqrt{n}}\hat{\bar{\boldsymbol{Z}}}^T \bar{\boldsymbol{X}}\phi = \frac{1}{\sqrt{n}}\hat{\bar{\boldsymbol{Z}}}^T (\boldsymbol{y^2} - \boldsymbol{y^1} - \bar{\boldsymbol{X}}\phi) = \frac{1}{\sqrt{n}}\hat{\bar{\boldsymbol{Z}}}^T ((\alpha_2 - \alpha_1)\boldsymbol{i} + \beta_2 \boldsymbol{G}^2 \boldsymbol{y^2} - \beta_1 \boldsymbol{G^1}\boldsymbol{y^1} +$$

$$+\tilde{\boldsymbol{\delta}}\boldsymbol{UR}+\gamma_2\boldsymbol{X^2_{TV}}-\gamma_1\boldsymbol{X^1_{TV}}+\delta_2\boldsymbol{G^2X^2}-\delta_1\boldsymbol{G^1X^1}+\Delta\epsilon-((\alpha_2-\alpha_1)\boldsymbol{i}+\beta_2\boldsymbol{G}^2\boldsymbol{y^2}-\beta_1\boldsymbol{G^1y^1}+\tilde{\boldsymbol{\delta}}\boldsymbol{UR}+$$

$$+\gamma_2\boldsymbol{X^2_{TV}}-\gamma_1\boldsymbol{X^1_{TV}}+\delta_2\boldsymbol{G^2X^2}-\delta_1\boldsymbol{G^1X^1})=\frac{1}{\sqrt{n}}\hat{\bar{\boldsymbol{Z}}}^T\Delta\epsilon$$

Hence,

$$\sqrt{n}(\hat{\phi}_{Lee}-\phi)=(\frac{1}{n}\hat{\bar{\boldsymbol{Z}}}^T\bar{\boldsymbol{X}})^{-1}\frac{1}{\sqrt{n}}\hat{\bar{\boldsymbol{Z}}}^T\Delta\epsilon$$

The following two statements have to hold to get the desired result:

$$plim\frac{1}{n}\hat{\bar{\boldsymbol{Z}}}^T\bar{\boldsymbol{X}}=plim\frac{1}{n}\bar{\boldsymbol{Z}}^T\bar{\boldsymbol{Z}}=\bar{\boldsymbol{J}}$$

$$\frac{1}{\sqrt{n}}\hat{\bar{\boldsymbol{Z}}}^T\Delta\epsilon\xrightarrow{D}\mathcal{N}(0,(\sigma^2_{\epsilon^1}+\sigma^2_{\epsilon^2})\bar{\boldsymbol{J}})$$

First, let's consider $\frac{1}{n}\hat{\bar{\boldsymbol{Z}}}^T\bar{\boldsymbol{X}}$. It is equivalent to $\frac{1}{n}[\boldsymbol{i},\boldsymbol{X^2_{TV}},\boldsymbol{X^1_{TV}},\boldsymbol{G^2X^2},\boldsymbol{G^1X^1},\boldsymbol{UR}$, $\mathbb{E}[\boldsymbol{G^1y^1}(\hat{\theta}^1_{2SLS})|\boldsymbol{X^1}],\mathbb{E}[\boldsymbol{G^2y^2}(\hat{\phi}_{2SLS})|\boldsymbol{X^2},\boldsymbol{X^1}]]^T[\boldsymbol{i},\boldsymbol{X^2_{TV}},\boldsymbol{X^1_{TV}},\boldsymbol{G^2X^2},\boldsymbol{G^1X^1},\boldsymbol{UR},\boldsymbol{G^1y^1}$, $\boldsymbol{G^2y^2}]$

First six rows do not consist any element of estimated vector of coefficients, and therefore, will not matter for the consistency argument.

Notice also that $\boldsymbol{G^1y^1}=\boldsymbol{G^1}(\boldsymbol{I}-\beta^1\boldsymbol{G^1})^{-1}\alpha^1+\boldsymbol{G^1}(\boldsymbol{I}-\beta^1\boldsymbol{G^1})^{-1}(\gamma^1\boldsymbol{I}+\delta^1\boldsymbol{G^1})\boldsymbol{X^1}+\boldsymbol{G^1}(\boldsymbol{I}-\beta^1\boldsymbol{G^1})^{-1}\epsilon_1$ and $\boldsymbol{G^2y^2}=\boldsymbol{G^2}(\boldsymbol{I}-\beta_2\boldsymbol{G^2})^{-1}[(\alpha_2-\alpha_1)\boldsymbol{i}+(\boldsymbol{I}-\beta_1\boldsymbol{G^1})((\boldsymbol{I}-\beta^1\boldsymbol{G^1})^{-1}\alpha^1+(\boldsymbol{I}-\beta^1\boldsymbol{G^1})^{-1}(\gamma^1\boldsymbol{I}+\delta^1\boldsymbol{G^1})\boldsymbol{X^1})+\tilde{\boldsymbol{\delta}}\boldsymbol{UR}+\gamma_2\boldsymbol{X^2_{TV}}-\gamma_1\boldsymbol{X^1_{TV}}+\delta_2\boldsymbol{G^2X^2}-\delta_1\boldsymbol{G^1X^1}]+\boldsymbol{G^2}(\boldsymbol{I}-\beta_2\boldsymbol{G^2})^{-1}\Delta\epsilon$ can be both split into two part: with and without error term.

Define $\mathbb{E}[\boldsymbol{G^1y^1}]\equiv\boldsymbol{G^1y^1}-\boldsymbol{G^1}(\boldsymbol{I}-\beta^1\boldsymbol{G^1})^{-1}\epsilon_1$ and $\mathbb{E}[\boldsymbol{G^2y^2}]\equiv\boldsymbol{G^2y^2}-\boldsymbol{G^2}(\boldsymbol{I}-\beta_2\boldsymbol{G^2})^{-1}\Delta\epsilon$

Consider now row six: $\frac{1}{n}(\mathbb{E}[\boldsymbol{G^1y^1}(\hat{\theta}^1_{2SLS})|\boldsymbol{X^1}])^T[\boldsymbol{i},\boldsymbol{X^2_{TV}},\boldsymbol{X^1_{TV}},\boldsymbol{G^2X^2},\boldsymbol{G^1X^1},\boldsymbol{UR}$, $\mathbb{E}[\boldsymbol{G^1y^1}],\mathbb{E}[\boldsymbol{G^2y^2}]]+\frac{1}{n}(\mathbb{E}[\boldsymbol{G^1y^1}(\hat{\theta}^1_{2SLS})|\boldsymbol{X^1}])^T[0,0,0,0,0,0,\boldsymbol{G^1}(\boldsymbol{I}-\beta^1\boldsymbol{G^1})^{-1}\epsilon_1,\boldsymbol{G^2}(\boldsymbol{I}-\beta_2\boldsymbol{G^2})^{-1}\Delta\epsilon]$.

By the assumed uniform boundedness of $X^1$, $X^2$ in absolute values as well as by the uniform boundness of the row and column sums of the matrices $G^1$, $G^2$, $(I-\beta^1G^1)^{-1}$ and $(I-\beta_2G^2)^{-1}$, by $\mathbb{E}[\Delta\epsilon]=0$ and by Lemmas A.6, A.7 and A.8 in Lee (2003), it can be shown that this row will have a limit in probability, which equals to corresponding row of $\bar{\boldsymbol{J}}$.

Similar argument holds for the row seven: $\frac{1}{n}(\mathbb{E}[\boldsymbol{G^2y^2}(\hat{\phi}_{2SLS})|\boldsymbol{X^1},\boldsymbol{X^2}])^T[\boldsymbol{i},\boldsymbol{X^2_{TV}},\boldsymbol{X^1_{TV}}$, $\boldsymbol{G^2X^2},\boldsymbol{G^1X^1},\boldsymbol{UR},\mathbb{E}[\boldsymbol{G^1y^1}],\mathbb{E}[\boldsymbol{G^2y^2}]]+\frac{1}{n}(\mathbb{E}[\boldsymbol{G^2y^2}(\hat{\phi}_{2SLS})|\boldsymbol{X^1},\boldsymbol{X^2}])^T[0,0,0,0,0,0,\boldsymbol{G^1}(\boldsymbol{I}-\beta^1\boldsymbol{G^1})^{-1}\epsilon_1,\boldsymbol{G^2}(\boldsymbol{I}-\beta_2\boldsymbol{G^2})^{-1}\Delta\epsilon]$. Therefore, the first statement is correct.

For the second statement consider $\frac{1}{\sqrt{n}}\hat{\bar{\boldsymbol{Z}}}^T\Delta\epsilon=[\boldsymbol{i},\boldsymbol{X^2_{TV}},\boldsymbol{X^1_{TV}},\boldsymbol{G^2X^2},\boldsymbol{G^1X^1},\boldsymbol{UR}$, $\mathbb{E}[\boldsymbol{G^1y^1}(\hat{\theta}^1_{2SLS})|\boldsymbol{X^1}],\mathbb{E}[\boldsymbol{G^2y^2}(\hat{\phi}_{2SLS})|\boldsymbol{X^2},\boldsymbol{X^1}]]^T\Delta\epsilon$.

None of the elements in $\hat{\bar{\boldsymbol{Z}}}$ consist $\Delta\epsilon$, therefore, since $\mathbb{E}[\Delta\epsilon]=0$, the expectation of the whole term gives 0, which concludes the consistency part of the proof.

Moreover, the variance can be written as $(\sigma_{\epsilon^1}+\sigma_{\epsilon^2})\mathbb{E}[\frac{1}{n}\hat{\bar{\boldsymbol{Z}}}^T\hat{\bar{\boldsymbol{Z}}}]$. By the same Lemmas as before, it can be shown that $plim \mathbb{E}[\frac{1}{n}\hat{\bar{\boldsymbol{Z}}}^T\hat{\bar{\boldsymbol{Z}}}] = plim \frac{1}{n}\bar{\boldsymbol{Z}}^T\bar{\boldsymbol{Z}} = \bar{\boldsymbol{J}}$, which concludes the proof of normality.■.

### Discussion of 2.4.2, step 2.

I am approaching the estimation of the second step also adopting the 2SLS procedure discussed for the first step. First, the model (5) can be rewritten in the following way:

$$\Delta \boldsymbol{y} = (\alpha_2 - \alpha_1)\boldsymbol{i} + \beta_2 \boldsymbol{G}^2\boldsymbol{y^2} - \beta_1 \boldsymbol{G^1 y^1} + \tilde{\boldsymbol{\delta}}\boldsymbol{UR} + \gamma_2 \boldsymbol{X^2_{TV}} - \gamma_1 \boldsymbol{X^1_{TV}} + \delta_2 \boldsymbol{G^2 X^2} -$$

$$-\delta_1 \boldsymbol{G^1 X^1} + \Delta\epsilon$$

Then:

$$(\boldsymbol{I} - \boldsymbol{G^1})\Delta\boldsymbol{y} = \beta_2(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G}^2\boldsymbol{y^2} - \beta_1(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1 y^1} + \tilde{\boldsymbol{\delta}}(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{UR} + \gamma_2(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{X^2_{TV}} -$$

$$- \gamma_1(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{X^1_{TV}} + \delta_2(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^2 X^2} - \delta_1(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1 X^1} + (\boldsymbol{I} - \boldsymbol{G^1})\Delta\epsilon \qquad (10)$$

Recall: $\bar{\boldsymbol{X}} = [(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{X^2_{TV}}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{X^1_{TV}}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^2 X^2}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1 X^1}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{UR}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1 y^1}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^2 y^2}]$.

And $\boldsymbol{M} = [(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{X^2_{TV}}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{X^1_{TV}}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^2 X^2}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1 X^1}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{UR}, \mathbb{E}[(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1 y^1}(\boldsymbol{\hat{\theta}^1_{2SLS}})|\boldsymbol{X^1}], (\boldsymbol{I} - \boldsymbol{G^1})(\boldsymbol{G^2})^2\boldsymbol{X^2}]$.

I modify (9), taking expectations given $\boldsymbol{X^2}$ and recalling $\mathbb{E}[\Delta\epsilon] = 0$:

$$(\boldsymbol{I} - \beta_2\boldsymbol{G^2})\mathbb{E}[(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{y^2}|\boldsymbol{X^2}] = (\boldsymbol{I} - \beta_1\boldsymbol{G^1})(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{y^1} + \tilde{\boldsymbol{\delta}}(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{UR} + \gamma_2(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{X^2_{TV}} -$$
$$-\gamma_1(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{X^1_{TV}} + \delta_2(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^2 X^2} - \delta_1(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1 X^1}$$

$$\mathbb{E}[(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{y^2}|\boldsymbol{X^2}] = (\boldsymbol{I} - \beta_2\boldsymbol{G^2})^{-1}[(\boldsymbol{I} - \beta_1\boldsymbol{G^1})(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{y^1} + \tilde{\boldsymbol{\delta}}(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{UR} + \gamma_2(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{X^2_{TV}} -$$
$$-\gamma_1(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{X^1_{TV}} + \delta_2(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^2 X^2} - \delta_1(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1 X^1}]$$

Let $\mathbb{E}[(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^2 y^2}(\phi)|\boldsymbol{X^2}, \boldsymbol{X^1}] = \boldsymbol{G^2}(\boldsymbol{I} - \beta_2\boldsymbol{G^2})^{-1}[(\boldsymbol{I} - \beta_1\boldsymbol{G^1})\mathbb{E}[(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{y^1}(\theta^1)|\boldsymbol{X^1}] + \tilde{\boldsymbol{\delta}}(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{UR} + \gamma_2(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{X^2_{TV}} - \gamma_1(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{X^1_{TV}} + \delta_2(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^2 X^2} - \delta_1(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1 X^1}]$, where $\mathbb{E}[(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{y^1}(\theta^1)|\boldsymbol{X^1}] = (\boldsymbol{I} - \beta_1\boldsymbol{G^1})^{-1}(\boldsymbol{I} - \boldsymbol{G^1})(\gamma_1\boldsymbol{I} + \delta_1\boldsymbol{G^1})\boldsymbol{X^1}$.

Then I also define the following vector $\bar{\boldsymbol{Z}} = [(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{X^2_{TV}}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{X^1_{TV}}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^2 X^2}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1 X^1}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{UR}, \mathbb{E}[(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1 y^1}(\theta^1)|\boldsymbol{X^1}], \mathbb{E}[(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^2 y^2}(\phi)|\boldsymbol{X^2}, \boldsymbol{X^1}]$

I propose the following estimation procedure:

**First,** compute the 2SLS estimator for $\phi = (\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma_1, \gamma_2, \delta_1, \delta_2)$ of the (7), using vector of instruments $\boldsymbol{M}$ and vector of covariates $\bar{\boldsymbol{X}}^{\boldsymbol{1}}$, as defined above. $\hat{\phi}^1_{2SLS} = (\bar{\boldsymbol{X}}^T\boldsymbol{P_M}\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}^T\boldsymbol{P_M}(\boldsymbol{y^2} - \boldsymbol{y^1})$, where $\boldsymbol{P_M} = \boldsymbol{M}(\boldsymbol{M}^T\boldsymbol{M})^{-1}\boldsymbol{M}^T$ is a projection matrix.

**Second,** define $\hat{\bar{\boldsymbol{Z}}} = \bar{\boldsymbol{Z}}(\hat{\phi}_{2SLS}) = [(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{X^2_{TV}}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{X^1_{TV}}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^2 X^2}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1 X^1}, (\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{UR}, \mathbb{E}[(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^1 y^1}(\hat{\theta}^1_{2SLS})|\boldsymbol{X^1}]], \mathbb{E}[(\boldsymbol{I} - \boldsymbol{G^1})\boldsymbol{G^2 y^2}(\hat{\phi}_{2SLS})|\boldsymbol{X^2}, \boldsymbol{X^1}],$

where $\mathbb{E}[(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{G^1}\boldsymbol{y^1}(\hat{\theta}^1_{2SLS})|\boldsymbol{X^1}] = (\boldsymbol{I}-\hat{\beta}_{1,2SLS}\boldsymbol{G^1})^{-1}(\boldsymbol{I}-\boldsymbol{G^1})(\hat{\gamma}_{1,2SLS}\boldsymbol{I}+\hat{\delta}_{1,2SLS}\boldsymbol{G^1})\boldsymbol{X^1}$, with $\hat{\theta}^1_{2SLS}$ obtained as the estimation of the first stage on the first step.
and $\mathbb{E}[(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{G^2}\boldsymbol{y^2}(\hat{\phi}_{2SLS})|\boldsymbol{X^2},\boldsymbol{X^1}] = \boldsymbol{G^2}(\boldsymbol{I}-\hat{\beta}_{2,2SLS}\boldsymbol{G^2})^{-1}[(\boldsymbol{I}-\hat{\beta}_{1,2SLS}\boldsymbol{G^1})\mathbb{E}[(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{y^1}(\hat{\theta}^1_{2SLS})|\boldsymbol{X^1}] + \hat{\boldsymbol{\delta}}_{2SLS}(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{UR} + \hat{\gamma}_{2,2SLS}(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{X^2_{TV}} - \hat{\gamma}_{1,2SLS}(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{X^1_{TV}} + \hat{\delta}_{2,2SLS}(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{G^2}\boldsymbol{X^2} - \hat{\delta}_{1,2SLS}(\boldsymbol{I}-\boldsymbol{G^1})\boldsymbol{G^1}\boldsymbol{X^1}]$

**Finally,** we use $\hat{\hat{\boldsymbol{Z}}}$ as a new vector of instrument to estimate (7). Then the following consistent estimator is obtained: $\hat{\phi}_{Lee} = (\hat{\hat{\boldsymbol{Z}}}^T\bar{\boldsymbol{X}})^{-1}\hat{\hat{\boldsymbol{Z}}}^T(\boldsymbol{y^2}-\boldsymbol{y^1})$.

## B. Additional Tables and figures

Table B.1: Distribution of the number of friends in samples

| # of friends | Long study, year 1 | | Long study, year2 | | Short study | |
|---|---|---|---|---|---|---|
| 0 | 17 | 5.29% | 26 | 8.12 % | 9 | 4.39 % |
| 1 | 1 | 0.31% | 14 | 4.37% | 3 | 1.46 % |
| 2 | 5 | 1.56% | 34 | 10.62% | 4 | 1.95 % |
| 3 | 28 | 8.72% | 39 | 12.19% | 17 | 8.29 % |
| 4 | 32 | 9.97% | 56 | 17.5% | 28 | 13.66 % |
| 5 | 39 | 12.15% | 55 | 17.19% | 34 | 16.59 % |
| 6 | 41 | 12.77% | 39 | 12.19% | 34 | 16.59 % |
| 7 | 150 | 46.73% | 33 | 10.31% | 28 | 13.66 % |
| 8 | 2 | 0.62% | 0 | 0.00% | 21 | 10.24 % |
| 9 | 1 | 0.31% | 0 | 0.00% | 14 | 6.83 % |
| 10 | 3 | 0.93% | 0 | 0.00% | 4 | 1.95 % |
| 11 | 0 | 0.00% | 0 | 0.00% | 3 | 1.46 % |
| 12 | 0 | 0.00% | 0 | 0.00% | 1 | 0.49 % |
| 13 | 1 | 0.31% | 0 | 0.00% | 3 | 1.46 % |
| 14 | 0 | 0.00% | 0 | 0.00% | 2 | 0.98 % |

Table B.2: Unified State Exams statistics

| Subject | Number of participated | Average grade |
|---|---|---|
| Mathematics | 305 | 59.87 |
| Russian | 305 | 79.85 |
| Biology | 2 | 71.5 |
| Chemistry | 1 | 80 |
| Computer Science | 49 | 76.96 |
| Economics | 27 | 32.52 |

| | | |
|---|---|---|
| Foreign Language | 272 | 70.64 |
| Geography | 4 | 67 |
| History | 78 | 70.94 |
| Law | 20 | 69.4 |
| Literature | 20 | 69.35 |
| Orientalism | 2 | 75 |
| Physics | 49 | 58.45 |
| Social Studies | 269 | 71.01 |

Table B.3: Descriptive statistics

| Variable | Mean | St.Dev. | Min | Max |
|---|---|---|---|---|
| Average grade, wave 1 | 7.20 | 0.94 | 4.58 | 9.35 |
| Average grade, wave 2 | 7.23 | 1.13 | 4.50 | 9.86 |
| Retakes (dummy) | 0.33 | 0.47 | 0 | 1 |
| Retakes (number) | 0.684 | 1.25 | 0 | 6 |
| Ability | 183.6 | 70.09 | 106 | 355 |
| Gender (f) | 0.67 | 0.47 | 0 | 1 |
| Tuition, wave 1 (private) | 0.18 | 0.38 | 0 | 1 |
| Tuition, wave 2 (private) | 0.184 | 0.39 | 0 | 1 |
| Economics department | 0.328 | 0.47 | 0 | 1 |
| Management department | 0.272 | 0.45 | 0 | 1 |
| Computer Science department | 0.26 | 0.44 | 0 | 1 |
| Working status, wave1 (not working) | 0.804 | 0.39 | 0 | 1 |
| Working status, wave2 (not working) | 0.74 | 0.44 | 0 | 1 |
| Higher Education of mother | 0.796 | 0.4 | 0 | 1 |
| Higher Education of father | 0.624 | 0.49 | 0 | 1 |
| Single parent family | 0.2 | 0.40 | 0 | 1 |
| Family with more than 1 kid | 0.54 | 0.50 | 0 | 1 |
| Living conditions, wave 1 (dormitory) | 0.16 | 0.37 | 0 | 1 |
| Living conditions, wave 2 (dormitory) | 0.172 | 0.38 | 0 | 1 |

Figure B.1: Distribution of friends in Short and Long surveys

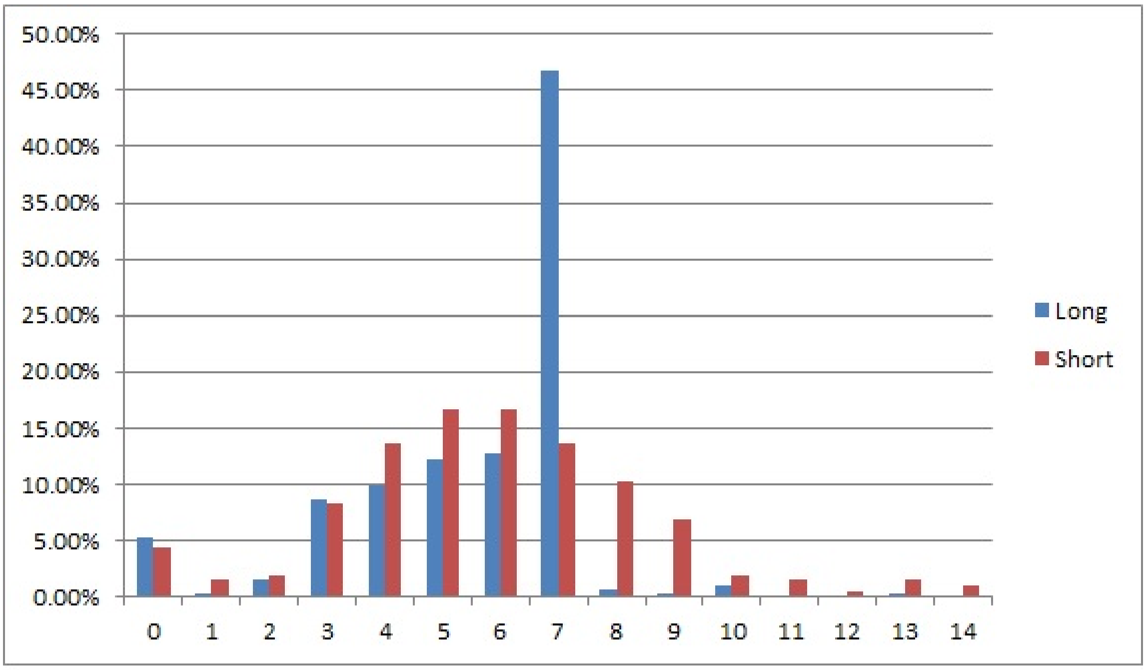



Table B.4: Results for the models with reciprocal links and best friends, no correlated effects

| Variable | Recipr., (1) | Recipr., (2) | Best, (3) | Best, (4) |
|---|---|---|---|---|
| Constant | | -0.2318• | | -0.3127** |
| Unexpected Retake | 0.1320 | -0.0097 | 0.0469 | 0.0768 |

| | | | | |
|---|---|---|---|---|
| Endogenous effect, period 1 | 0.0180 | 0.0111 | 0.0231 | -0.0467** |
| Endogenous effect, period 2 | 0.0480* | 0.0407* | 0.0818*** | 0.0215 |
| *Time-variant own controls* | | | | |
| Tuition, w1 | 0.0317 | | -0.0771 | |
| Tuition, w2 | -0.1547 | | -0.2518 | |
| Working status, w1 | | -0.0909 | | -0.1235 |
| Working status, w2 | | 0.1483* | | 0.1510* |
| *Network's controls* | | | | |
| Economics, w1 | 0.0778 | -0.0214 | 0.0082 | 0.1953 |
| Economics, w2 | -0.4701** | -0.5337** | -0.4869** | -0.4096** |
| Computer Science, w1 | | -0.4977** | | |
| Computer Science, w2 | | -0.3899$^{\bullet}$ | | |
| Working status, w1 | -0.2272 | | -0.1881 | |
| Working status, w2 | -0.1623 | | -0.4340** | |
| Siblings, w1 | | | | 0.2611* |
| Siblings, w2 | | | | -0.0080 |
| Sample size | 250 | 250 | 250 | 250 |
| BIC | -224.67 | -221.77 | -221.41 | -218.07 |

*** - p-value $< 0.01$, ** - p-value $< 0.05$, * - p-value $< 0.1$, $^{\bullet}$ - p-value $< 0.15$

Table B.5: Results for the models with best friends and reciprocal links, with correlated effects

| Variable | Recipr.,(1) | Recipr.,(2) | Best,(3) | Best,(4) |
|---|---|---|---|---|
| Unexpected Retake | -0.1212 | -0.0913 | 0.0406 | -0.1365 |
| Endogenous effect, period 1 | -0.0081 | -0.1188 | 0.0235 | -0.0404 |
| Endogenous effect, period 2 | 0.0498 | 0.0016 | 0.0811 | -0.0262 |
| *Time-variant own controls* | | | | |
| Tuition, w1 | | 0.0394 | | 0.1946 |
| Tuition, w2 | | -0.1933 | | -0.0608 |
| Working status, w1 | 0.0296 | | -0.0248 | |
| Working status, w2 | 0.0493 | | 0.1437 | |
| *Network's controls* | | | | |
| Abilities, w1 | | -0.0004 | | |
| Abilities, w2 | | -0.0029 | | |
| Tuition, w1 | -0.5923 | | | 0.2977 |
| Tuition, w2 | -0.4462* | | | 0.1171 |
| Economics, w1 | | | 1.0059 | |
| Economics, w2 | | | -0.2755 | |
| HE of mother, w1 | | | | -0.3578 |
| HE of mother, w2 | | | | -0.2662 |
| Single Parent, w1 | -0.0083 | | | |
| Single Parent, w2 | -0.1985 | | | |

| | | | | |
|---|---|---|---|---|
| Siblings, w1 | | | 0.2174 | |
| Siblings, w2 | | | 0.1179 | |
| Sample size | 250 | 250 | 250 | 250 |
| BIC | -191.02 | -192.71 | -189.98 | -183.06 |

*** - p-value < 0.01, ** - p-value < 0.05, * - p-value < 0.1, • - p-value < 0.15

Table B.6: List of classes with retakes in the sample

| Class | Department | Total No. of retakes | Important classes |
|---|---|---|---|
| Algebra | Computer Science | 21 | No |
| Architecture of Computer Systems | Computer Science | 2 | No |
| Architecture of ECM | Computer Science | 1 | No |
| Basics of computer technology and programming | Computer Science | 3 | Yes |
| Discrete Mathematic | Computer Science | 8 | No |
| Discrete Mathematic | Economics | 2 | No |
| Economic Theory and Institutional Analysis | Management | 28 (in 2 terms) | Yes |
| Economic Theory and Institutional Analysis | Computer Science | 12 | No |
| Economic Theory Basics | Economics | 27 (in 3 terms) | Yes |
| Economics | Computer Science | 3 | No |
| English and other languages | All departments | 9 | No |
| Geometry and Algebra | Computer Science | 8 | Yes |
| History of economic thoughts | Economics | 1 | Yes |
| History of foreign state and law | Law | 2 | Yes |
| Introduction to software engineering | Computer Science | 3 | Yes |
| Judicial power and law enforcement | Law | 1 | No |
| Life safety | All departments | 3 | No |
| Linear Algebra | Economics | 28 | No |
| Mathematical Analysis | Computer Science | 68 (in 2 terms) | Yes |
| Mathematical Analysis | Economics | 12 | Yes |
| Mathematics | Management | 31 (in 2 terms) | Yes |
| Methods of financial and economic computations | Economics | 1 | No |
| Microeconomics | Computer Science | 18 (in 2 terms) | Yes |
| Philosophy | Management | 6 | Yes |
| Roman Law | Law | 1 | No |
| Socio-Economic Statistics | Economics | 2 | No |

| | | | |
|---|---|---|---|
| Sociology | Management | 1 | Yes |
| Theoretical basics of computer technology | Computer Science | 9 (in 2 terms) | No |
| Theory of state and law | Law | 4 | Yes |