Discrete Graphical Models in Social Mobility Research -
A Comparative Analysis of American, Czechoslovakian and Hungarian Mobility before the Collapse of State Socialism

Renáta Németh[1]
Tamás Rudas[2]

THE FOLLOWING TEXT IS THAT OF THE ORGINIAL SUBMISSION.
THE FINAL PUBLISHED VERSION WILL BE SLIGHTLY DIFFERENT.

Abstract

Variants of path models have been widely used for the analysis of the social status attainment process. The methods presented here differ from earlier approaches in several ways. Social status is considered a categorical variable and path models are developed starting from graphical models, using the marginal log-linear approach. Overall model fit may be tested by standard techniques. Under these models, the status attainment process is completely characterized by a set of parameters that measure the strengths of the relevant effects. This is in sharp contrast with estimating and interpreting ad hoc parameters, without paying attention to overall model fit and to other effects influencing the process. The method is applied to the social status attainment process in the USA, Hungary and Czechoslovakia at the end of the last century, and shows that policies in the latter socialist countries to prevent status inheritance had little success.

---

[1] Corresponding author: Renáta Németh, Department of Statistics, Faculty of Social Sciences, Eötvös Loránd University, Pázmány Péter sétány 1/A, H-1117, Budapest, Hungary.
Email: nemethr@caesar.elte.hu
[2] Department of Statistics, Faculty of Social Sciences, Eötvös Loránd University, Hungary

1

"There has been a revolution in the study of social mobility: the once dominant Blau-Duncan paradigm has been overthrown by log-linear modeling. [...] The log-linear revolution was a noble experiment, and at first seemed to offer a bright new future beyond Blau and Duncan. [...] In the end, I believe Goldthorpe's experiment failed. A decade and more has passed. Endless models have been fitted; legions of design matrices passed in review; chi-squares marshaled. But we have learnt little new of substance. Frequencies in a crude cross-classification of father's occupation by son's occupation are described in loving detail, with some ad-hoc interpretation but little theory. [...] This compares poorly with the statistically simpler but conceptually more sophisticated results of the Blau-Duncan paradigm."

Jonathan Kelly: The failure of a paradigm (1990)

Introduction

The motto of our paper reveals the long-standing need for covering structural relationships as Blau and Duncan did but defining social status in Goldthorpe's way with categorical variables. Discrete graphical models which this paper deals with are capable to achieve these two goals simultaneously. They generalize log-linear models and are able to model direction of influences.

Graphical models (classic works are Whittaker, 1990, Cox and Wermuth, 1996 and Lauritzen, 1996) provide a well visualizable and easily interpretable representation of complex associations. The models are based on graphs in which nodes represent variables and missing edges represent conditional independence statements. Modifications of the hypotheses can be easily followed by drawing new or deleting old edges. Take as an example the classic model of Figure 1. There is no edge between father's education and offspring's occupation which means that they are conditionally independent, controlled for the predictors of occupation, father's occupation and offspring's education. This assumption is generally plausible for Western countries, but may fail in the case of socialist Eastern-European societies. For example in Czechoslovakia before the collapse of the State Socialist system, if higher status children could not study as a consequence of negative

2

educational discrimination, they found other ways to achievement with the help of inherited cultural capital and aspirations (Boguszak et al., 1990). Father's education had a direct effect on offspring's occupation; therefore the graph of Figure 1 should be completed with an arrow.
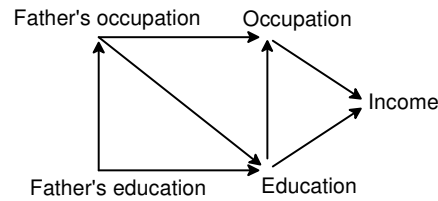
Figure 1. A status attainment model.

In this paper, our focus is on categorical variables, but graphical models can also handle continuous data (e.g. Cox and Wermuth, 1996). Continuous data with normality assumed and categorical data turned out to be quite different from the perspective of graphical modeling. The study of discrete models is still an active research area (e.g. Drton, 2009, Rudas et al., 2010, Marchetti and Lupparelli, 2011), with some questions still open. However, several important results are already available; this paper introduces them through a concrete sociological example rather than through mathematics. The example concerns social mobility, however, extension of the present model to other research areas is straightforward. Németh and Rudas (2013) present the underlying theory in more detail.

Graphical models

Graphical models are models based on graphs in which nodes represent random variables, and missing edges represent conditional independence assumptions, hence the graphs provide a compact representation of characteristics of joint probability distributions. If the graph has only directed edges (arrows), but contains no directed cycles, is called a directed acyclic graph (DAG, also called as Bayesian network or belief network). The conditional independencies defining a DAG can be read off using the rule that a variable is conditionally independent of its non-descendants, given its parents. The graph of Figure 1 is a DAG, and the rule implies that the following conditional independencies define the model:

$$I \perp E'O' \mid OE,$$
$$O \perp E' \mid O'E, \tag{1}$$

where *I* stands for income, *E* stands for education, *O* stands for occupation, and *E'* and *O'* stand for father's characteristics. $O \perp E' | O'E$ means that father's education has no direct effect on his offspring's occupation controlled for the mediator variables, father's occupation and offspring's education. $I \perp E'O' | OE$ means that *E'* and *O'* do not have a direct effect neither separately nor jointly on *I*, controlled for *O* and *E*. The latter statement points to an important difference between discrete and Gaussian graphical models. The pairwise independence statements $I \perp E' | OE$ and $I \perp O' | OE$ imply $I \perp E'O' | OE$ only in the case of normally distributed variables, but not in the case of categorical data. In association with this, in the case of Gaussian data only effects associated with the edges exist, e.g., the effects of *E* and *O* on *I* might be separated. But in the discrete case this assumption does not hold automatically, rather needs to be imposed. Discrete graphical models fulfilling this assumption are inspired by Goodman (1973) and were discussed within the unified framework of marginal log-linear modeling by Rudas et al. (2006). Rudas et al. derived these models from DAG models and, following Goodman and emphasizing the parallelism with Gaussian path models, called them discrete path models. This important difference between continuous and categorical variables points to the fact that concerning conditional associations, categorical data allow for richer structure than continuous data do under the almost automatic assumption of multivariate normality.

In the paper DAG models and path models derived from DAGs are considered. There are two other main types of graphical models which differ with regard to their edges, that is, with regard to the association structure capable to describe (Lauritzen, 1996). A graph with only undirected edges is called an undirected graph (also called Markov network or Markov random field). Variables in this case are considered on equal footing. Undirected graphs and DAGs were extended to permit both directed and undirected edges; such graphs are called chain graphs. Chain graphs allow both response-predictor and symmetric association relations between variables. In the last decade different competing approaches to chain graphs emerged, see e.g. Andersson et al. (2001), Drton (2009) and also Németh and Rudas (2013).

Social mobility research context

Sociological application of graphical models may arise in the field of intergenerational social mobility. Several shifts occurred in the last decades in research on social mobility, and the successive generations of researchers remarkably differed from a methodological viewpoint (Ganzeboom, Treiman and Ultee, 1991, Treiman and Ganzeboom, 2000, Wolbers et al., 2011). The first generation of mobility research (one of the main works is

Lipset and Bendix, 1959) focused on cross-tabulation of father's occupation by son's occupation, detecting patterns of mobility and measuring the openness of societies within this table. The second generation exchanged these tables for path models describing the whole status attainment process (Blau, Duncan, 1967). They operationalized occupation with continuous rather than categorical variables. Path models can handle more than two variables; hence intervening variables like son's education could be also involved in the analysis. They „opened the black box of mobility" (expressive words of Kelley, 1990) constituted of the two-way table of the former generation. The third generation found again the categorical operationalization theoretically more adequate. They turned back to the two-way table of the first generation, but used log-linear models which allow finer analysis of the table (Featherman and Hauser, 1978, Goldthorpe, Payne and Llewellyn, 1978, Goodman, 1973). The fourth generation does not represent a unified paradigm, neither from the perspective of their research questions nor from that of their methods. Methods of this generation important in the context of discrete graphical modeling are unidiff models (Erikson and Goldthorpe, 1992, Xie, 1992); extensions of path models to categorical variables, primarily within the framework of structural equation modeling (a good summary is Menard, 2010, chapter 8) and different generalizations of log-linear models continuing the third generation's thinking (Zelterman and Youn, 1992, Rudas, Klimova, 2011).

It is conspicuous how the viewpoint of successive generations altered regarding the operationalization of variables. Researchers divided into those who favor a class approach (categorical variables) and those who favor a hierarchical approach (continuous variables). The choice is not a technical one: it follows from the researcher's sociological theory, and it has far reaching consequences on the kinds of analyses that are usually performed with the data. None of the two approaches is superior; the common argument that continuous coding containing more information is more precise, fails. The continuous variable approach assumes that only one dimension can capture the differences between occupational groups and it allows an unlimited number of groups. The categorical variable approach assumes that there exist a number of clearly distinguishable classes based on, for example, labor market position, skill level and sectoral differences. Classes are assumed to be not necessary ordered and relatively homogeneous internally. Regarding the statistical methods, while the continuous approaches usually seek simplicity and structure in some kind of a linear relationship, the categorical approach allows for much richer structures. The categorical approach shows us where the typical channels for or barriers to movement are in the social structure, providing clues about the stratification process. These "channels" and "barriers" (Blau and Duncan's terminology, 1967:117) cannot be revealed within the

continuous approach, because it averages different recruitment patterns specific to different classes.

Another alternation can also be observed in the history of mobility research: sometimes the research question of "how much mobility" (macro perspective, collective mobility, typical for the first generation), other times the question of "how mobility works" (micro perspective, individual process, typical for the second generation) was put forward. The first approach defines one specific measure of mobility, which makes across-country and across-time comparisons easier. Such a single measure has its statistical advantage as well, since it provides great statistical power to discern slight trends. For this very reason, the third generation's log-linear models were criticized for estimating too many parameters and not being powerful instruments in comparative researches (Vallet, 2001). To solve this problem, unidiff models for two-way mobility tables were proposed (Erikson and Goldthorpe, 1992, also known as log-multiplicative layer effect model, Xie, 1992). The name, „unidiff" comes from „uniform difference": the model assumes that all odds ratios of the table change in the same direction over time; this common change is measured by the unidiff parameter.

Single measures of mobility make comparisons easier, but do not provide explanation. For this reason the two approaches could be well combined: process models like graphical models could describe the mechanisms which are behind the trends detected by aggregated measures.

The earliest measures answering the question of "how much mobility" were functions of the margins of the occupational table, i.e. the fathers' and the offsprings' occupational structure, therefore they were inadequate for comparative analyses. This problem was recognized during the first generation and several proposals were put forward to solve it (Ganzeboom et al., 1991). The third generation's log-linear models also achieved this requirement, since the log-linear parameters describing margins (main effects) and associations (interaction effects) are variation independent, which means that they can have any value independently from each other (Rudas, 1998). Parameters defining discrete graphical models can also be chosen to be variation independent (Bergsma, Rudas, 2002). On the contrary, the unidiff parameter is affected by structural changes (Breen and Luijkx, 2004).

Among methods answering „how mobility works" are path models and their generalization, structural equation modeling (SEM). SEM, similarly to graphical modeling, represents the associations by a graph. However, the two methods define this graph in different ways. Graphical models characterize the association structure with the lack of some edges, which

correspond to conditional independences. SEM, on the contrary, focuses on existing rather than missing edges and estimates direct effects pertaining to the edges. In SEM, a missing edge does not in general imply independence, neither conditional nor unconditional (Cox and Wermuth, 1993, Whittaker, 1990, Cox and Wermuth, 2001). Graphical modeling is preferable if the research question concerns conditional independences (which formerly attained resources are already irrelevant at a given step of the status attainment process).

The other main difference to graphical models is the way in which categorical variables are modeled. The SEM approach, developed originally for Gaussian variables, has been extended to handle categorical variables in many ways; they most often are treated as a manifestation of an underlying continuous variable (see as a review Kupek, 2005, 2006). On the contrary, graphical models treat categorical data as inherent categories. Finally, both methods aim to identify direct and indirect effects of one variable on another, however, graphical modeling does not provide an additive decomposition of the total effect. Despite these differences, several new results at the intersection of the two areas emerged in the last years. An example is Koster (1999), who showed that the independence structure defined with any structural equation model for Gaussian variables can be read off from a corresponding graph.

To sum up, selection of the appropriate method cannot be reduced to a statistical question. Background concepts and research questions affect when graphical models are preferable to other alternatives or can complement them by introducing a new viewpoint.

Marginal log-linear parameterization

If one decides to operationalize variables as categorical and to use graphical modeling, a basic question is how to parameterize the contingency table in order to easily define graphical models on it. This is an important issue, since (1) a graphical model could be given by restricting the parameters (preferably in an easy way, setting some of them to zero). Additionally, (2) it is very often the case that one moves from a poorly fitting model to a slightly less restrictive model in the hope of a better model fit, by adding new edges to the graph, one after the other. Or on the contrary, the initial model is a well fitting one, and one searches for a slightly more restrictive model in the hope of still good model fit, by removing edges in the graph. One can obtain the new model by omitting some (or introducing new) parameter restrictions. Furthermore, (3) interpretation of a model is usually based on estimated parameter values. An improper parameterization may consist of variationally dependent parameters which make the interpretation of the parameters

misleading. Finally, (4) some parameterizations lead to the standard results from asymptotic distribution and testing theory to fail.

These problems can be avoided by defining graphical models with properly chosen marginal log-linear parameters (Bergsma and Rudas, 2002). These parameters are marginal, because they are defined not in the whole table rather within a marginal of it. Let us take as an example the 2X2X2X2 table of O' (father's occupation), E (educational level), O (occupation) and I (income). Recall that the usual log-linear parameter of an effect measures the association among the variables, conditioned on the categories of all other variables and averaged over these categories. For example the parameter of the EI effect pertaining to the $E=1$ and $I=1$ indices can be computed as[3]:

$$\lambda_{*1*1}^{O'EOI} = \frac{1}{4} \sum_{i,k} \log \left(\frac{p_{i1k1} p_{i2k2}}{p_{i1k2} p_{i2k1}} \right)^{1/4}, \tag{2}$$

where p_{ijkl} is the probability of the cell $O'=i$, $E=j$, $O=k$ and $I=l$. Note that the expression in the brackets is the odds ratio computed within the conditional table. The parameter may take any number, positive or negative. In case of positivity, on the average there are more observations in the four cells with $E=1$, $I=1$ indices than there would be if E and I were conditionally independent (holding O' and O fixed, people with educational level of 1 are more likely to be in the first income category); and vice versa, in case of negativity having the first educational level is a barrier to achieving the first income category, when O' and O are controlled for.

If, instead of this, one is interested in the association measured in a marginal rather than the entire table, marginal log-linear parameters can be defined in a similar way. The EI effect, for example, can be measured within the EI marginal after marginalization over O' and O:

$$\lambda_{11}^{EI} = \log \left(\frac{p_{+1+1} p_{+2+2}}{p_{+1+2} p_{+2+1}} \right)^{1/4}, \tag{3}$$

or within the EOI marginal after marginalization over O':

$$\lambda_{1*1}^{EOI} = \frac{1}{2} \sum_{k} \log \left(\frac{p_{+1k1} p_{+2k2}}{p_{+1k2} p_{+2k1}} \right)^{1/4}, \tag{4}$$

---

[3] In this paper, as in the most applications of log-linear modeling, λ parameters are calculated using effect coding (Agresti 2002).

where "+" is a marginalization operator. The parameters in (2), (3) and (4) all pertain to the *EI* effect, but apparently, their values heavily depend on the marginal from which they are computed. While the parameter λ_{11}^{EI}, measuring the total effect of education on income, is expected to take a high positive value; the parameter λ_{1*1}^{EOI} is smaller, since it measures the direct effect of education on income controlled for occupation.

Graphical models cannot be parameterized with usual log-linear parameters, since the conditional independence restrictions defining the model in general do not involve all variables, rather certain subsets of them. However, marginal log-linear parameters of the above kind can be used to parameterize such models. As it is readily implied by the results of Bergsma and Rudas (2002), marginal log-linear parameterizations which meet some simple criteria assure that missing edges correspond to some of the parameters being zero, and the non-restricted parameters of the model can be interpreted in a straightforward way. The way these combinatorial properties can be achieved was shown by Rudas et al. (2006) for DAGs and by Rudas et al. (2010) for a class of chain graphs, and this is illustrated next.

As an example, let us take the DAG of Figure 1. Following Rudas et al. (2006), let us consider a complete ordering of the variables which suits the direction of the arrows. Let this ordering be *E'O'EOI*, where *E*, *O*, *I* and *O'* is as before, while *E'* is father's education. The ordering gives the marginals used in the parameterization: *E'O'*, *E'O'E*, *E'O'EO* and *E'O'EOI*. Each subset of the five variables makes up an effect, the marginal within which an effect is computed is the smallest possible one (i.e. the parameterization is hierarchical, see Table 1).The conditional independencies in (1) can be obtained by setting those parameters to zero whose effect contains non-connected nodes.

Table 1. Parameterization of the DAG of Figure 1**.**

Marginal	*E'O'*	*E'O'E*	*E'O'EO*	*E'O'EOI*
Effect set to zero			*E'O, E'O'O, E'EO, E'O'EO*	*E'I, O'I, E'O'I, E'EI, O'EI, E'O'EI, E'OI, O'OI, E'O'OI, E'EOI, O'EOI, E'O'EOI*
Free effect	*empty, E', O', E'O*	*E, E'E, O'E, E'O'E*	*O, O'O, EO, O'EO*	*I, EI, OI, EOI*

9

The free parameters contain all information about the distributions in the model. The model can be parameterized in other ways as well, but the one obtained this way avoids the problems listed in the beginning of this section (for the proof see Bergsma and Rudas, 2002). Namely, the parameters are variation independent and may be interpreted separately, maximum likelihood estimates exist and have an asymptotic normal distribution, and the likelihood ratio statistic for testing the model has an asymptotic chi-squared distribution with the number of zero parameters as the number of degrees of freedom. Data analysis happens in a straightforward way: data are fitted to the model by forcing some parameters to be zero, and in case of good fit the status attainment process is interpreted with the help of the estimated values of the free parameters. This method is in sharp contrast with estimating and interpreting ad hoc parameters, without paying attention to overall model fit and to other effects influencing the process.

The marginals of the free parameters can be simplified in a way which makes the interpretation easier while keeping the parameterization's favorable characteristics unchanged. Namely, if the above zero parameters are fixed, then those variables which are neither elements of the effect nor parents of the last variable in the marginal can be removed from the marginal without any change in the value of the parameters. For example, the strength of the $O'O$ effect is the same whether computed within the $E'O'EO$ or within the $O'EO$ marginal, because E' is neither an element of $O'O$ nor the parent of O. Without proving this, we refer to the fact that conditional independence is equivalent to a simplification of the conditional probability function (e.g. Lauritzen, 1990):

$$X \perp Y \mid Z \Leftrightarrow p(x \mid y, z) = p(x \mid z).$$

After simplifications, we obtain the final parameterization of Table 2, where the marginals of the free parameters contain only one variable with its parents.

Table 2. Simplified parameterization of the DAG of Figure 1.

Marginal	$E'O'$	$E'O'E$	$O'EO$	$E'O'EO$	EOI	$E'O'EOI$
Effect set to zero				$E'O,$ $E'O'O,$ $E'EO,$ $E'O'EO$		$E'I, O'I, E'O'I, E'EI,$ $O'EI, E'O'EI, E'OI,$ $O'OI, E'O'OI, E'EOI,$ $O'EOI, E'O'EOI$
Free effect	*empty,* $E', O',$	$E, E'E,$ $O'E,$	$O, O'O,$ $EO,$		$I, EI,$ $OI,$	

10

	E'O'	E'O'E	O'EO		EOI	

Separate interpretability of the parameters permit that one could obtain new models by introducing new parameter restrictions. For example, if one wants to derive a discrete path model by introducing the assumption that automatically holds for Gaussian path models and Gaussian DAGs, namely, that only effects associated with the edges exist, then effects with more than two elements should be also set to zero. Less formally this assumption states that parents with common child do not have a joint effect on the child over and above their separate effects. The remaining free parameters parameterizing the distributions in the path model are listed in Table 3.

**Table 3.** Parameterization of the path model of Figure 1.

Marginal	E'O'	E'O'E	O'EO	E'O'EO	EOI	E'O'EOI
Effect set to zero		E'O'E	O'EO	E'O, E'O'O, E'EO, E'O'EO	EOI	E'I, O'I, E'O'I, E'EI, O'EI, E'O'EI, E'OI, O'OI, E'O'OI, E'EOI, O'EOI, E'O'EOI
Free effect	*empty, E', O', E'O'*	*E, E'E, O'E, O'E*	*O, O'O, EO*		*I, EI, OI*	

Status attainment in the USA, Hungary and Czechoslovakia

The model of Figure 1 is a variant of the classic Blau-Duncan model; its original source is Duncan et al. (1968), but made widely known by Treiman (1970), whose often cited modernization hypothesis was defined using this model. According to Treiman's modernization hypothesis, economical-technological development leads to a more open society as the increasing bureaucratization of work makes direct inheritance of positions more difficult: the job market of a developed economy rewards abilities achieved through formal education (Treiman, 1970). Treiman expressed his hypothesis in a formal way referring to the steps of the status attainment process: as societies develop, the direct influence of father's occupation on education and occupation declines, similarly to the education-income direct effect, while the education-occupation and occupation-income direct effects become stronger.

In what follows, the model Figure 1 is fitted to data from the International Social Survey Programme, choosing the year of 1992 in order to detect the effects of socialist regimes on mobility. The analysis covers two former socialist states of Central Europe (Czechoslovakia, Hungary) and a liberal welfare state, the USA. Regarding these three countries, Treiman and other economic theories (e.g. Lipset and Zetterberg, 1959, Blau and Duncan, 1967, Jonsson, 1992) would predict that social mobility is highest in the economically more developed USA. However, apart from industrialization other factors, like political intervention might influence status attainment, possibly in a different direction. Socialist regimes had the proclaimed goal of creating an egalitarian society; and they favored the advancement of people with working class backgrounds. According to Boguszak et al. (1990) these policies could not completely achieve their goals in Czechoslovakia. They claim that inherited aspirations and cultural capital compensated for educational discrimination against politically undesirable higher status students. Thus regarding the USA-Hungary-Czechoslovakia comparison, political and economic approaches do not yield a consistent forecast. Our study aims to contribute to this question by fitting Treiman's model and estimating the effects appearing in his modernization hypothesis. These effects correspond to the free parameters in Table 2.

Wording their finding more formally, Boguszak et al. (1990) hypothesized that occupation is directly influenced by father's education. The hypothesis corresponds to the addition of the arrow *E'O* to the graph (Figure 2).
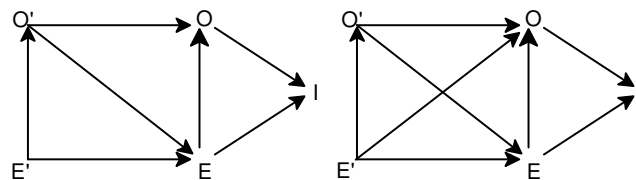
Figure 2. Treiman's model and its modification.

A new arrow added always means some ease in the conditional independence restrictions; in the present case the new graph implies only one of the two statements listed in (1):

$I \perp O'E' | OE.$

Since the conditional independence defining the model is given within only one marginal, the model is a log-linear one, which can be parameterized with the usual log-linear

parameters. However, the DAG-parameterization described above has the advantage of yielding better interpretable parameters, because it uses different marginals following the structure of the graph. Table 4 shows the parameters obtained this way.

Table 4. Parameterization of the modified Treiman model of Figure 2.

Marginal	*E'O'*	*E'O'E*	*E'O'EO*	*EOI*	*E'O'EOI*
Effect set to zero					*E'I, O'I, E'O'I, E'EI, O'EI, E'O'EI, E'OI, O'OI, E'O'OI, E'EOI, O'EOI, E'O'EOI*
Free effect	*empty, E', O', E'O*	*E, E'E, O'E, E'O'E*	*O, O'O, EO, O'EO, E'O, E'O'O, E'EO, E'O'EO*	*I, EI, OI, EOI*	

In the analysis these two models of status attainment process will be tested. The ISSP 1992 data are restricted to those at least 25 years old with non-missing data for each variable. The sample size is 725 for Czechoslovakia, 857 for Hungary and 524 for the USA. The five-dimensional table of the joint distribution contains 28, 23 and 28 empty cells, respectively; therefore a value of 0.5 is added to each cell.

Income is defined as a binary variable (above/below median); it is derived from the respondent's average monthly income using the country-specific median as a cut point to allow cross-country comparison. Education is also binary (with or without a college degree). The questionnaires defined father's occupation as the one he had when the respondent was 16 years old, while respondent's occupation is her/his present job or the last job if she/he did not have one. We use a hierarchical three-class version of the EGP classification following the recommendation of Erikson and Goldthorpe (1992, pp.44-46.):

1. upper class (professionals, administrators, officials, managers, higher-grade technicians),
2. middle class (routine non-manual workers, pretty bourgeoisie, skilled workers, and for offsprings: farmers and other self-employed workers in primary production),
3. lower class (non-skilled workers, agricultural and other workers in primary production, and for fathers: farmers and other self-employed workers in primary production).

The five variables yield a table with 72 cells, which can be parameterized with 72 parameters, some of which are set to zero by the model. The number of non-redundant zero parameters equals to the number of degrees of freedom of the likelihood ratio test statistics (L^2) that is used to test model fit. Since the log-linear parameters summed in any of its indices are 0, for a k level index, there are $k-1$ non-redundant parameters. Therefore the zero effects in Table 1 define four non-redundant parameters if they contain both O and O'. There are two non-redundant parameters if the effect contains only one of O and O', and there is only one non-redundant parameter if the effect does not contain them. This yields altogether 42 non-redundant parameters for model 1 (Treiman's original model) and 30 for model 2 (modified Treiman's model). Since some empty cells may be structural (i.e. also empty in the population), the real degrees of freedom is between these values and the ones reduced by the number of empty cells, rather closer to the former ones.

Model 1 is nested within model 2, therefore the difference of the likelihood ratio statistics can be used to compare them. The Bayesian Information Criterion (BIC) is also computed with taking the degrees of freedom at its maximum value. As the real degrees of freedom differs from this value to a different extent country by country, our BIC is proposed for within-country comparisons.

Table 5 shows the results. The column df contains maximal and minimal values of the degrees of freedom, and according to this, the column p contains two, maximal and minimal, values. The real p values are between these, rather closer to the larger ones.

Table 5. Fit statistics for model 1 and model 2.

	Model 1				Model 2				Model 2 | Model 1			
	L^2	df	p	BIC	L^2	df	p	BIC	ΔL^2	Δdf	p	ΔBIC
HU	38.1	42/14	.643/.000	-247.8	23.3	30/2	.802/.000	-180.9	14.8	12	.253	66.9
CS	38.8	42/19	.614/.005	-239.9	25.5	30/7	.700/.000	-173.5	13.3	12	.348	66.4
USA	25.6	42/14	.978/.030	-241.9	14.7	30/2	.991/.000	-176.3	10.9	12	.538	65.6

Both models fit well in case of each country. Comparing model 1 to model 2, reduction in model fit is not significant even in the case of socialist countries, hence the more parsimonious model 1 is to be chosen. That is, the arrow $E'O$ may be omitted from the graph.

In the next step the path model parameterized in Table 3 is fitted: the question is whether higher than first-order effects of Model 1 are relevant in the status attainment process. According to Table 6, the path model fits well in each country. The BIC prefers the path model to model 1, so does L^2 with the only exception of Hungary. Since the p value pertaining to the difference in L^2 is nearly 0.05 for Hungary, we choose the more parsimonious and well-fitting path model for each country.

Table 6. Fit statistics for model 1 and its path version.

	Model 1				Path model				Path model | Model 1			
	L^2	df	p	BIC	L^2	df	p	BIC	ΔL^2	Δdf	p	ΔBIC
HU	38.1	42/14	.643/.000	-247.8	53.9	50/22	.327/.002	-286.4	15.8	8	.045	-38,6
CS	38.8	42/19	.614/.005	-239.9	41.9	50/27	.786/.006	-289.8	3.1	8	.926	-49,9
USA	25.6	42/14	.978/.030	-241.9	28.0	50/22	.995/.150	-290.4	2.4	8	.965	-48,5

The estimated free parameters (listed in Table 3) pertaining to the arrows describe the country-specific mobility patterns (Table 7-Table 13). Recall that these values characterize average conditional effects, where the conditioning set is defined by the marginal within which the effects were computed. Greater absolute values denote stronger effects; positive/negative values increase/decrease the chance of falling into the given cell. The signs *, ** and *** denote significance at the 0.05, 0.01 and 0.001 level, respectively.

Table 7 presents the estimates for the effect pertaining to the arrow $E'O'$, which is the (total) $E'O'$ effect computed within the $E'O'$ marginal. According to the results, there exists a strong, statistically significant relationship between fathers' education and occupation. The association is weakest in the USA and strongest in Hungary. In the latter country, occupational classes are closed at top and bottom: the upper class is rather recruited from those with a degree, while the lower class is recruited from non-degree individuals. Czechoslovakia is closed at the top.

Table 7. Estimates of parameter $\lambda_{ij}^{E'O'}$.

$E'-O'$	USA			Hungary			Czechoslovakia		
	U	M	L	U	M	L	U	M	L
Coll	0.78***	-0.29**	-0.49***	0.98***	-0.18	-0.80***	0.97***	-0.54***	-0.43***

| No coll | -0.78*** | 0.29** | 0.49*** | -0.98*** | 0.18 | 0.80*** | -0.97*** | 0.54*** | 0.43*** |

There are two arrows pointing to education in the status attainment model: father's education and his occupation. Their estimated (direct) effects, controlled for each other, are shown in
Table 8 and Table 9. The *E'E* association (conditioned on and averaged over *O'*) is significant and positive; the weakest effect is observed in the USA, and the strongest in Hungary. One's chance of getting a degree is largely dependent on the father's education, even controlling for paternal occupation. The latter factor, father's occupation has a weaker but positive direct effect conditioned on father's education. The association is equally strong in the post-socialist countries and slight in the USA.

Table 8. Estimates of parameter $\lambda_{i*k}^{E'O'E}$.

E'-E	USA		Hungary		Czechoslovakia	
	Coll	No coll	Coll	No coll	Coll	No coll
Coll	0.26***	-0.26***	0.56***	-0.56***	0.39***	-0.39***
No coll	-0.26***	0.26***	-0.56***	0.56***	-0.39***	0.39***

Table 9. Estimates of parameter $\lambda_{*jk}^{E'O'E}$.

O'-E	USA		Hungary		Czechoslovakia	
	Coll	No coll	Coll	No coll	Coll	No coll
U	0.21*	-0.21*	0.30**	-0.30**	0.32***	-0.32***
M	-0.07	0.07	-0.02	0.02	-0.03	0.03
L	-0.14	0.14	-0.28***	0.28***	-0.30***	0.30***

Occupation is affected by paternal occupation and education; Table 10 and Table 11 present the estimates for these direct effects. In the USA, paternal occupation has a weak effect: the American seems to be the most open of the three societies. Occupational inheritance is more marked at bottom in Hungary and at top in Czechoslovakia. Again, the effect is stronger in the post-socialist countries. The same is true for the effect of education (Table 11). The association is weakest in the USA, where, contrary to the post-socialist countries, the recruitments into the middle and lower class are similar ($\lambda_{*22}^{O'EO} \approx \lambda_{*23}^{O'EO}$).

Table 10. Estimates of parameter $\lambda_{j*l}^{O'EO}$.

O'-O	USA			Hungary			Czechoslovakia		
	U	M	L	U	M	L	U	M	L
U	0.36***	-0.17	-0.19	0.23*	-0.03	-0.20	0.45***	-0.05	-0.40**
M	-0.09	0.09	0.00	0.22*	0.09	-0.31**	-0.24**	0.10	0.14
L	-0.27**	0.09	0.19*	-0.45***	-0.06	0.51***	-0.21*	-0.05	0.26**

Table 11. Estimates of parameter $\lambda^{O'EO}_{*kl}$.

E-O	USA			Hungary			Czechoslovakia		
	U	M	L	U	M	L	U	M	L
Coll	0.50***	-0.23	-0.26	0.85***	-0.24*	-0.61***	0.82***	-0.38***	-0.44***
No coll	-0.50***	0.23	0.26	-0.85***	0.24*	0.61***	-0.82***	0.38***	0.44***

In the USA, the effect of education on income is completely mediated by occupation: the direct effect is almost zero and insignificant (Table 12), while the effect of occupation is significant and strong (Table 13). The picture is different for the post-socialist countries, where education has a strong direct effect over occupation, but the effect of occupation is much weaker in Czechoslovakia. Another feature of the USA/post-socialist opposition is the pattern in the *OI* effect: in the USA, upper class individuals are much more likely to earn over the median income, but the middle and lower classes do not really differ from this aspect. In Hungary, the gap is less deep between the upper and middle classes, but the middle and lower classes are far from each other.

Table 12. Estimates of parameter λ^{EOI}_{k*m}.

E-I	USA		Hungary		Czechoslovakia	
	High	Low	High	Low	High	Low
Coll	0.12	-0.12	0.348***	-0.348***	0.257***	-0.257***
No coll	-0.12	0.12	-0.348***	0.348***	-0.257***	0.257***

Table 13. Estimates of parameter λ^{EOI}_{*lm}.

O-I	USA		Hungary		Czechoslovakia	
	Coll	No coll	Coll	No coll	Coll	No coll
U	0.51***	-0.51***	0.25***	-0.25***	0.13*	-0.13*
M	-0.26***	0.26***	0.12*	-0.12*	0.03	-0.03

| L | -0.26*** | 0.26*** | -0.36*** | 0.36*** | -0.16** | 0.16** |

The graph of Figure 3 sums up these results by representing the size of the effects by the width of the arrows.
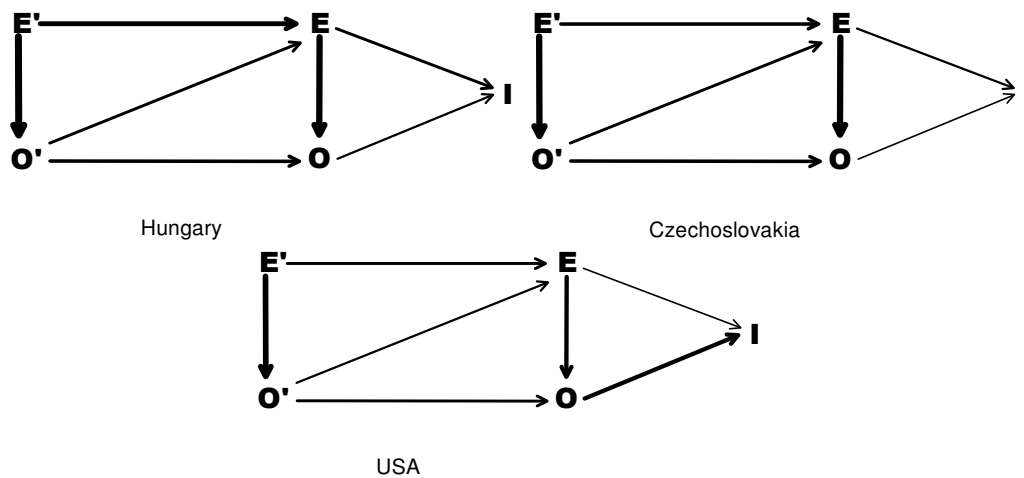
Figure 3. Results.

There are clear tendencies in the estimates:

1. a child with a lower status family faces more barriers to educational success in the post-socialist countries than in the USA (arrows *E'E* and *O'E*), and

2. in these countries, the recruitment process into occupational classes is more strongly affected by education (arrow *EO*), than in the USA;

3. contrary to the post-socialist countries, American income is strongly determined by occupation (a thick *OI* arrow) and

4. in the USA, occupation completely mediates the effect of education (a thin *EI* arrow); finally,

5. there is a deeper gap between the upper class and the other two classes in the USA than in the post-socialist countries, but the middle and lower classes are

closer to each other than their Eastern European counterparts (estimates of effect *EO* and *OI*).

Points 1, 3 and 4 imply that, in Treiman's sense, the American society is more developed than the Hungarian or the Czechoslovakian ones. Only point 2 contradicts to this conclusion, since in Treiman's theory the job market of a developed economy more markedly rewards abilities achieved through formal education. Finally, in Hungary and Czechoslovakia, policies to weaken intergenerational educational transmission (effect *E'E*) were rather unsuccessful.

References

Andersson SA, Madigan D and Perlman MD (2001) Alternative Markov properties for chain graphs. *Scandinavian Journal of Statistics*, 28(1): 33-85.

Bergsma W, Rudas T (2002) Marginal models for categorical data. *The Annals of Statistics* 30(1): 140-159.

Blau PM and Duncan OD (1967): *The American Occupational Structure*. New York: Wiley and Sons.

Boguszak M, Gabal I and Matějů P (1990) Ke koncepcím vývoje sociální struktury v ČSSR. *Sociologický časopis* 26(3): 168- 186.

Breen R and Luijkx R (2004) Conclusions. In: Breen R (ed) *Social Mobility in Europe*. Oxford: Oxford University Press, pp.383-446.

Cox DR and Wermuth N (1993) Linear dependencies represented by chain graphs. *Statistical Science*, 8(3): 204-218.

Cox DR and Wermuth N (1996): *Multivariate Dependencies – Models, Analysis and Interpretation*. London: Chapman & Hall.

Cox DR and Wermuth N (2001) Some statistical aspects of causality. *European Sociological Review*, 17(1): 65–74.

Drton M (2009) Discrete chain graph models. *Bernoulli* 15(3): 736-753.

Erikson R and Goldthorpe JH (1992): *The Constant Flux*. Oxford: Calderon.

Featherman DL and Hauser RM (1978): *Opportunity and Change*. New York: Academic Press.

Ganzeboom HBG; Treiman DJ and Ultee WC (1991) Comparative Intergenerational Stratification Research: Three Generations and Beyond. *Annual Review of Sociology*, 17(1): 277-302.

Goldthorpe JH, Payne C and Llewellyn C (1978) Trends in class mobility. *Sociology* 12(3): 441-68.

Goodman LA (1973) The analysis of multidimensional contingency tables when some variables are posterior to others: a modified path analysis approach. *Biometrika,* 60(1): 179-192.

Jonsson JO (1992): *Towards the Merit-Selective Society?* Stockholm: Swedish Institute for Social Research.

Kelley J (1990) The failure of a paradigm: Log linear models of social mobility. In: Modgil C (ed) *John Goldthorpe: Consensus and Controversy.* London, England: Falmer Press, pp.319-46.

Klimova A and Rudas T (2012) Coordinate free analysis of trends in British social mobility. *Journal of Applied Statistics* 39(8): 1681-1691.

Koster JTA (1999) On the validity of the Markov interpretation of path diagrams of Gaussian structural equations systems with correlated errors. *Scandinavian Journal of Statistics*, 26(3): 413–431.

Kupek E (2005) Log-linear transformation of binary variables: A suitable input for structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal* 12(1): 35-47.

Kupek E (2006) Beyond logistic regression: structural equations modelling for binary variables and its application to investigating unobserved confounders. *BMC Medical Research Methodology* 6(13)

Lauritzen SL (1996): *Graphical Models*. Oxford: Clarendon Press.

Lipset SM and Bendix R (1959): *Social Mobility in Industrial Society.* Berkeley: University of California Press.

Lipset SM and Zetterberg HL (1959) Social mobility in industrial societies. In: Lipset SM and Bendix R (eds): *Social Mobility in Industrial Society.* Berkeley: University of California Press, pp.11-75.

Marchetti GM and Lupparelli M. (2011) Chain graph models of multivariate regression type for categorical data. *Bernoulli* 17(3): 827-844.

Menard S (2010): *Logistic Regression: From Introductory to Advanced Concepts and Applications*. Thousand Oaks, CA: Sage.

Németh R and Rudas T (2013) On the application of discrete marginal graphical models. *Sociological Methodology*, to appear.

Rudas T, Bergsma WP and Németh R (2006) Parameterization and estimation of path models for categorical data In: Rizzi A and Vich M (eds) *COMPSTAT 2006 Proceedings in Computational Statistics*. Heidelberg: Physica, pp.383-394.

Rudas T, Bergsma WB and Németh R (2010) Marginal log-linear parameterization of conditional independence models. *Biometrika* 97(4): 1006-1012

Rudas T (1998): *Odds Ratios in the Analysis of Contingency Tables*. Thousand Oaks: Sage.

Treiman DJ (1970) Industrialization and social stratification. In: Laumann EO (ed) *Social Stratification, Research and Theory for the 1970s*. Indianapolis: Bobbs-Merill, pp.207-34.

Treiman DJ, Ganzeboom HBG (2000) The Fourth Generation of Comparative Stratification Research. In: Quah S and Sales A (eds) *The International Handbook of Sociology*. Thousand Oaks, CA: Sage, pp. 123-150.

Vallet L-A (2001) Forty Years of Social Mobility in France: Change in Social Fluidity in the Light of Recent Models. *Revue française de sociologie* Supplement: An Annual English Selection, 42: 5-64.

Whittaker J (1990): *Graphical Models in Applied Multivariate Statistics*. New York: John Wiley & Sons.

Wolbers MHJ, Luijkx R and Ultee W (2011) Educational attainment, occupational achievements, career peaks. *European Societies* 13(3): 425-450.

Xie Y (1992) The log-multiplicative layer effect model for comparing mobility tables. *American Sociological Review* 57(3): 380-95.

Zelterman D and Youn TI (1992) Indicator models for social mobility tables. *Computational Statistics & Data Analysis*, 14(1): 39-53.