Incorporating Information into Models: A Methodology and an Estimation Using Czech Voucher Privatization Data

Jan Hanousek and Radek Laštovička
CERGE and EI

January 1994

Abstract

Transition economies and the underlying processes that have occurred during the last four years in the eastern European countries are in the spotlight of leading economists. Unfortunately, the theory of economic transition can hardly explain these processes; even the intuition is often missing. On the other hand, there are large data sets which record the development of different transition economies. The question is how to utilize the information hidden in the recorded data and properly reflect the evolution of the processes.

This paper presents a sequence of steps which enable the incorporation of information about the behaviour of agents into statistical models when using a large data set with many variables. We assumed that nothing concrete could be said in advance about behaviour; only the given rules of the game are utilized.

The proposed methodology is followed by its application to data from the first wave of voucher privatization in the Czech Republic. Surprisingly, we found that the seemingly inexperienced population behaved rationally. In addition, several intuitive, easily understandable dynamic properties of the voucher privatization scheme have been identified.

Abstrakt

Ve středu pozornosti ekonomů jsou dnes transformujícíse ekonomiky východní Evropy a procesy v nich probíhající v posledních čtyřech letech. Bohužel, ekonomická teorie dnes ještě zdaleka není schopna tyto procesy dostatečně vysvětlit. Dokonce často schází i samotná intuice. Na druhé straně, dnes se už začínají objevovat rozsáhlé datové soubory zachycující jednotlivé transformační procesy. Otázkou je jak odkrýt informaci obsaženou v takovýchto datech a zároveň správně zachytit podobu zkoumaných procesů.

Autoři článku prezentují poslopnost kroků jak do modelu zabudovat dostupnou informaci o chování ekonomických subjektů pro případ velkých datových souborů zachycujících mnoho proměnných. Prezentovaný postup neklade žádné dodatečné předpoklady na chování účastníků, pouze využívá pevně daná pravidla procesu.

Metodologie je doplněna o praktický příklad postupu ukázaný na datech z první vlny kupónové privatizace českých podniků. Překvapivým zjištěním je vysoce racionální chování investorů nemajících žádné zkušennosti s kapitálovým trhem. Také poznatky o dynamickém chování sytému jsou jasně a jednoduše interpretovatelné.

The authors gratefully acknowledge the assistance of Dana Sapatoru in editing the English.

I. Introduction

The current transformation processes in central and eastern European countries can be an important impulse for economic theory. A wide variety of policy approaches have been adopted here and different economic developments have been accomplished. Many economic theories have been confronted with practice. Some results of the economic transformation have been recorded and are publicly available. Newly established capital markets are well-mapped, firm level data for the periods before and after privatization exist for several countries and different privatization methods, and many monopolies have started to function in a market-oriented economic environment under different regulatory mechanisms or even without state regulation. Huge datasets have therefore been gathered, but only a few are exploited by economists.

However, these data have some features in common. First, they usually contain a large number of variables, some of which are redundant or not closely related to the topic of interest. Second, consistent data exist only for a very small number of periods. And third, there is a lack of theory and sometimes even intuition to facilitate the understanding of the underlying economic processes.

The paper presents a methodology of dealing with such data, applied to the Czech voucher privatization scheme, which can be considered an exceptional realization of a closed market economy with a finite horizon (millions of active economic agents and direct convergence to equilibrium). It is a very simple system, characterized by strict and clear rules, publicly known and unique prices for each good in each period, a finite and constant number of participants and goods, a specific currency, and five distinct periods. The data set used contains prices, supply of and aggregate demand for each good for each of the five periods.

Briefly, the voucher privatization system works as follows: hundreds of firms are selected to be fully or partly privatized through vouchers; every adult citizen is allowed to enter the voucher privatization scheme by paying a small administrative fee; prices of shares are announced and participants express their demands by directing their vouchers to firms. If the aggregate demand for a particular firm is higher than the supply of shares, demands are cancelled and the price of shares is raised for further bidding in the next round. Otherwise, shares are exchanged for vouchers, and the remaining shares are offered at a lower price. Within five rounds this system succeeded in the exchange of 98% of vouchers for shares and was discontinued.[1]

[1] We are not going to restate the rules of voucher privatization in Czechoslovakia here. For those who are not familiar with these rules, we strongly recommend reviewing the subject in some of the publications listed in our references before reading any further.

In contrast with the very limited literature based on analysis of data from post-communist countries, there are quite a few papers which examine voucher privatization in the former Czechoslovakia. Some interesting results are published in The Privatization Newsletter of Czechoslovakia. Švejnar and Singer (1993) examine the differences between Czech and Slovak data, and the differences among small, individual, and large institutional investors. They estimate demand and supply separately, in each round, using a Tobit model. The main conclusion of their analysis is that investment behaviour is surprisingly rational for an economy which was led by Communists for forty years.

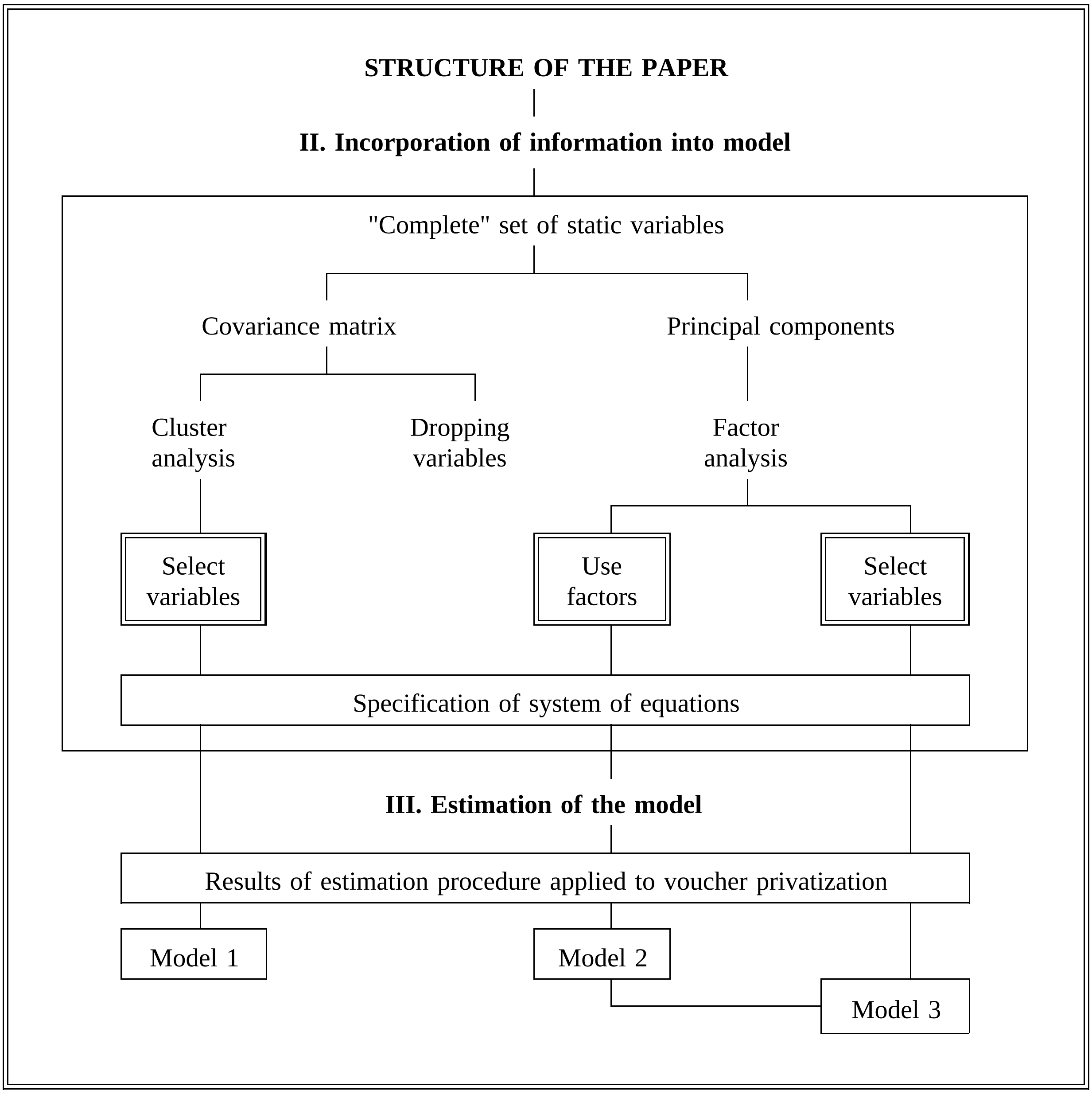
As regards the literature on closed economies, we found the intertemporal substitution theory to be the subject most related to our topic of analysis. Models of the substitution of current consumption for consumption in later periods were first introduced by Samuelson (1958). They are presented in standard modern Macroeconomic textbooks, such as (Blanchard, Fischer (1989). The similarity to voucher privatization consists of no influence from the outside world (a closed economy) and the common assumption that there is some given wealth to which each consumer is entitled at the beginning of the time horizon. The initial wealth is then spent during several subsequent periods.

Rather than modifying a standard model and estimating its parameters, we focused on putting forward a possible sequence of steps to extract as much information as possible about the behaviour of economic agents. We attempted to utilize all the information initially incorporated in the data and all that we knew about the system. Very few arbitrary assumptions were made about the behaviour of the agents.

We are interested in the dynamic behaviour of an equation system consisting of the supply, the price and the demand, as opposed to the study and estimation of the fixed and random effects of demand and price equations. Three models differing in static variables in demand functions are presented. Using estimated models we construct several hypotheses and test them. Namely, we show that the price setting applied by the price authority changes for every round. The models indicate that investors' behaviour changes too, and that the effects of the initial firm characteristics (static variables) are constant over time with relatively low levels of significance for the decision-making process.

Section II justifies the choice of the model specification and formulates the regression equations. Methods of cluster and factor analysis are used to summarize the effects of a large number of static variables. Section III reports the results of the estimations and hypotheses testing. Some concluding remarks are presented in section IV. A somewhat more detailed structure of the remaining part of the paper is shown, in diagram form, in Figure 1.

Figure 1. The structure of the paper



II. Incorporation of information into the model

There are several types of initial information - recorded data, the rules of the process, and economic theory and intuition. Typical features of economies in transition are large datasets with many variables, rules usually somehow different from the rules of similar processes in developed economies, and a lack of theory and intuition. While rules, theory and intuition determine the appropriate functional form, data sample properties lead to the choice of the method of estimation. Thus, the functional form is specific for each case and no general methodology exits. Therefore a deeper analysis is necessary. For illustrative purposes, we use the voucher privatization process.

Since almost no theory of voucher privatization exists, we focus our examination of three crucial dimensions: demand, supply and price. We believe that the functional forms describing these variables should be chosen according to specific rules limiting the behaviour of agents in the economy, rather than on our intuition, which is very weak in this case. Our aim is to show how to reflect all the known rules and, at the same time, to impose as few additional assumptions as possible.

While the most popular economic models are based on demand and supply, which determine price, in voucher privatization supply was exogenous. The seller (the state) set the prices, and the buyers (investors) expressed their demands based on the knowledge of both supply and price. Therefore, price did not have the of a variable characterizing equilibrium. This role was played by demand, which reflected the market equilibrium given price and supply. Our model has three basic equations - supply, price, and demand.

The **supply** side of voucher privatization was established in the same way for each round. The number of shares of each firm in each round - the supply - was either the supply in the previous round (in the case of excess demand) or the supply in the previous round minus demand in that same round.

$$Supply(t) \equiv Supply(t-1) - Sold(t),$$

The **price** - the number of voucher points per share - was set by the Price Committee according to an unknown algorithm based on the ratio of demand to supply in the previous round and some other variables. However, we do not know the exact algorithm used by the Price Committee for determining prices. They announced that they adjusted prices mainly on the basis of the ratio of demand to supply in the previous round. They also promised to exclude any potential individual interventions from the price setting process as much as possible and use only the results of previous rounds.

Thus, the price for a given round can be expressed with the help of a function such as:

$$Price(t) = F\left(\frac{Demand}{Supply}(t-1),\ Price(t-1),\ \boldsymbol{U}\right),$$

where $\boldsymbol{U}$ stands for other unknown explanatory variables.

Demand, i.e. the number of shares demanded by investors, is the most interesting equation to be estimated. However, only publicly known information can be used, since no data about the private information and preferences of the investors is available. As Švejnar and Singer pointed out, the decisions of investors in voucher privatization are surprisingly rational, in other words, they had evidently collected information. Therefore, we believe that the crucial role in demand is played by the demands expected in the previous round (well-known theory of steady investors), current price and supply, and some starting values.

$$Demand(t) = G\left(Supply(t),\ Price(t),\ Demand(t-1),\ \boldsymbol{V}\right),$$

where $\boldsymbol{V}$ stands for other dynamic and static (starting) explanatory variables.

Since the initial (starting) information is incorporated in a very large number of variables, the problem of selecting only some of them arises. Incorporating a large number of variables directly into the regression equations would lead to high multicollinearity.

However, there exists a generally applicable methodology of dealing with this problem, which is very common in the study of economies in transition: suppose we have a dataset reporting many variables, some of which do not play any role in the behaviour of agents, some are directly useful and others could be useful only after rearrangement. Suppose also that we do not have enough *ex-ante* intuition to select those which should be taken into account. Under these circumstances, we propose the following sequence of steps: first, we rearrange the variables in a form which can be clearly interpreted in the decision process of the actors. This means computing indexes, rescaling, etc.; second, we use correlations to reduce the number of variables. The simple way is to select one variable from a group of strongly correlated variables. A more advanced procedure is to use a summarization method based on correlations. Both summarization methods proposed below are increasingly popular in econometrics. Moreover, cluster analysis as well as factor analysis (the method of principal components) is widely used in other disciplines, for example, sociology.

Cluster analysis is a sequential process.[2] Every step consists in finding the two variables which are most heavily correlated and merging them into one new cluster. There are two methods for selecting representative variables through cluster analysis. The first one utilizes the fact that this technique does not only merge variables, but it also orders them. The ordering is based on their similarity, so that the first one is the most similar to the second one and most different from the last one in the ordered sequence. We can pick every n-th variable starting with the first one and ending with the last one. Through this procedure the entire spectrum of the effects of variables is covered, but their particular grouping into clusters is ignored.

The second method of selection is based on picking one variable for each cluster (usually the variable from the centre of the cluster). The number of clusters is not selected *ad hoc*, but on the basis of the agglomeration schedule of the cluster analysis. The agglomeration schedule shows a chosen measure of distance (it is common to use the squared Euclidean distance). Jumps in this distance then indicate a greater dissimilarity between merged clusters, i.e. an "optimal" number of clusters.

The advantage of clustering is that it provides a good graphic interpretation of results. However, its disadvantage is the sequentiality: once a cluster is formed, it cannot be split; it can only be combined with other variables and clusters. This property induces the merging of even quite unrelated variables.

Factor analysis[3] is probably a more useful tool. It does not suffer from the difficulties described above and, in addition, eliminates problems of multicollinearity. The method of principal components extracts a chosen number of factors, which are orthogonal and can be represented as a linear combination of initial variables. For every prespecified number of factors, the similarity of the effect of the initial variables is examined from the beginning, regardless of the computations for another prespecified number of factors.

---

[2] Many examples of the use of cluster analysis can be found, for instance, in Romesburg (84). In Anderberg (73), a deeper discussion of the calculation of measures of distances and similarities is provided.

[3] For more details about factor analysis see Jonassen, Peres (60), Stoetzel (60), and Kaiser (74).

Since devised factors usually do not provide a good interpretation, it is necessary to rotate them. Orthogonality is not damaged by the rotation procedure.[4] The rotation gives more weight to some variables in a factor and less to others. A desirable result of the method of principal components with subsequent rotation is a small number of factors with reasonable interpretation. Note that any factor is interpreted according to the meaning of the linear combination that forms it.

There are two ways to decrease the number of variables through factor analysis. It is possible either to employ factors directly, or to pick the variables with the highest contribution to the formation of factors as the representative variables. The use of devised factors directly makes sense only if they are easily interpretable.

In our concrete case, we normalize some variables in order to compare firms of different size, and we use growth indexes for some variables. The information published for each firm includes the following data: number of employees, profit, debt and sales, separately for 1989, 1990 and 1991. However, these numbers have rather higher informative value if we normalize them using the total number of shares of a firm (TNS).[5] Rather then using separate numbers for each year, we operate with the number for 1991, and with the changes between 1991 and 1990, and between 1990 and 1989, which captures trends. The list of variables is given in Table A.1. in the appendix.

Additional initial information available is the region, industry and percentage of shares transformed to non-voucher shareholders. In order to convey this, we constructed eight regional and ten industrial dummy variables, and eight variables

---

[4] The latent variables - principal components - are orthogonal, but usually they are not easily interpretable. The next step, thus, in the investigation is to use factor analysis (see Harman (67)). We rotate the principal component (i.e. multiply it by a rotation matrix) the to get the so-called factors. Because of the definition of the rotation matrix, orthogonality still holds and, in addition, any linear combination of variables forming a factor can then be interpreted.

[5] There are four possible measures in the database which might reflect the size of a firm: number of employees, book value, sales, and number of emitted shares. For our purpose, the number of employees is misleading for large holdings, because they report only those employed in the coordination centre. Sales are useless for financial institutions and misleading for relatively small trading firms with extremely large sales. For the most part of firms, the number of shares closely corresponds to the book value, but in some cases it is adjusted based on estimates of the real value of a firm. Therefore, we use the total number of shares of a firm as a normalization variable.

Note that correlations between variables are not changed by normalization. It follows that the cluster analysis is the same for normalized variables as before normalization, while factor analysis is not equivariant for this operation.

reporting the percentage of shares transferred to non-voucher shareholders. The last eight variables are not dummy variables in a pure sense, but in practice they are very similar, since their value is mostly zero and in some cases lies between one and ninety-nine. Therefore, it is natural to analyze them together with the other dummy variables. The list is given in Table A.2. in the appendix.

Because it is easier to interpret clusters and factors, we run both procedures separately for real value static variables and dummy variables. The results of the cluster analysis are presented in Figures 2 and 3.

The dendrograms show how the cluster is formed and the plot coefficients rescaled (scale 0-25). To understand the dendrogram note that in Figure 2, the variables ΔP90 and ΔI90 are the closest and are thus merged into the first cluster; the nearest variable (to this cluster) is P91, and so forth.

Figure 2. Hierarchical cluster analysis - real value static variables

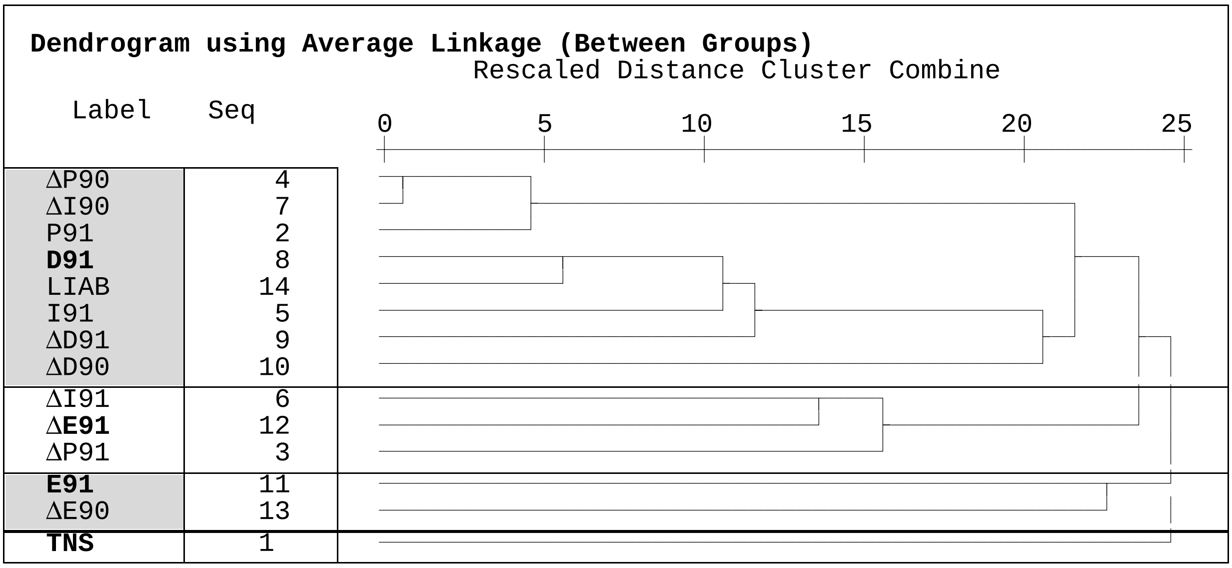
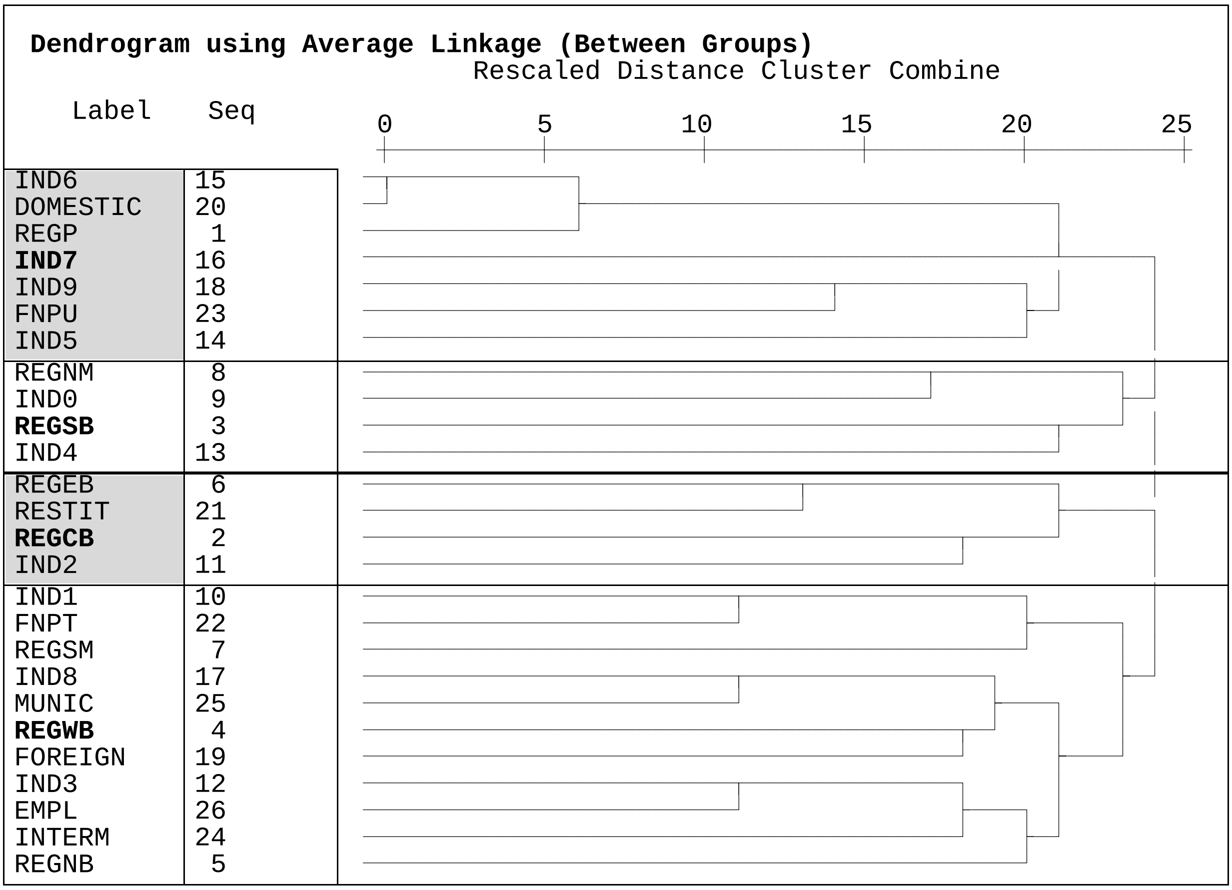


Figure 3. Hierarchical cluster analysis - dummy variables



The number of clusters selected according to squared Euclidean distances is reported in the agglomeration schedule. For instance, for the real value static variables jumps in distances indicate between two and five as a reasonable number of clusters (see Appendix, Table B.1., column "Coefficient"). We decided to select four. Similarly, we chose four clusters as a reasonable number for the dummy variables. Representative variables resulting from the analysis are selected from the centre of each cluster (variables in bold in Figures 2 and 3).

We run the factor analysis for several reasonably small numbers of factors. We try to find the factors which can be easily interpreted and, at the same time, explain a sufficient amount of variation. Finally, we considered four factors for static variables (75% of variation of all static variables) and three factors for dummy variables. The results are reported in Tables 1 and 2 below.

Table 1: Rotated factor matrix for real value static variables:

Variable	FSTAT1	FSTAT2	FSTAT3	FSTAT4
ΔP90	.99611	-.01860	.00882	-.00199
I91	.99538	.01394	.01073	-.00287
ΔI90	.99477	.02856	.00877	.02083
ΔP91	-.99208	.01654	.00948	-.00177
ΔI91	-.94150	.13826	.07125	.00893
P91	.84423	.14723	-.03318	-.02055
D91	-.00163	.91917	-.01313	.16555
LIAB	.01898	.87231	-.08629	.22008
ΔD91	.00102	.81890	.02768	-.36870
E91	.01874	.07797	-.78373	.01245
ΔE91	-.02715	.03957	.67981	-.10876
TNS	.00142	-.01401	.40639	.02379
ΔD90	-.00189	.20246	.08262	.87171
ΔE90	-.00495	-.04770	-.10267	.53298

Table 2: Rotated factor matrix for dummy variables:

Variable	FDUMMY1	FDUMMY2	FDUMMY3
IND6	**.72642**	-.03150	-.09721
REGP	**.71562**	-.15593	-.11887
DOMESTIC	**.67741**	-.07616	-.02845
RESTIT	.18553	-.05667	-.09432
REGSM	-.15244	-.06320	.08110
INTERM	.14990	.10528	.13075
REGSB	-.11180	-.08063	-.01883
IND7	.10090	-.09898	-.09861
IND0	-.05849	.02688	-.05048
IND3	-.09202	**.70729**	.04562
EMPL	.11106	**.55662**	.14751
FNPU	-.12597	.37131	.26766
REGWB	-.08181	.30806	-.13580
IND8	.05641	.27774	.06830
REGNB	-.10145	.22689	-.04462
IND1	-.09756	.10525	-.02681
FOREIGN	.06020	.08460	.01166
IND2	-.20624	-.50050	**.68669**
IND4	-.24012	-.18456	**-.65124**
REGEB	-.07793	-.06497	.40081
REGNB	-.25084	-.05984	-.31069
MUNIC	.03597	.19686	.28475
FNPT	-.05994	.00488	.27805
IND5	.01498	-.05892	-.21892
REGCB	-.06046	-.05831	.18584
IND9	.03214	-.09344	-.12927

Let us try to interpret the above factors:[6]

FSTAT1 represents sales and profits variables. Sales and profits are very highly correlated. Therefore we can call this factor the profit factor.

FSTAT2 can be interpreted as the debt factor, because the main role is played by two debt variables and liabilities.

FSTAT3 is denoted by two variables representing capital facilitation of labour (employees per share) and TNS. All these variables are equivalent to the size of a firm. Even ΔP91 is, since the largest firms have the greatest dismissal rate. For these reasons, we call FSTAT3 the size factor.

FSTAT4 we call the "before transformation trend factor." The real pro-market changes in the Czech economy started at the beginning of 1991, which means that the two-year lagged trends probably reflect the tendencies which firms adopted under the communist system.

Finding a good interpretation of factors representing dummy variables is always difficult. However, if we consider only the variables with the highest coefficients of contribution (it is standard to consider coefficients higher than 0.5), we get some reasonable combinations of variables:

FDUMMY1 is explained mainly by trade, the Prague region, and the high involvement of Czech capital. All these are typical for Prague. We call the "Prague dummy factor" FDUMMY1.

FDUMMY2 reflects the role of light industry and employee ownership. In fact, the proportion of employee ownership in light industry is rather higher than in other industries. Therefore we can characterize FDUMMY2 as the light industry dummy factor.

FDUMMY3 is formed mainly by the indicator for heavy industry together with the opposite (a negative sign is the coefficient) to construction. We call it the "heavy industry dummy factor."

Having solved the problem of too many initial variables, we can now concentrate on the precise formulation of the regression equations. Since we do not know much about the system, we try to focus on its specific features determined by the rules. This means that we are not concerned with the start-up of the process, but with its dynamics (we thus lose the first observation to create lagged dynamic variables).

---

[6] Positive coefficients mean that the variables have the same effect, while negative coefficients imply that the effect of one variable is in the opposite direction to that of the other variable.

As is standard, we approximate an unknown functional form of the price equation by a polynomial function consisting of the explanatory variables raised to different powers. Allowing for quite a "wild" path of prices, we use a polynomial function of the fourth power. The explanatory variables available for the price equation are: the lagged ratio of demand and supply to the powers 1 to 4; supply; lagged price (an autoregressive function of the first degree); the percentage of a firm's shares directed to voucher privatization, and the percentage of shares directed to voucher privatization which remain to be invested.

While supply and price depend exclusively on dynamic explanatory variables, demand could reflect the initial information. The dynamic explanatory variables which we consider are: supply, price, the percentage of shares sold up to a certain round, lagged demand, and the ratio of demand to supply in a previous round. We represent the static explanatory variables via the summarization methods presented above.

There data suffers from significant heteroscedasticity. To alleviate this problem, we apply a logarithmical transformation which helps in two ways: it reduces heteroscedasticity, and the prices after logarithmization are not sensitive to overall price levels varying across rounds.[7]

In order to avoid the problem of multicollinearity, we checked the correlations between explanatory variables for all the equations to be estimated. Because of the high correlation for some rounds with the logarithm of supply, we excluded from the price equation the logarithm of the percentage of shares directed to voucher privatization which remain to be invested. Similarly, we excluded from the demand equation the logarithm of the lagged ratio of demand to supply (highly correlated with the logarithm of price).

---

[7] There was a significant decline in the average price of shares in later rounds. In the first round, prices were 33.3 points per share, while in the fifth round the average price declined to about 18 points per share. The relationship between nominal prices for different rounds after logarithmization is equivalent to the relationship between real prices and the shift is absorbed in the constant term. This is the main reason why we do not need to transform nominal prices into real prices.

In spite of having time and cross-sectional data, we do not apply the panel approach for estimation.[8] Thus, we use the full information maximum likelihood method of estimation. We considered this approach to be the best to emphasize the general features of the system (something like a macro view), regardless of the fact that the panel approach is more powerful for studying individual firms.

The possibility of substituting current spending for future spending leads us to estimate all equations simultaneously. It is the way to model, in a closed economy with a finite time horizon, the fact that decision makers always have in mind their forthcoming actions.

Due to technical reasons, we omit from our analysis all firms which were fully sold in voucher privatization.[9] The final dataset contains 791 firms out of 987 Czech firms using vouchers, at least partly, for their transformation.

Finally, the model consists of three basic equations - supply, price and demand, which we estimate simultaneously for four rounds using the full information maximum likelihood method. It yields a dynamic system of twelve log-linear equations of the following form:

SUPPLY

$$\log(S_t) = ED_{t-1}*\log(S_{t-1}) + (1 - ED_{t-1})*\log(S_{t-1} - D_{t-1}),$$

PRICE

$$\log(P_t) = ß_{t0} + ß_{t1}\log(DS_{t-1}) + ß_{t2}\log(DS_{t-1})^2 + ß_{t3}\log(DS_{t-1})^3 + ß_{t4}\log(DS_{t-1})^4 +$$
$$ß_{t5}\log(S_t) + ß_{t6}\log(P_{t-1}) + \upsilon_t,$$

8 In this case, we apply the approach of Bhargava, Sargan (1983) to cover a short time period (5 rounds). It is well known (see e.g. Nerlove (1967, 1971) and Nickel (1981) among others) that standard panel data techniques lead to very high bias. Using Nickel's results, biases for autocorrelation coefficients for 5 periods are as follows:

	ρ (autocorrelation coeff.)				
	0.0000	0.2000	0.4000	0.6000	0.8000
bias	-0.2000	-0.2483	-0.3020	-0.3618	-0.4279

9 In the case of a fully-sold firm's supply, demand and prices in later rounds are missing. Missing values either cannot be properly handled by the statistical software or they greatly complicate the modelling.

DEMAND

$$\log(D_t) = \alpha_{t0} + \alpha_{t1}\log(PS_t) + \alpha_{t2}\log(S_t) + \alpha_{t3}\log(P_t) + \alpha_{t4}\log(D_{t-1})$$
$$+ \alpha_{t5}(\text{STATIC VARIABLES}) + \varepsilon_t,$$

where

S_t	supply of shares at round t;
P_t	price of shares at round t (measured in voucher points);
D_t	demand for shares at round t;
ED_t	0-1 indicator for large excess demand for shares at round t (0 if at least some demands satisfied, 1 if all demands cancelled);
DS_t	demand supply ratio at round t;
PS_t	percentage of shares sold up to round t;

Note that the lagged price is excluded from the price equation for round 2 because all prices in the first round are equal. Otherwise we would partition the constant term into two meaningless parts.

In our sample, the demand for shares is zero only for a single firm for a single round. Ignoring this case, we have only positive observations of demand, which allow us to take logarithms.

III. Estimation of the Model

In this section, we provide the results of the estimation of three models which differ according to the set of initial static variables and the demand equations. The model which derives from the cluster analysis consists of the supply and price equations presented in section II. and a demand equation of the following form:

DEMAND (MODEL 1):

$$\log(D_t) = \alpha_{t0} + \alpha_{t1}\log(PS_t) + \alpha_{t2}\log(S_t) + \alpha_{t3}\log(P_t) + \alpha_{t4}\log(D_{t-1}) + \alpha_{t5}D91 + \alpha_{t6}\Delta E91 + \alpha_{t7}E91 + \alpha_{t8}TNS + \alpha_{t9}IND7 + \alpha_{t10}REGSB + \alpha_{t11}REGCB + \alpha_{t12}IND3 + \varepsilon_t$$ [10]

We estimated a system of simultaneous equations for t=2, ..., 5. Obviously, before estimating the model, we looked at the correlation among the explanatory variables, which was satisfactorily low. The model has a semilogarithmic form (some static variables are zero or even negative and cannot be logarithmized). This means that the coefficients of the dynamic variables can be naturally interpreted as elasticities, while the coefficients of the static variables cannot be understood in this manner.

We expected to obtain the following results: in the price equation, we anticipated a strongly significant and positive coefficient of DS_{t-1}, and a decrease in the significance of the coefficient of DS_{t-1} for higher powers, indicating that DS_{t-1} is the crucial variable for price determination. We also believe we will find a negative coefficient of S_t: if we consider two firms equivalent in all respects except for the number of shares available in the market, the relatively scarce one obviously costs more. Furthermore, P_{t-1} should have a highly significant and positive coefficient reflecting the autoregressive character of price setting.

In the demand equation, we assume that PS_t will have a negative coefficient due to the effect of relative scarcity, not among different firms, but rather among supplies of shares of the same firm in different rounds. S_t and D_{t-1} were expected to have positive signs caused by the law of large numbers[11] and the autoregressive character of demand, respectively. We also believed that every

[10] The description of the initial variables (coefficients α_{t5} through α_{t12}) is found in the Appendix.

[11] Consider two firms, identical except for the number of shares offered for sale. According to the law of large numbers, which is appropriate in the case of several million agents taking part in this process, the firm with the greater amount of shares available attracts more investors than the smaller one.

market economy should feature a negative relationship between demand and price (the coefficient of P_t has to be negative and significant). This property is sometimes referred to as "the demand law."[12]

Although we feel comfortable with the role of dynamic variables, we had no idea what the signs of coefficients of initial variables would be. It could have been, for instance, that their entire effect was captured in the first round (which implies that the role of the initial variables was fully captured by dynamic variables for subsequent rounds) or that the effect of some variables was overshot and the effect of others was undershot (resulting in negative or positive signs in a following round, respectively).

The estimated coefficients are reported in Tables 3 and 4 below.

Table 3. Price equations.

Variable	Round 2	Round 3	Round 4	Round 5
$CONST_t$	4.170 (40.2)	1.459 (7.0)	1.293 (14.7)	1.188 (15.9)
DS_{t-1}	.835 (33.2)	.595 (17.0)	.300 (8.7)	.420 (16.8)
DS_{t-1}^2	.135 (14.0)	.016 (.8)	.077 (5.5)	.044 (3.8)
DS_{t-1}^3	-.023 (-14.1)	-.034 (-4.9)	.008 (.9)	.003 (1.3)
DS_{t-1}^4	-.005 (-10.9)	-.004 (-1.8)	-.002 (-.7)	.005 (.4)
S_t	-.068 (-6.6)	-.223 (-14.8)	-.097 (-14.7)	-.063 (-11.6)
P_{t-1}		1.269 (33.8)	.953 (53.9)	.926 (61.2)
R-squared	**.905**	**.900**	**.970**	**.967**

12 See, for example, Hyman (93).

Table 4. Demand equations

Variable	**Round 2**	**Round 3**	**Round 4**	**Round 5**
$CONST_t$	3.427 (6.3)	-.012 (-.0)	.202 (.3)	-.438 (-.6)
PS_t	.001 (.0)	-.188 (-2.5)	-.143 (-1.5)	-.256 (-2.4)
S_t	.359 (8.3)	.740 (15.4)	.570 (16.9)	.626 (26.4)
P_t	-.711 (-10.2)	-.519 (-10.7)	-.195 (-4.3)	.045 (.7)
D_{t-1}	.529 (16.0)	.433 (9.3)	.473 (12.7)	.440 (13.5)
D91	.029 (1.6)	.034 (1.2)	.001 (.0)	.023 (.6)
ΔE91	-1.574 (-.1)	4.196 (.3)	2.336 (.2)	-5.108 (-.5)
E91	-18.364 (-1.9)	1.732 (.2)	-10.089 (-1.4)	-5.777 (-.8)
TNS	-1E-8 (-.2)	3E-8 (.6)	-1E-8 (-.1)	-1E-9 (.0)
IND7	-.095 (-.3)	-.172 (-.7)	-.260 (-.9)	.025 (.1)
REGSB	-.013 (-.1)	-.053 (-.4)	.163 (1.6)	-.011 (-.1)
REGCB	.030 (.3)	.067 (.7)	.009 (.1)	-.147 (-2.0)
REGWB	.153 (1.5)	.041 (.4)	.052 (.6)	.072 (.9)
R-squared	**.754**	**.845**	**.925**	**.920**

Note that the estimated coefficients, as well as the t-statistics, correspond closely to our expectations. The single serious deviation is the violation of the demand law in the fifth round. A straightforward interpretation for this is that the investment points would have had no value after the fifth round; in this round, therefore, it was pointless for investors to look at the price. They preferred instead to spend all the points to which they were entitled, which lead to a slightly positive (although insignificant) coefficient of price in the demand equation.

On the basis of the above model, we formulated and tested several hypotheses. We think that, from the investors' point of view, not only was the method of price determination extremely important, but also the stability of the price algorithm used by the Price Committee was significant. The hypothesis we formulated is this: "The price equation was the same for all periods." The test fully rejected this hypothesis. Furthermore, an even stronger hypothesis was rejected. There are no two rounds with the same price equation. The price setters, it is very probable, changed their behaviour in each round.

For the demand side of the system, we formulated a hypothesis similar to the one for the price equation: "All demand equations are the same." This hypothesis was also rejected, as well as the hypotheses concerning the equivalence between any two demand equations. The conclusion is that investors, on average, also changed their behaviour from round to round. Probably, some "learning by doing" took place during the process.

Looking at the estimated coefficients and their t-statistics, we also formulated some general hypotheses concerning the role of static variables in the demand equation: "The following variables may be omitted from the demand equation: ΔE91, TNS, IND7, REGSB, REGCB, and the variables D91, E91 and REGWB have coefficients constant in time". This hypothesis can not be rejected.[13] It says, in other words, that the effect of all starting variables is either negligible or constrained to be the same in all rounds.

Using factors given in section II., we can rewrite the demand equation as:

DEMAND (MODEL 2):

$$\log(D_t) = \alpha_{t0} + \alpha_{t1}\log(PS_t) + \alpha_{t2}\log(S_t) + \alpha_{t3}\log(P_t) + \alpha_{t4}\log(D_{t-1}) +$$
$$\alpha_{t5}FSTAT1 + \alpha_{t6}FSTAT2 + \alpha_{t7}FSTAT3 + \alpha_{t8}FSTAT4 +$$
$$\alpha_{t9}FDUMMY1 + \alpha_{t10}FDUMMY2 + \alpha_{t11}FDUMMY3 + \varepsilon_t,$$

where FSTAT1-FSTAT4 and FDUMMY1-FDUMMY3 are factors related to real value static variables and dummy variables, respectively (see Tables 1 and 2).

The correlation between dynamic variables and factors is extremely low in this case (remember that the correlation between factors is zero by definition). The model now has the same supply and price equations as before, only the estimated demand equations have changed. The estimated coefficients are reported in Appendix, Tables C.1 and C.2. Note that this model explains a higher proportion

[13] The joint test using Wald statistics yields W = 25.941; the corresponding p-value (χ^2 with 29 degrees of freedom) is P = 0.628.

of variance compared with the previous one. Factors represent static variables more thoroughly than the variables selected from clusters.

We tested a similar series of hypotheses to that for the model resulting from cluster analysis, but we found the estimated coefficients to be very similar. Thus, the interpretation of coefficients is the same as before. The tests lead us again to the conclusion that the price equations and demand equations were dissimilar for all rounds. One of the comprehensive hypotheses which restricts the role of initial variables represented by "latent" factors is: "From the FSTAT1, FSTAT3, FSTAT4 variables may be omitted from the demand equation, and the coefficients of PS_t and FSTAT2 are constant in time." We could not reject this hypothesis at any reasonable level of significance.[14] The accepted hypothesis states that the single factor of real value static variables which should be considered is the debt factor and its effect is constant over time.

For those who do not like to work with artificially constructed explanatory variables - factors - in the regression equation, we suggest the following step. We can consider factor analysis only as a tool for selecting a few explanatory variables. On the basis of the above tests, we can conclude that variables contributing mainly to FSTAT1, FSTAT3 and FSTAT4 do not have significant explanatory power. We replace FSTAT2 and the factors of dummy variables in the model with variables mainly contributing to them.

A problem arises with D91 and LIAB, which are heavily correlated. They both represent different kinds of debts of a firm. The easiest solution is to add them together.

The demand equation becomes:

DEMAND (MODEL 3):

$$\log(D_t) = \alpha_{t0} + \alpha_{t1}\log(PS_t) + \alpha_{t2}\log(S_t) + \alpha_{t3}\log(P_t) + \alpha_{t4}\log(D_{t-1}) +$$
$$\alpha_{t5}(D91 + LIAB) + \alpha_{t6}\Delta D91 + \alpha_{t7}IND6 + \alpha_{t8}REGP +$$
$$\alpha_{t9}DOMESTIC + \alpha_{t10}IND3 + \alpha_{t11}EMPL + \alpha_{t12}IND2 + \alpha_{t13}IND4 + \varepsilon_t,$$ [15]

[14] The joint test using Wald statistics yields W = 10.849; the corresponding p-value (χ^2 with 18 degrees of freedom) is P = 0.901.

[15] The description of the initial variables (coefficients α_{t5} through α_{t13}) can be found in appendix A.

The results of our estimation procedure are presented in Tables D.1. and D.2. in the Appendix. The estimated coefficients do not correspond so closely to those of the model with factors. In particular, it looks like the coefficient of $\log(PS_t)$ varied over time; in other words, it captured more of the variance because of the relatively pure properties of the model with factors representing initial variables in comparison with the model with factors directly. The hypothesis we tested was: "The coefficients of $\log(PS_t)$ are constant in time". This hypothesis was rejected on a 5% significance level.[16]

We also claimed the following about the properties of other coefficients of initial variables: "The variable D91 + LIAB can be completely omitted from the demand equation, while variables ΔD91, IND6 and IND3 can be omitted for rounds three to five, and variables DOMESTIC, EMPL and IND4 have time indifferent coefficients.". This hypothesis could not be rejected at a usual significance level.[17]

---

[16] The joint test using Wald statistics yields W = 10.446; the corresponding p-value (χ^2 with 3 degrees of freedom) is P = 0.015.

[17] The joint test using Wald statistics yields W = 29.017; the corresponding p-value (χ^2 with 22 degrees of freedom) is P = 0.144.

IV. Conclusions

The authors attempted to overcome the problem of limited understanding by using a quite general approach, making as few additional assumptions as possible. The derivation of the regression equations was based upon general rules of the process investigated, and related information was allowed to have an impact on the behaviour of participants.

The objective of including all the information available in the model brought about the difficulties of high multicollinearity and the loss of a large number of degrees of freedom. Therefore, we introduced two possible methods to summarize the effects of variables effects - cluster and factor analyses.

Moreover, a simultaneous estimation of the system of equations captures the possibility of foresight (or intertemporal substitution); in other words, simultaneous estimation takes into account the possibility of rational expectations. The influence of expectations on a future development can be easily shown by running a simple OLS regression instead of a simultaneous estimation and comparing the correlations of the error terms (see Appendix, Table E.1.; one can notice significant correlations across periods).

The results of the practical estimations presented in section IV and Appendices C and D allow us to conclude that the models should be considered as a simple but meaningful approximation of behavioral patterns in the voucher privatization process. From the three possibilities proposed for handling too many variables, the largest part of variance is captured by the model with factors.

References:

Anderberg, M. R. 1973. "Cluster analysis for applications". New York: Academic Press.

Bhargava, A., Sargan J.D.: "Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods" *Econometrica* 51 (1983), 1635-1659.

Blanchard, O. J.,Fischer, S. 1989. "Lectures on Macroeconomics". Massachusetts: MIT Press.

Hanousek, J. (1994). "Analysis of Demand Equation in the Voucher Privatization: Learning by Doing", CERGE, Charles University, Mimeo.

Harman, H. H. 1967. "Modern factor analysis", 2nd ed. Chicago: University of Chicago Press.

Hyman, D. N. 1993. "Modern Microeconomics", 3rd ed. Boston: Irwin.

Jonassen, C. T., Peres, S. H. 1960. "Interrelationships of dimensions of community systems". Columbus: Ohio State University Press.

Kaiser, H. F. 1974. An index of factorial simplicity. "Psychometrika 39": 31-36.

Kotrba J., Švejnar J. 1993. "Rapid and Multifaceted Privatization: Experience of the Czech and Slovak Republics". CERGE Working Paper No. 36, April 1993.

Lahiri, K., Schmidt, P.: "On the Estimation of Triangular Structural Systems" *Econometrica*, 46 (1978), 1217-22.

Marcinčin A. 1993. "Determinants of Demand and Prices in the First Wave of Voucher Privatization". Privatization Newsletter of Czechoslovakia, No. 16, May 1993.

Nerlove, M.: "Experimental Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross-Sections" *Economic Studies Quarterly*, 18 (1967), 42-74.

Nerlove, M.: "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross-Sections" *Econometrica*, 39 (1971), 359-387.

Nickel, S.: "Biases in Dynamic Models with Fixed Effects" *Econometrica*, 49 (1981), 1417-1426.

Romesburg, H. C. 1984. "Cluster analysis for researchers". Belmont, California: Lifetime Learnings Publications.

Samuelson, P. A. 1958. "An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money". Journal of Political Economy No. 66, December 1958.

Stoetzel, J. 1960. A factor analysis of liquidor preference of French consumers. "Journal of Advertising Research 1(1)": 7-11.

Švejnar, J., Singer M. 1993. "Using Vouchers to Privatize an Economy: The Czech and Slovak Case. CERGE Working Paper No. 36, April 1993.

Appendix

Content:

Table A.1. Real value static variables.

Name	Description of the variable
TNS	Total number of shares of a firm
LIAB	Liabilities in 1991 divided by TNS
E91	Number of employees in 1991 divided by TNS
ΔE91	Number of employees in 1991 minus number of employees in 1990 all divided by TNS
ΔE90	Number of employees in 1990 minus number of employees in 1989 all divided by TNS
P91	Profit in 1991 divided by TNS
ΔP91	Profit in 1991 minus profit in 1990 all divided by TNS
ΔP90	Profit in 1990 minus profit in 1989 all divided by TNS
I91	Sales in 1991 divided by TNS
ΔI91	Sales in 1991 minus the sales in 1990 all divided by TNS
ΔI90	Sales in 1990 minus the sales in 1989 all divided by TNS
D91	Debts to banks in 1991 divided by TNS
ΔD91	Debts to banks in 1991 minus debts to banks in 1990 all divided by TNS
ΔD90	Debts to banks in 1990 minus debts to banks in 1989 all divided by TNS

Table A.2. Dummy variables.

Name	Description of the variable
Ownership structure:	
FOREIGN	The percentage of shares for direct sale to a predetermined foreign owner
DOMESTIC	The percentage of shares for direct sale to a predetermined domestic owner
FNPT	The percentage of shares for transfer to the National Property Fund for a temporary period
FNPU	The percentage of shares for transfer to National Property Fund for an indeterminate period
RESTIT	The percentage of shares for transfer to restituents (former owners)
INTERM	The percentage of shares for transfer to an intermediator (usually a bank) which will sell the shares later
MUNIC	The percentage of shares for free transfer to municipalities
EMPL	The percentage of shares for sale to employees
Industries:	
IND1	Agriculture
IND2	Heavy industry and mining
IND3	Light industry
IND4	Construction
IND5	Transportation and telecommunications
IND6	Trade
IND7	R & D
IND8	Services, culture and education
IND9	Financial and state institutions
IND0	Others
Regions:	
REGP	Prague
REGCB	Central Bohemia
REGSB	Southern Bohemia
REGWB	Western Bohemia
REGNB	Northern Bohemia
REGEB	Eastern Bohemia
REGSM	Southern Moravia
REGNM	Northern Moravia

Table B.1. Agglomeration schedule of the cluster analysis (of real value static variables)

Stage	Clusters Combined Cluster 1	Cluster 2	Coefficient	Stage Cluster 1st Appears Cluster 1	Cluster 2	Next Stage
1	12	13	.975998	0	0	2
2	12	14	.930029	1	0	6
3	5	10	.923202	0	0	10
4	8	9	.920686	0	0	7
5	3	4	.912612	0	0	7
6	12	15	.872972	2	0	9
7	3	8	.866818	5	4	10
8	1	2	.864104	0	0	11
9	11	12	.798625	0	6	13
10	3	5	.797266	7	3	12
11	1	7	.747882	8	0	12
12	1	3	.646002	11	10	14
13	6	11	.505167	0	9	14
14	1	6	- .345251	12	13	0

Table B.2. Variation explained by a given number of factors (of real value static variables)

COMPONENT	ITERATIONS	EIGENVALUE	CUMULATIVE R-SQUARED
1	21	5.5511458	0.39651041
2	24	2.4052623	0.56831486
3	185	1.3054905	0.66156418
4	121	1.1747507	0.74547495
5	217	1.0207959	0.81838894
6	76	0.94668797	0.88600951
7	29	0.74788436	0.93942982
8	26	0.42702298	0.96993146
9	33	0.21679156	0.98541657
10	20	0.14161029	0.99553159
11	8	0.55936864E-01	0.99952708
12	25	0.34755851E-02	0.99977534
13	74	0.17698026E-02	0.99990175
14	4	0.13752548E-02	0.99999999

Table C.1. Price equation (MODEL 2):

Variable	**Round 2**	**Round 3**	**Round 4**	**Round 5**
$CONST_t$	4.184 (40.0)	1.387 (7.0)	1.275 (14.8)	1.198 (16.2)
DS_{t-1}	.838 (34.0)	.630 (18.8)	.292 (9.2)	.408 (17.5)
DS_{t-1}^2	.137 (14.8)	.014 (.7)	.078 (5.5)	.044 (4.1)
DS_{t-1}^3	-.023 (-14.5)	-.036 (-5.2)	.009 (1.0)	.003 (1.4)
DS_{t-1}^4	-.005 (-11.9)	-.004 (-1.9)	-.002 (-.7)	.000 (.4)
S_t	-.070 (-6.7)	-.215 (-15.4)	-.095 (-14.5)	-.063 (-11.5)
P_{t-1}		1.271 (35.9)	.951 (58.7)	.920 (64.9)
R- squared	**.906**	**.906**	**.970**	**.968**

Table C.2. Demand equations (MODEL 2):

Variable	**Round 2**	**Round 3**	**Round 4**	**Round 5**
$CONST_t$	4.514 (9.0)	.788 (1.2)	.440 (.7)	-.249 (-.3)
PS_t	-.084 (-2.2)	-.219 (-3.1)	-.162 (-1.8)	-.274 (-2.7)
S_t	.280 (6.8)	.708 (15.4)	.553 (16.6)	.616 (26.9)
P_t	-.916 (-12.0)	-.597 (-12.7)	-.217 (-4.2)	.018 (.2)
D_{t-1}	.591 (17.6)	.427 (9.9)	.477 (12.8)	.445 (13.8)
FSTAT1	-.012 (-.0)	.009 (.0)	.010 (.0)	-.014 (-.0)
FSTAT2	.041 (1.5)	.033 (1.0)	.003 (.0)	.014 (.3)
FSTAT3	.013 (.4)	-.003 (-.1)	.009 (.3)	.0022 (.1)
FSTAT4	-.037 (-.9)	-.014 (-.4)	.017 (.5)	.016 (.6)
FDUMMY1	.123 (3.9)	.090 (3.1)	.0092 (.3)	.030 (1.0)
FDUMMY2	.141 (4.0)	.127 (4.8)	.047 (1.9)	.063 (2.3)
FDUMMY3	.083 (3.0)	.029 (1.0)	.022 (1.0)	-.037 (-1.7)
R-squared	**.780**	**.858**	**.926**	**.921**

Table D.1. Price equation (MODEL 3):

Variable	**Round 2**	**Round 3**	**Round 4**	**Round 5**
$CONST_t$	4.191 (40.3)	1.364 (6.9)	1.276 (15.1)	1.202 (16.0)
DS_{t-1}	.836 (33.8)	.627 (18.7)	.283 (9.0)	.404 (17.0)
DS_{t-1}	.136 (15.1)	.016 (.8)	.079 (5.6)	.045 (4.0)
DS_{t-1}	-.023 (-14.4)	-.036 (-5.2)	.009 (1.0)	.003 (1.3)
DS_{t-1}	-.005 (-11.9)	-.004 (-1.8)	-.002 (-.7)	.61E-3 (.4)
S_t	-.070 (-6.8)	-.214 (-15.1)	-.094 (-14.7)	-.063 (-11.5)
P_{t-1}		1.273 (36.1)	.947 (59.1)	.918 (63.2)
R-squared	**.906**	**.905**	**.971**	**.968**

Table D.2. Demand equations(MODEL 3):

Variable	**Round 2**	**Round 3**	**Round 4**	**Round 5**
$CONST_t$	3.806 (8.2)	.280 (.4)	.315 (.5)	-.535 (-.8)
PS_t	-.018 (-.5)	-.213 (-3.0)	-.129 (-1.4)	-.257 (-2.4)
S_t	.324 (8.3)	.709 (14.9)	.564 (17.0)	.617 (25.5)
P_t	-.818 (-11.5)	-.574 (-11.9)	-.210 (-4.2)	.045 (.7)
D_{t-1}	.548 (16.9)	.462 (10.1)	.475 (12.8)	.457 (14.0)
D91 + LIAB	-.007 (-.8)	.003 (.4)	.0009 (.1)	-.001 (-.1)
ΔD91	.135 (2.1)	-.007 (-.1)	.010 (.1)	.007 (.1)
IND6	.213 (1.9)	.012 (.1)	-.035 (-.3)	.080 (.8)
REGP	.202 (2.9)	-.046 (-.6)	-.107 (-2.0)	-.034 (-.5)
DOMESTIC	.0002 (.0)	.007 (3.2)	.33E-3 (.1)	.002 (1.1)
IND3	.348 (3.6)	.004 (.0)	-.076 (-.9)	.021 (.2)
EMPL	.026 (1.7)	.044 (3.0)	.58E-3 (.0)	.010 (.7)
IND2	.101 (1.3)	-.109 (-1.3)	-.120 (-1.9)	-.075 (-1.2)
IND4	-.118 (-1.4)	.020 (.2)	-.265 (-4.1)	.045 (.6)
R-squared	**.777**	**.854**	**.927**	**.920**

E.1. Correlation matrix of error terms related to OLS

	EP2	EP3	EP4	EP5	ED2	ED3	ED4	ED5
EP2	1.0000							
EP3	-0.282**	1.0000						
EP4	-0.0395	0.149**	1.0000					
EP5	-0.0119	0.0444	0.034	1.0000				
ED2	-0.177**	0.091**	-0.0362	0.097**	1.0000			
ED3	0.254**	-0.418**	-0.0195	0.0010	0.0246	1.0000		
ED4	0.154**	-0.092**	-0.430**	0.0001	0.0524	0.140**	1.0000	
ED5	0.065*	-0.0310	0.147**	-0.297**	0.0074	0.233**	-0.077*	1.0000

NUMBER OF OBSERVATIONS: 791
1-tailed Signif: * - .05 ** - .01 (via Fisher z-transformation)

where

EP2,..,EP5	OLS residuals of price equation for rounds 2,..., 5
ED2,..,ED5	OLS residuals of demand equation for rounds 2,..., 5.