Pre-election Polls, Election Results, and Validity of Measurement before the 2002 Elections

Martin Kreidl & Tomáš Lebeda

SP 03:8 **2003**

This work was reviewed by: Doc. PhDr. Hynek Jeřábek, CSc.
PhDr. Ladislav Cabada, Ph. D.

Work on this project was kindly supported by the Grant Agency of the Academy of Sciences of the Czech Republic (grant no. S7028204).

ISBN 80–7330–043-5

Acknowledgements

The authors would also like to thank Jindřich Krejčí and both reviewers for their helpful comments and critiques of earlier drafts. An earlier version of the first chapter was presented at a seminar at the Institute of Sociology, Czech Academy of Sciences in Prague on 6. March 2003.

Obsah

Předvolební výzkumy, volební výsledky a validita měření před parlamentními volbami v roce 2002

Martin Kreidl & Tomáš Lebeda

Abstrakt

V tomto textu se pokoušíme empiricky zhodnotit kvalitu měření volebních preferencí před českými parlamentními volbami v roce 2002. Popisujeme jak zveřejněné výsledky, tak měřící techniky používané různými výzkumnými agenturami a z různých hledisek hodnotíme jejich validitu. Používáme dva základní, i když zásadně odlišné, přístupy: hodnocení konstruktové validity a hodnocení kriteriální validity. V první části textu používáme pokročilé statistické techniky původně určené k měření konstruktové validity škál k identifikaci rozdílů v dosažené validitě mezi jednotlivými agenturami. V této analýze používáme sadu standardizovaných měřících nástrojů a argumentujeme, že veškeré zjištěné rozdíly ve validitě musejí být připsány terénní práci a dalším faktorům, které jsou specifické pro každou agenturu. Tento přístup identifikuje rozdíly v kvalitě měření a terénní práci mezi agenturami. Druhá sekce textu popisuje a zhodnocuje kriteriální validitu instrumentů používaných jednotlivými agenturami a výsledků, kterých je těmito instrumenty dosaženo. Tato část textu se zaměřuje na volební předpovědi a srovnává je se skutečnými volebními výsledky.

Klíčová slova

Kvalita měření, validita, konstruktová validita, kriteriální validita, výzkumy volebních preferencí, volební prognózy, parlamentní volby

Pre-election Polls, Election Results, and Validity of Measurement before the 2002 Elections

Martin Kreidl & Tomáš Lebeda

Abstract
This text seeks to empirically assess the quality of measurement of voting preferences before the parliamentary elections in the Czech Republic in the year 2002. It looks at the results and the measurement techniques employed by various public opinion poll agencies and evaluates their validity. It uses two rudimentary and yet rather different approaches, one focusing on the construct validity of measurement, and the other focusing on its criterion validity. First, the authors use advanced statistical instruments originally intended to measure the construct validity of scales, to look at differences in the achieved validity between agencies. In this exercise the authors use a standardised set of measurements and argue that all other sources of measurement errors and diferences in validity must be attributed to fieldwork and other procedures that are agency-specific. This approach points out the existing differences in the quality of work that individual agencies supply. The second section of the text describes and assesses the criterion validity of the instruments employed and the results produced by respective research agencies. It focuses on election predictions and compares them with actual election results.

Key words
Quality of measurement, validity, construct validity, criterion validity, party preference polls, election forecasts, parliamentary elections

Erhebungen vor Wahlen, Wahlergebnisse und die Validität der Messungen vor den Parlamentnswahlen im Jahr 2002

Martin Kreidl & Tomáš Lebeda

Abstraktum
In dem Text versuchen wir, die Qualität der Wahlvorhersagen vor den tschechischen Parlaments Wahlen im Jahre 2002 empirisch zu bewerten. Wir beschreiben sowohl die veröffentlichten Ergebnisse als auch die verwendeten von den verschiedenen Meinungsforschungsinstituten verwendeten Messtechniken, deren Richtigkeit wir aus verschiedenen Blickwinkeln bewerten. Wir verwenden zwei grundlegend verschiedene Herangehensweisen: die Bewertung der konstruktiven Validität und die Bewertung der Kriterien-Validität . Im ersten Teil verwenden wir fortgeschrittene statistische Techniken, die ursprünglich zur Messung der konstruktiven Validität von Skalen gedacht war, für die Identifizierung der Unterschiede in der erreichten Validität zwischen den Instituten. In dieser Analyse verwenden wir eine Reihe standardisierter Messwerkzeuge und argumentieren damit, dass alle gefundenen Unterschiedein der Validität, erstens der Arbeit im Terrain zugeschrieben werden müssen und zweitens Faktoren, die für die verschiedenen Institute spezifisch sind. Diese Vorgehensweise zeigt Unterschiede in der Messqualität und der Arbeit im Terrain zwischen den Instituten auf. Der zweite Teil des Textes bewertet die Kriterien-Validität der verwendeten Instrumente der jeweiligen Institute und ihrer damit erzielten Ergebnisse. Dieser Abschnitt konzentriert sich auf die Wahlvorhersagen und vergleicht sie mit den Wahlergebnissen.

Schlüsselbegriffe
Qualität der Messung, Validität, konstruktive Validität, Kriterien-Validität, Erfassung von Trends für Wahlen, Wahlprognosen, Parlamentswahlen

Introduction

The strength of any empirical research, and of social science research in particular, is entirely dependent on the quality and precision with which it is able to measure its concepts. Without a precise measurement there is no way to be able to accurately test theories and come to any relevant and unbiased conclusions. There are no sophisticated statistical procedures that can salvage bad data. It is for this reason that social science students are usually exposed to the key terms of measurement theory in their very first introductory course of survey research methods and have to master not only its terminology, but also its practical application. Furthermore, both applied and academic researchers must deal with issues related to the quality of measurement on an almost daily basis. They design measurement instruments, test them and redesign them until they achieve a reasonable degree of reliability and validity. Moreover, they spend hours designing sampling and other survey procedures to acquire an as accurate as possible picture of social phenomena in the population they wish to study. Mastery of both measurement and sampling procedures is a highly praised skill in the scientific community.

However, the issue of the quality of measurement seldom receives any deserved attention from the wider public. Interested citizens, community representatives, local and national politicians, administrators, business representatives, marketing specialists as well as journalists, public intellectuals and commentators usually consume polls and their results without actually inquiring about the details of the actual sampling and measurement techniques. The one occasion when even non-professionals and semi-professionals pay considerable attention to the technical aspect of polling occurs in connection with the measuring of voting preferences. A wave of interest relating to measurement quality surfaces particularly in connection with the approach of any elections. The Czech public were last witness to this phenomenon in the first half of 2002. At the time, some journalists attempted to systematically explain the variation appearing in the results of the public opinion research produced by different Czech agencies, especially the most frequently cited CVVM, STEM and TNS Factum. Consequently, the different methods used for obtaining measurements of voting preferences also came to be the focus of interest. For example, on April 19, 2002, *Hospodářské noviny* went over a report issued by ČTK, which confirmed the differences among the agencies in their measurement of voting preferences (čtk 2002). Lipold (2002) pointed to the inaccuracy that is inherent to election estimates, while Šídlo (2002), in a similarly inclined article, clarified, in relative detail, how estimates of voting preferences and election prognoses are produced and the methodology that is used to achieve this. One author, who concealed his or her identity behind the editorial sign luk (2002), later approached the subject of this paper most closely in an article published in *Právo*. The author speculated that the variation in voting preferences was the result of the different methods of questioning that were used by the individual agencies. Finally, Černý (2002) later opined that the outcome of any election prognosis is even based on the context of the specific questionnaire and the subjects of other questions that the interviewer asks a respondent in the survey. The debate was then joined by Seidlová (2001, 2002), who is also the director of the CVVM.

However, none of the authors went so far as to question or, conversely, to highlight the status of the fieldwork of any one agency. Nonetheless, the discussion surrounded a legitimate methodological question. Therefore, the question we pose in this publication is whether it is possible to *empirically compare the quality of the measurements of individual agencies and to confirm that one agency conducts its fieldwork better than another or obtains more valid results than its competition?*

In order to assess the quality of measurement of the main research agencies we combine two rather different and yet, we believe, complementary approaches. In the first section of the text, we use advanced statistical instruments originally intended to measure construct validity of scales to look at differences in achieved validity between agencies. In this exercise we use a standardised set of measurements and argue that all other sources of measurement errors and differences in validity must be attributed to the fieldwork and other procedures that are agency-specific. While this approach enables us to identify differences in the quality of work that individual agencies supply, it is impossible to use it to rank their performance, because we have no external criteria showing the actual validity of the scale we use. But the simple fact that there exist substantial differences in the validity of the scale and therefore in the quality of work of respective agencies is remarkable.

The second section of the text takes a different approach. It describes and measures the criterion validity[1] of instruments employed and results produced by research agencies. More specifically, it looks at election predictions publicised by agencies and compares them with actual election results. It also speculates about the sources of deviations, compares their sizes and directions to see if they appear to have a systematic component. Because there exists a clear criterion against which election predictions can be contrasted, we can gain rather strong evidence, that indeed not all results that were circulated in the public sphere before the 2002 election were of the same quality.

The 2002 elections were evidently the most abundant in Czech history in terms of research on voting preferences. In addition to the regular surveys that some agencies publicise on roughly a monthly basis, there was literally a surge in the number of commercial surveys sponsored by the media. These "exclusive" polls were commissioned not only by the largest television and radio stations but also by some daily newspapers and internet servers. It appears, however, that the quality of surveys varied considerably. The second section analyses the work of three of the most frequently cited agencies from a "user's" perspective and is devoted to the actual surveying of pre-election preferences. It presents the pre-election surveys these agencies carried out, compares their results, and assesses them in relation to the actual election outcome.

[1]See e. g. Schutt (2004) for the exact definition of ''construct validity'' and ''criterion validity''.

Measuring Construct Validity and its Application to Evaluate the Quality of Work done by Research Agencies

The Quality of the Measurements and the Fieldwork Procedures

The quality of the statements about society that are based on empirical evidence depends in a critical way on the quality of the data that stand behind them. Data quality is defined by two key components: representativeness and the quality of the measurement. Representativeness indicates how well the sample relates to the basic population referred to by the research, while the quality of the measurements is indicated by its reliability and validity. By reliability we mean how accurate the measurements of the variable that we are measuring are, while validity refers to the ability to actually measure the concept that we are trying to measure (see e.g. Řehák 1998a; Schutt 2004).

The level of importance that quality measurements have come to signify has in recent years evoked a wave of interest among social scientists. Although the classic model for obtaining measurements was laid out in the monograph by Lord and Novick (1968), the subject of measurement quality only came to be more broadly applied methodologically and popularised in the 1980s and especially in the 1990s (e.g. Alwin 1989; Saris, Meurs 1990; Scherpenzeel 1995). The theory of measurements was introduced into the Czech academic sphere in a paper by Jeřábek (Jeřábek 1992), and later, the empirical tools for determining the quality of measurements was presented in a series of articles by Řehák and his colleagues (Řehák 1998a, 1998b, Řehák, Bártová, Hamanová 1998).

Interest in the subject of measurement quality is not restricted to the field of sociology. Recent years have revealed an interest in measuring the reliability and verifying the validity of measuring instruments also among other fields of the social sciences; for example, in social demography (e.g. Wu, Martin, Long 1999), social psychology (e.g. Lynch 2000; Rees, Hardy, Ingledew, Evans 2000; Russell 1996), criminology (Smith, Patterson 1984), research on education (e.g. Huang, Michael 2000), psychology (e.g. Ward 1994) and social work (Reed-Ashcraft, Kirk, Fraser. 2001), and also in medicine and research on public health (Bjorner, Kristensen 1999; Sukhwinder, Kuttalaliangam, Seneviratna, Orrell 1999).

The Sources of Measurement Error

The possible sources of poor measurement vary considerably and range from the features of the actual instruments of measurement to the conditions in which the measuring is conducted. In sociology, the most common instrument of measurement is the questionnaire. The reliability of the measurements obtained through a questionnaire is influenced by the very design and the length of the questionnaire, as well as by the wording of the questions it contains, the order the questions are presented in, the application of cards, etc. Contextual influences include the method whereby the questionnaires are administered (self-administered vs. face-to-face interviews), the interviewers' training, their performance and their neutrality when posing questions, and their credibility and reserve in reacting to the responses. The preconditions for obtaining quality measurements are also founded on mutual anonymity between the interviewer and the respondent, the rule of conducting only one interview per household, and ruling out the possibility of repeatedly questioning a single person (e.g. Alwin 1989; Řehák 1998a).

For the analysis in this paper I have made use of compatible measuring instruments in order to evaluate the quality of the work of the best known Czech agencies engaged in the collection of

sociological data and measuring voting preferences, specifically looking at the work of the private agencies STEM (Centre for Empirical Studies) and TNS Factum, and the academic Public Opinion Research Centre (CVVM). I conducted a comparison with the aid of the MIMIC model, which is intended for the verification of the construct validity of scales. In estimating a model on the basis of data from each agency and comparing the attained level of validity of the scale, using the results we can also compare the quality, or better put, the equivalence of the work of the agencies, as the only source of difference in the observed level of validity, owing to the given design, can be the work of the agencies themselves. Thus it is not a matter of conducting a complex test of the quality of the measurements in the agencies, but involves rather one of its aspects and disregards other, e.g. the varied sizes of the samples used, etc. Nor is the exercise capable of tracing any potential variability in the quality of measurements over time.

I used data drawn from a methodological survey conducted in the months prior to the elections in 2002. A set of standard questions was added to the omnibus surveys of all three agencies, which were also used for measuring voting preferences, and from this set of questions I used a five-point battery for measuring the position of the respondent on an "objective" left-right scale of political orientations (see below), a seven-point scale running from left (1) to right (7) to measure subjective political orientation, and a ten-point scale of subjective social status running from the highest status (1) to the lowest (10), and a four-point scale to measure education (elementary, vocational, complete secondary, post-secondary). These are questions that in the Czech and international contexts have been used on numerous occasions in the past (see e.g. Evans, Heath, Lalljee 1996; Kreidl 1998; Matějů, Vlachová 2000). The methodological survey indicated above was conducted as part of the project "The Quality of Studies of Voting Preferences", which is currently being conducted at the Institute of Sociology AS CR and is supported by the Grant Agency of the Academy of Sciences of the Czech Republic as part of the Programme for the Support of Applied Research and Development.

As the wording of the questions, the measuring scales that were used, and other tools and terms presented in the questions (May 2002) were in all contexts (i.e. agencies) compatible to a maximum degree, and as the agencies make use of very similar sampling procedures, it is possible to ascribe the only observed differences in the quality of the measurements to a difference in the quality of the work of the individual agencies. A source of the variation in quality could, for example, be the quality of the interviewer network, interviewer training, the social composition of the network, the neutrality and credibility of the interviewers, the strict observance of the instructions for carrying out the research, not repeating interviews in the same households repeatedly, conducting interviews only with respondents the interviewers do not know, etc.

Another source of poor quality measurements could also of course be the poor quality of data processing in the agency itself, the data coding and internal checks. The quality of measurements is also affected by the overall length of an interview (i.e. the number of additional questions in the omnibus survey), or by the context of the additional questions in the research, as these could lead to a short-term change in attitude, or could evoke certain responses, initiate a process of forming a new attitude, or even increase or decrease the willingness of the respondents to participate in the research.[2]

[2]The influence of these contextual effects can never be completely ruled out, or standardised, even though it would be useful, for example, in a methodological survey of this kind. As the contexts of the question on voting preferences change in the omnibus surveys from month to month, it is likely that the quality of the data will also change with it. Therefore, it would be a useful to repeat an experiment similar to ours more often and to thereby acquire more robust results.

At first glance, all three agencies offer quite similar conditions in their monthly omnibus surveys. The basic method applied is a standardised interview conducted in personal, face-to-face contact with the respondent, sometimes with the aid of cards. Cards were also used in the questions in our methodological experiment. The respondent's answers are recorded by the interviewer in pre-printed blanks on the questionnaire. All three firms use the method of quota sampling with almost identical quotas. TNS Factum uses 5 quotas: regions- NUTS3, size of place of residence, age (5 categories), sex and education (4 categories). TNS Factum did not provide us with a technical report to accompany the data sample; therefore, we do not know the details of the breakdown of the age groups and residential-size categories in the quota grid. STEM also uses quota sampling with five quotas: regions (defined as 8 regions at the 'NUTS2' level), size of place of residence (6 categories), sex, age (4 categories) and education (4 categories). CVVM uses 5 quotas: region (8 regions according the old administrative specification), size of place of residence (6 categories), sex, age (4 categories) and education (also 4 categories).

Neither STEM nor CVVM offer clients weights in their studies, while TNS Factum offers two alternative ways of weighting. The first is the socio-demographic weight, which weighs the results according to the basic sampling quotas. The second weight modifies the results according to region, size of place of residence, and voting behaviour in 1998. Details on how the weights were formed were not provided along with the data sample, which is a somewhat surprising approach. The weights are not used in the analysis of the data from TNS Factum.

The research by STEM was conducted in the first week of May 2002 on a total of 2080 respondents. The research by CVVM was carried out on the same dates using a sample of 1083 respondents in total. In the week beginning May 13, 2002, TNS Factum conducted its research on 1014 respondents. Given that the individual agencies delimit the age of the basic population differently (15 years and over vs. 18 years and over[3]), we decided to unify the age distribution and to include only respondents aged 18 and over in the analysis (972 respondents at TNS, 2077 respondents at STEM, and 991 respondents at CVVM). Once respondents with missing data had been excluded, there remained for the analysis 754 respondents in TNS Factum, 1817 respondents in STEM and 709 respondents in CVVM. Given that in the comparison of models (see below) we are using test statistics on the basis of chi-square, the size of which depends on the size of the sample, we standardised the sample at 709 respondents from each agency. These 709 respondents were selected from the numbers of respondents at TNS Factum and STEM using a probability method. For a comparison, of course, we also present the results of the test without this delimitation of the sample (see below).

[3]Though STEM indicates the basic population in the research to be aged 18 and over, we found 3 respondents in the sample who indicated a younger age. Even this is an indication of the attention STEM pays to cleaning data before providing them to the client.

Methods of Measuring Reliability and Validity

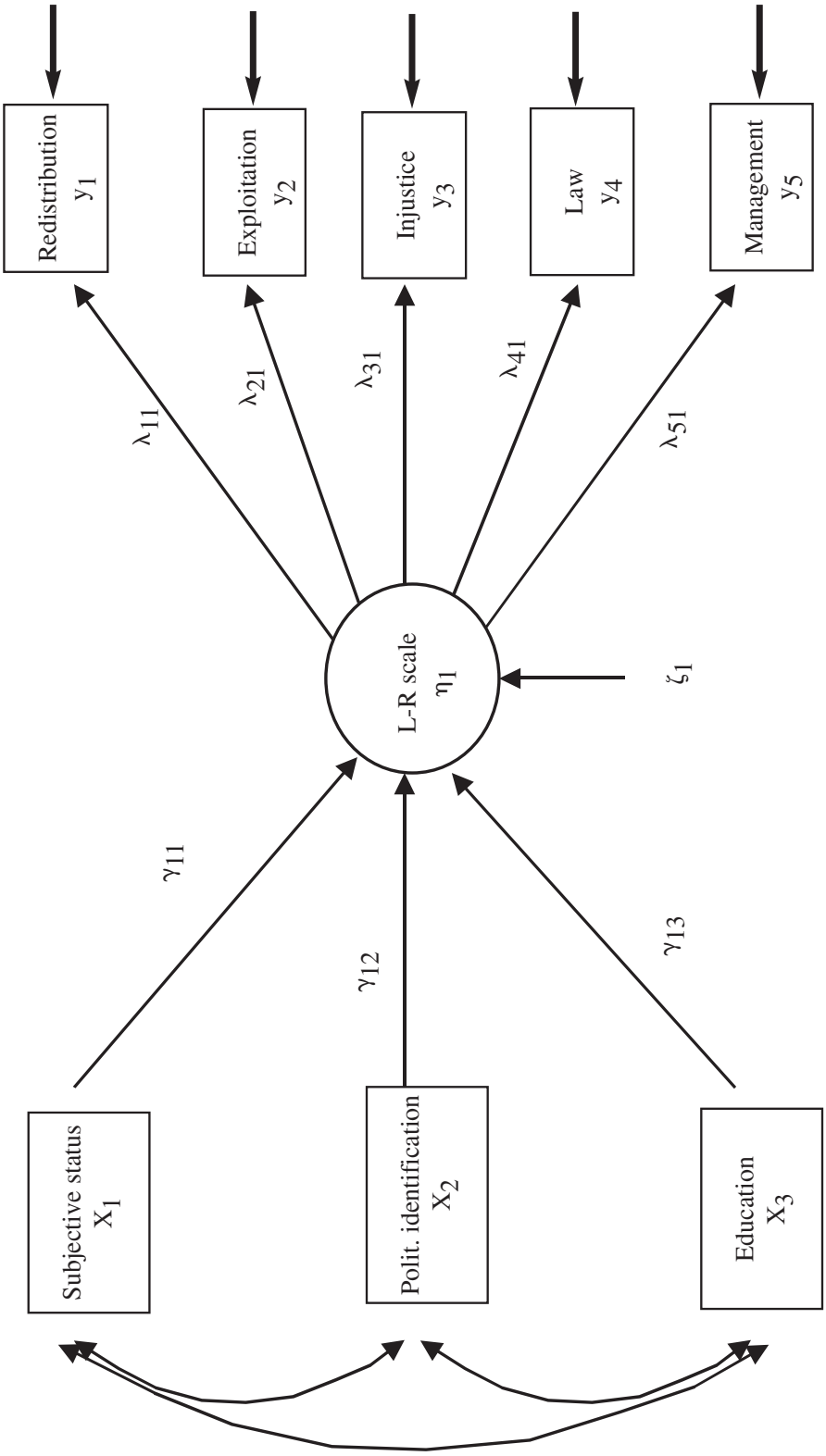
There are numerous methods that can be used to measure the reliability and validity of measurements. Among the classic tools used for evaluating the quality of individual measuring items, there is the method of the Quasi simplex model, the RMM method (repeated multimethod model) and, clearly the most advanced, the MTMM design (multitrait- multimethod model), which alone also enables – in addition to measuring reliability – the measurement of validity, albeit conceived differently than it is in this paper (see e.g. Alwin 1989; Scherpenzeel 1995). If we intend to evaluate the quality of a multi-item scale, the spectrum of instruments that are applicable expands to include a tool for measuring the 'correlatedness' of individual items in the scale (e.g. the Cronbach's alfa; see e.g. Matějů, Vlachová 2000), and a tool for measuring the convergent and discriminatory validity of scales, such as exploratory and confirmatory factor analysis and their various versions, modifications and extensions (e.g. Bell, Lee 2002; Campbell, Arthur 1997; Jones et al. 2002; Lynch 2000; Russell 1996; Zhu 2000). The final tool is then the MIMIC model, used for determining the construct validity of scales (e.g. Bagozzi 1978; Edmundson, Koch 1993; Jöreskog, Sörbom 1975; Ruble, Stout 1990; Williams 1994). It is this method that I have chosen for the analysis.

The MIMIC model (MIMIC is an abbreviation for Multiple Indicators - MultIple Causes model) is probably the most common current tool for measuring the construct validity of measuring instruments. What the MIMIC model does is to test the validity of a scale in a specific theoretically relevant context, which, in the case of the most simple version of the MIMIC model, can be a situation where there is one latent scale. This kind of simple model is used, for example, by Hodge and Treiman (1968), who have tested the dependency of social participation (measured with the aid of three indicators) on various ladders of social status (income, employment, education). The principle behind the MIMIC model is a statistical test of the assumption that the influence of measured causes on the measured indicators is not direct but rather mediated through one latent variable. The test enables us to entirely reject a scale as being invalid in the given context. Here we are primarily interested in a comparison of the results from the individual agencies. Although we could entirely reject, or accept, the scale in all three data samples, we can still then observe whether the results from the agencies differ. A difference in the level of validity indicates a difference in the quality of the agency's work. However, as we do not know the genuine validity of the scale in Czech society, we cannot confirm which agency attained the best outcome and which conversely did worst. We could only reach this kind of conclusion on the basis of different analyses, such as those for example offered in the second substantive part of this publication.

Summarising the key features of the MIMIC model:

It was probably Hauser and Goldberger (1971) who first presented a methodological discussion of the MIMIC type of models. Jöreskog and Goldberger (1975) further showed that it is possible to view the MIMIC model like a multi-variate regression model, with certain systemic restrictions (the specific form of which is described below).

Figure 1. MIMIC model for testing the construct validity of the left-right scale of political attitudes



In the specific case here the MIMIC model at the outset of the analysis is presented in figure 1. The model contains 8 directly measured variables and one latent variable η. Of the 8 directly measured variables 3 are exogenous (x) and 5 are endogenous (y). Endogenous variables are items that are standardly used for measuring the so-called objective left-right political orientation (see e.g. Evans, Heath, Lalljee 1996; Matějů, Vlachová 1998, 2000). It refers to the amount of agreement (expressed in a four-point scale, where 1 is strongly agree, 2 means somewhat agree, 3 somewhat disagree and 4 strongly disagree) indicated with regard to the following five statements (the abbreviated expressions used in the figures are indicated in the brackets):

1. The government should redistribute the income of those who are better off to those who are worse off. (redistribution)

2. Owners of large companies get rich at the expense of labourers. (exploitation)

3. Normal working people do not get their fair share of the national wealth. (injustice)

4. There exists one set of laws for the rich and another for the poor. (law)

5. The management of a company will always try to get the better of an employee whenever they get the chance. (management)

These five attitudes are, in compliance with the theory, interpreted as the five indicators of the latent construct (η_1). Finally, the model contains two types of errors – errors in the equations (ζ_i) and errors of measurement (ϵ_i). The entire proposed model agrees with previous analyses of determinants of political attitudes in Czech society (e. g. Kreidl 2000; Matějů, Vlachová 2000), and therefore it is a suitable tool for our analysis.

The relations between the variables in the model can be expressed in two equations - the structural equation and the measurement equation.

The structural equation can be written as:[4]

$$\eta = \Gamma X + \zeta, \quad (1)$$

where η is the column vector of (generally) m latent variables, X is the column vector of n observed exogenous variables, Γ is the matrix (m x n) of the parameters, which expresses the effects of the exogenous variables (x_i) on the vector of m latent endogenous variables (η) and ζ is the column vector of (generally) m disturbances in the measurement of m latent constructs.
The model of measurement is then:

$$Y = \Lambda \eta + \epsilon, \quad (2)$$

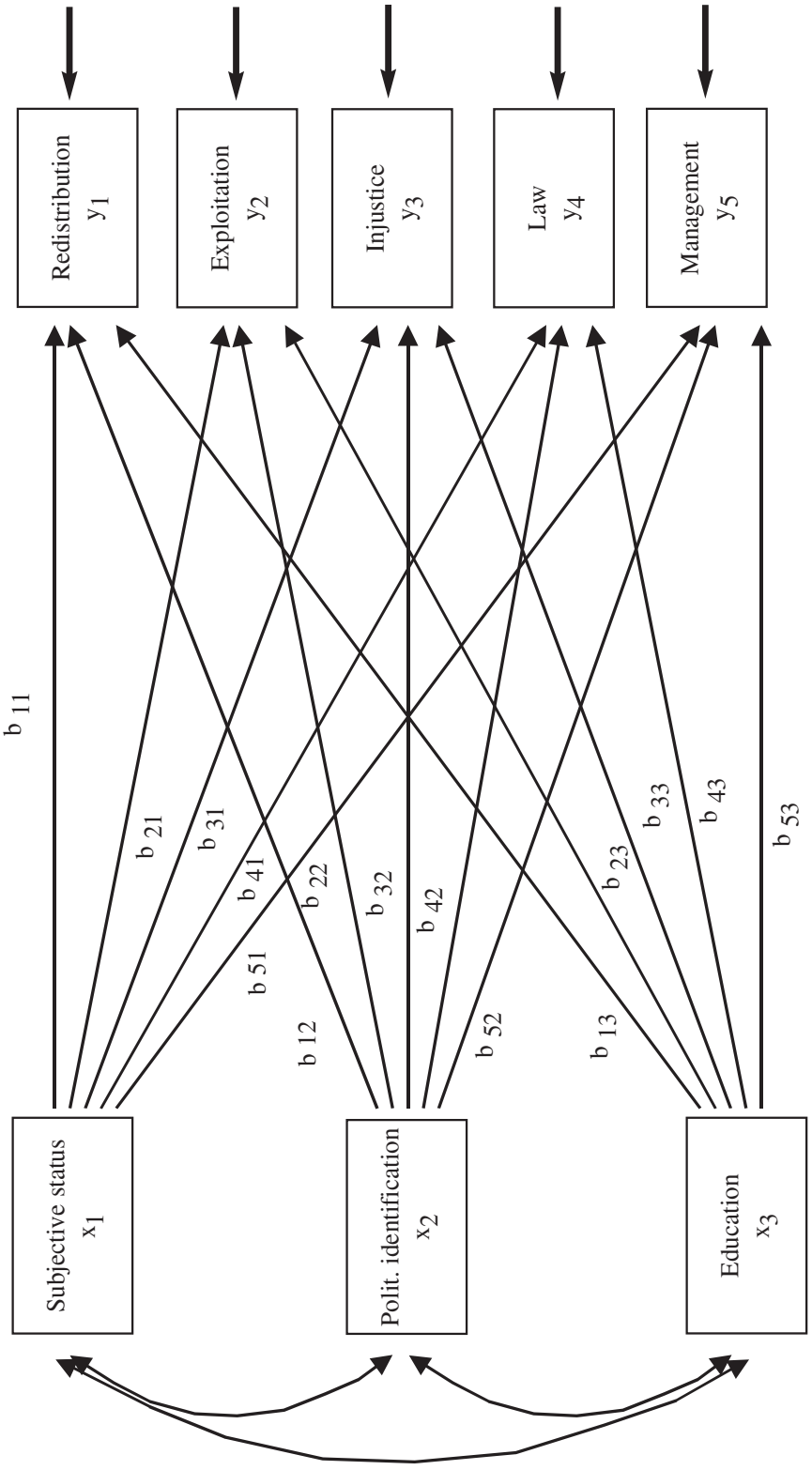
where Y is the column vector of p observed indicators, η is the column vector of (generally) m latent variables, Λ is (p x m) the parameter matrix containing the factor loadings and ϵ is the column vector of p errors of measurement among the indicator variables.

[4]I use notation common in structural models, e.g. in the LISREL manual whenever I describe a structural model (see Jöreskog, Sorbom 1989).

Measurement of the Validity of Scales with the Aid of the MIMIC Model

An evaluation of the construct validity of the scale can be made through a statistical comparison between the MIMIC model, which assumes the existence of a latent scale that mediates the relationship between the exogenous variables and the indicator variables, and the model which does not assume the existence of the scale and specifies only direct causal ties between the exogenous and endogenous variables measured. Based on the results of this statistical test, we can either accept the MIMIC model as an accurate and succinct summary of structural relationships between variables, or we reject it and therefore consider the presumed scale invalid.

Figure 2. Unrestricted regression model assuming direct effect of exogenous variables on the indicators of the left-right scale



In the case we are dealing with in this paper, the model devoid of the assumed existence of the scale would appear as illustrated in figure 2. This model works with each indicator item independently and does not anticipate any scaling restrictions, i.e. it leaves each endogenous variable to react individually to the exogenous variable, while the relative strength of the effects of individual exogenous variables can in each of the equations change randomly. Each of the direct arrows in figure 2 corresponds to one of the total of 15 regression coefficients in the model. In subtracting the average from all explanatory variables it applies that:

$$y_1 = b_{11} x_1 + b_{12} x_2 + b_{13} x_3 + \epsilon_1$$
$$y_2 = b_{21} x_1 + b_{22} x_2 + b_{23} x_3 + \epsilon_2$$
$$y_3 = b_{31} x_1 + b_{32} x_2 + b_{33} x_3 + \epsilon_3$$
$$y_4 = b_{41} x_1 + b_{42} x_2 + b_{43} x_3 + \epsilon_4$$
$$y_5 = b_{51} x_1 + b_{52} x_2 + b_{53} x_3 + \epsilon_5 \quad (3)$$

In the MIMIC model, where y_1 is the referential indicator for η and therefore $\lambda_{11} = 1$, it applies that:

$$y_1 = \eta + \epsilon_1$$
$$y_2 = \lambda_{21} \eta + \epsilon_2$$
$$y_3 = \lambda_{31} \eta + \epsilon_3$$
$$y_4 = \lambda_{41} \eta + \epsilon_4$$
$$y_5 = \lambda_{51} \eta + \epsilon_5 \quad (4)$$

and at the same time that:

$$\eta = \gamma_{11} x_1 + \gamma_{12} x_2 + \gamma_{13} x_3 + \zeta \quad (5)$$

The MIMIC model is nested in the non-restricted regression model. MIMIC (fig. 1) assumes the following restrictions within the regression model (fig. 2):

$b_{11} = \gamma_{11}$	$b_{12} = \gamma_{12}$	$b_{13} = \gamma_{13}$
$b_{21} = \gamma_{11} \lambda_{21}$	$b_{22} = \gamma_{12} \lambda_{21}$	$b_{23} = \gamma_{13} \lambda_{21}$
$b_{31} = \gamma_{11} \lambda_{31}$	$b_{31} = \gamma_{12} \lambda_{31}$	$b_{33} = \gamma_{13} \lambda_{31}$
$b_{41} = \gamma_{11} \lambda_{41}$	$b_{42} = \gamma_{12} \lambda_{41}$	$b_{43} = \gamma_{13} \lambda_{41}$
$b_{51} = \gamma_{11} \lambda_{51}$	$b_{52} = \gamma_{12} \lambda_{51}$	$b_{53} = \gamma_{13} \lambda_{51}$ (6)

From which the following restrictions emerge:

$$b_{21}/b_{11}=b_{22}/b_{12}=b_{23}/b_{13} \quad (7)$$

$$b_{31}/b_{11}=b_{32}/b_{12}=b_{33}/b_{13} \quad (8)$$

$$b_{41}/b_{11}=b_{42}/b_{12}=b_{43}/b_{13} \quad (9)$$

$$b_{51}/b_{11}=b_{52}/b_{12}=b_{53}/b_{13} \quad (10)$$

Instead of the 15 regression coefficients (see equation (3)) the MIMIC model has only 7 coefficients (γ and λ in equations (4) and (5)), which stem from the indicated 8 additional restrictions (see (7), (8), (9) and (10). These restrictions are known as "proportionality restrictions". *These show that the relative effects of the independent variables are the same for all items in the battery.* For example, x_1 can have a greater effect than x_2, but it nonetheless applies that the relative size of the effects x_1 and x_2 is the same for y_1, y_2, y_3, y_4 and y_5. The validity of the proportionality restrictions can be tested through a comparison of both models using a chi-square with 8 degrees of freedom. A higher chi-square value, while the size of the sample is standardised, signifies a greater departure from the proportionality restrictions and therefore a lower degree of validity of the measurements.

Discussion of the Results and Conclusions

The specific results of the tests for the validity of the proportionality restrictions in the MIMIC model are presented in table 1. For the purpose of illustration we present here not only the results of the test for each agency using the standardised sample, but also summary results for all the agencies, and the results again for each agency but using a non-standardised sample. However, as indicated above, in order to compare the quality of the agencies' work, the key data are the results presented in panel A of table 1, which presents the results of the test conducted on the standardised samples.

Table 1. Test of the validity of proportionality restrictions in the MIMIC model by research agency

	L^2	df	N	P
A. standardized sample sizes				
CVVM	4.60	8	709	0.7991
STEM	7.81	8	709	0.4519
TNS	8.41	8	709	0.3942
All	5.38	8	2127	0.716
B. complete samples				
CVVM	4.60	8	709	0.7991
STEM	12.52	8	1817	0.1293
TNS	10.02	8	754	0.2637
All	9.04	8	3280	0.3387

Note: Results produced by CVVM are the same in both panels of table 1, because CVVM had the smallest sample and therefore its sample was not reduced and served as the baseline for other agencies to achive the same size of 709 respondents.

Panel A in table 1 clearly reveals the substantial differences among the individual agencies in terms of the degree of validity the scale has attained. In the test for the validity of the proportionality restrictions of the MIMIC model, for the entire sample of respondents L^2 is 5.38, with 8 degrees of latitude (p value 0.716). By comparison, for the sub-sample of respondents used by CVVM, L^2 is 4.60 (p= 0.799), for the sub-sample of respondents used by STEM, L2 is 7.81 (p=0.452), and for the sub-sample used by TNS Factum, L^2 is 8.41 (p=0.394). Clearly, the proportionality restrictions are not falsified by the test, and therefore, we have to conclude that the scale is construct valid. This conclusion holds not only for the merged data file, but for each survey agency.

However, we do register considerable differences in the quality of the MIMIC model in the data collected by the individual agencies. As we standardised all aspects of the models in the analysis, with the exception of the actual work of each agency, the varying quality of the models is evidence of the varying quality of their work. However, this test does not measure which agency is better or of higher quality; it only demonstrates that there exist demonstrable qualitative differences between the individual agencies.

The main problem with this type of evaluation is that we actually do not know from any external source what is the true validity of the scale we have used and therefore we can't assess which of the three agencies came the closest to the real values in their data. While we have established that there indeed exist demonstrable differences in data collected by individual agencies, we propose that other methods be used to determine the actual ranking of agencies in terms of their performance. An exam-

ple of such an exercise is offered in the subsequent section of this publication, which compares electoral predictions made by individual agencies with the true outcome of the elections. Because election results provide an independent source of information about voters' behaviour, they are an obvious criterion to assess the validity of measurement of individual fieldwork agencies.

Let us now make a preliminary statement about the quality of measurement by individual agencies based on what we have just presented and what an interested reader may read in full in the second part of this publication. The second section of this text documents that it was mostly surveys conducted by TNS Factum and STEM, and in particular regional predictions, that did not pass the test of criterion validity, while CVVM fared much better. The sources of the difference are multiple, but are clearly agency-specific. The most likely sources of mis-measurement rest in the fieldwork procedures. Therefore, we conclude that, overall, CVVM must have been superior in one or more of the following areas: interviewers' training and ability including their performance while interviewing, their neutrality when asking questions, and their credibility and reserve in reacting to the responses; strict observance of the principle of mutual anonymity between the interviewer and the respondent; absolute respect of the rule of conducting only one interview per household; avoidance of repeated questioning of a single person. While we have little information about how interviewers are trained, controlled and supervised in individual agencies, it is clear from the latter section of this text that interviewers working for each agency had very variable workloads, ranging from the necessity to conduct a few interviews per month (CVVM) to a quest to complete dozens of interviews every week (TNS Factum, STEM). Under such severe working requirements the incentives to "make one's life easier" increase dramatically, not to mention the real, or only perceived, lower likelihood of efficient control of the fieldwork. Moreover, inevitable tiredness results in a more error-prone behaviour and may thus contribute to lower performance of interviewers while in the field.[5] Consequently, we would like to advise pre-election polls-users to bear in mind that less (surveys) is actually sometimes more (validity). An increased number of electoral predictions and conducted surveys is produced at the cost of their reduced quality. Polling managers, on the other hand, should reconsider the value of positive and negative PR, which is produced by increased media coverage at the expense of making invalid predictions.

Also, it is necessary to bear in mind that the test presented here does not represent a complex evaluation of the quality of the data that the indicated agencies routinely produce. As stated in the introduction, the quality of data is dependent on numerous circumstances, including the quality of the sample and the questions that are employed. The quality of the inference of the sample for the basic population also depends on the size of the sample and the ability to produce, when necessary, strictly a probability sample. The last variable in the comparison of voting preferences is then the instrument that was used itself, i.e. the question that the agency uses to ascertain party preferences. Of course, in this regard, there is no clear consensus in the academic community concerning which method of questioning is better and more accurate in ultimately predicting the election results.

[5]Unfortunately, we have no information about what percentage of interviewers work for more than one fieldwork agency at the same time. While some overlap is likely and is monitored by the fieldwork managers there is usually still a substantial proportion of the interviewers who work for only one agency. Therefore, we believe that our "extreme workload" argument is still a valid one.

References

„Agentury se rozcházejí v odhadech volebních výsledků.“ 2002. *Hospodářské noviny,* 19. 4. 2002.

Bagozzi, Richard P. 1978. „The Construct Validity of the Affective, Behavioral, and Cognitive Components of Attitude by Analysis of Covariance Structure." *Multivariate Behavioral Research* 13: 9- 31.

Bell, S. and C. Lee 2002. „Development of the Perceived Stress Questionnaire for Young Women." *Psychology, Health & Medicine* 7: 189- 201.

Bjorner, Jakob Bue and Tage Sondergaard Kristensen. 1999. „Multi-Item Scales for Measuring Global Self-Rated Health. Investigation of Construct Validity Using Structural Equations Models." *Research on Aging* 21: 417- 439.

Bohrnstedt, Georg W. 1977. „Use of Multiple-Indicators Multiple-Causes (MIMIC) Model." *American Sociological Review* 42: 656- 663.

Černý, Jiří 2002. „Proč průzkumy selhaly.“ *Lidové noviny, příloha Orientace,* 20. 6. 2002.

Campbell, Todd and James Arthur 1997. „The Factor Structure of the Bem Sex-Role Inventory (BSRI): Confirmatory Factor Analysis of Long and Short Forms." *Educational & Psychological Measurement* 57: 118- 124.

Edmundson, Elizabeth and William W. Koch 1993. „A Facet Analysis Approach to Content and Construct Validity." *Educational & Psychological Measurement* 53: 351- 368.

Evans, G., A. Heath and M. Lalljee 1996. „Measuring left-right and libertarian-authoritarian values in the British electorate". *British Journal of Sociology* 47: 93- 112.

Gallo, Joseph J. and James C. Anthony 1994. „Age Differences in the symptoms of Depression: a Latent Trait Analysis." *Journal of Gerontology* 49: 251- 234.

Gilles, David E. A. 1999. „Modeling the Hidden Economy and the Tax-Gap in New Zealand." *Empirical Economics* 24: 621- 640.

Hauser Robert M. and Arthur S. Goldberger 1971. „The Treatment of Unobservable Variable in Path Analysis." Pp. 81- 177 in: *Sociological Methodology.* Edited by H. L. Costner. San Francisco: Jossey- Bass.

Hodge, Robert W. and Donald J. Treiman 1968. „Social Participation and Social Status." *American Sociological Review* 33: 723- 740.

Huang, Chiungjung and William B. Michael 2000. „A Confirmatory Factor Analysis of Scores on a Chinese Version of an Academic Self-concept Scale and its Invariance Across Groups." *Educational and Psychological Measurement* 60: 722- 786.

Jeřábek, Hynek 1992. „Koncepce měření v sociálních vědách.“ *Sociologický časopis* 28: 103 - 118.

Inverarity, James M. 1976. „Populism and Lynching in Louisiana, 1889-1896: a Test of Erikson’s Theory of the Relationship Between Boundary Crises and Representative Justice.“ *American Sociological Review* 41: 262- 280.

Jones, Nicole Tuomi, Peter Ji, Mary Beck and Niels Beck 2002. „The Reliability and Validity of the Revised Conflict Tactics Scale (CTS2) in a Female Incarcerated Population.“ *Journal of Family Issues* 23: 441- 457.

Jöreskog, Karl G. and Arthur S. Goldberger 1975. „Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable." *Journal of American Statistical Association* 70: 631- 639.

Jöreskog, Karl G. and Dag Sorbom 1989. LISREL 7. *A Guide to the Program and Applications. 2nd Edition. Chicago: SPSS Inc.*

Kreidl, Martin 1998. „Changes in the Perception of Poverty and Wealth: Czech Republic Between 1991- 1995." *Czech Sociological Review* 6: 73- 97.

Kreidl, Martin. 2000. „Změny v percepci životního úspěchu, bohatství a chudoby." Pp. 191-224 in: *Nerovnosti, spravedlnost a politika. Česká republika 1991-1998.* Edited by Petr Matějů and Klára Vlachová. Praha: Sociologické nakladatelství.

Lipold, Jan 2002. „Proč se prognózy pokaždé aspoň trochu mýlí." *Hospodářské noviny* 10.6. 2002.

Lord, Frederic M. and Melvin R. Novick. 1968. *Statistical Theories of Mental Test Scores.* Reading: Addison-Wesley Publishing Company Inc.

„Volební průzkumy bývají divoké." 2002. *Právo*, 1. 6. 2002.

Lynch, Scott M. 2000. „Measurement and Prediction of Aging Anxiety." *Research on Aging* 22: 533- 558.

Mast Benjamin T., Susan E. MacNeill and Peter A. Lichtenberg. 2002. „A MIMIC Model Approach to Research in Geriatric Neuropsychology: The Case of Vascular Dementia." *Aging Neuropsychology and Cognition* 9: 21- 37.

Matějů, Petr and Klára Vlachová et al. 2000. *Nerovnost, spravedlnost a politika. Česká republika 1991-1997.* Praha: Sociologické nakladatelství.

Reed-Ashcraft, Kellie, Raymond S. Kirk and Mark W. Fraser 2001. „The Reliability and Validity of the North Carolina Family Assessment Scale." *Research on Social Work Practice* 11: 503- 520.

Rees, Tim, Lew Hardy, David K. Ingledew and Lynne Evans 2000. „Examination of the Validity of the Social Support Survey using Confirmatory Factor Analysis." *Research Quarterly for Exercise and Sport* 71: 322- 330.

Ruble, T. L. and Stout, D. E. 1990. „Reliability, Construct Validity, and Response- Set Bias of the Revised Learning- Style Inventory (LSI- 1985)." *Educational & Psychological Measurement* 50: 619- 629.

Russell, Daniel W. 1996. „UCLA Loneliness Scale (Version 3): Reliability, Validity, and Factor Structure." *Journal of Personality Assessment* 66: 20-40.

Řehák, Jan 1998a. „Kvalita Dat I. Klasický model měření reliability a jeho praktický aplikační výzkum." *Sociologický časopis* 34: 51- 60.

Řehák, Jan 1998b. „Kvalita dat II. Přístupy ohodnocování výzkumných instrumentů založené na modelování kovariančních struktur." *Sociologický časopis* 34: 195- 204.

Řehák, Jan, Irena Bártová and Jana Hamanová 1998. „Kvalita dat III. Empirické výsledky měření reliability pro vybrané míry a stupnice." *Sociologický časopis* 34: 363- 372.

Saris, W. E. and A. van Meurs (Eds.) 1990. *Evaluation of Measurement Instruments by Meta-Analysis of Multi-Trait Multi-Method Studies.* Amsterdam: North- Holland.

Scherpenzeel, Annette 1995. *A Question of Quality. Evaluating Survey Questions by Multitrait-Multimethod Studies.* Leidschendam: KPN Research

Seidlová, Adéla 2001. „Volební preference politických stran CVVM.“ *Neviditelný pes* 21. 11. 2001.

Seidlová, Adéla 2002. „Výzkumy stranických preferencí.“ *Neviditelný pes* 20. 3. 2002.

Shutt, Russel K. 2004. *Investigating the Social World. The processes and Practice of Research. 4th Edition.* Thousand Oaks: Pine Forge Press.

Smith, Douglas A. and E. Britt Patterson 1984. „Applications and a Generalization of MIMIC Models to Criminological Research.“ *Journal of Research in Crime and Delinquency* 21: 333-352.

Stapleton, David C. 1978. „Analyzing Political Participation Data with a MIMIC Model.“ *Sociological Methodology* 9: 52- 74.

Sukhwinder S. Shergill, Kuttalaliangam K. Shankar, Knightley Seneviratna and Martin W. Orrell. 1999. „The Validity and Reliability of the Health of the Nation Outcome Scales (HoNOS) in the Elderly.“ *Journal of Mental Health* 8: 511- 521.

Ward, Edward A. 1994. „Construct Validity of Need for Achievement and Locus of Control Scales.“ *Educational and Psychological Measurement* 54: 983- 992.

Warren, John Robert, and Robert M. Hauser 1997. „Social Stratification Across Three Generations: New Evidence from the Wisconsin Longitudinal Study.“ *American Sociological Review* 62: 561- 572.

Williams, Jannice E. 1994. „Anxiety Measurement: Construct Validity and Test Performance.“ *Measurement & Evaluation in Counseling & Development* 27: 302-307.

Wu, Lawrence L., Steven P. Martin, and Daniel A. Long. 1999. Comparing Data Quality of Fertility and First Sexual Intercourse Histories. Center for Demography and Ecology WP 99/08. Madison: University of Wisconsin.

Zhu, Weimo 2000. "Which Should It Be Called: Convergent Validity or Discriminant Validity?" *Research Quarterly for Exercise and Sport* 71: 190- 194.

Surveys of Voting Preferences prior to the 2002 Parliamentary Elections

A Comparison of the Research Conducted by Three Research Agencies – CVVM, STEM and TNSF[6]

The Surveys Conducted

During the period in the run-up to the 2002 election, there were three types of election surveys encountered most often: *party preference polls, election forecasts,* and *voting preference polls*. The way this terminology is used is not arbitrary. The three public opinion agencies – CVVM, STEM, and TNSF – have agreed on the exact meaning of these terms, and they use them in an identical manner in order to avoid any misunderstandings.[7] The results gathered in each type of research are processed using a different kind of methodology in each case, and are capable of providing us with an image of electoral preferences from "a different point of view".

Party preference polls tell us how the adult population declares its decisions in elections. The basis (100 %) is comprised of the responses provided by all adults included in the study who are entitled to vote. Alongside specific responses, referring by name to a particular party, there can also be found answers like "I don't know who I'd vote for", "I wouldn't vote at all", and other, similar responses. Party preference polls attempt to provide an image of the voting stance of the entire adult population. This, however, is by no means identical to the group of people who actually vote, that is, the adult citizens who ultimately do take part in the elections. The record low in voting participation in the 2002 elections (58 %) clearly demonstrates that an even smaller group of citizens than that which declares itself in favour of a certain party is what ultimately determines the actual outcome of the elections. Party preferences only give information on declared attitudes. These, however, can differ from final decisions. What is practically impossible to trace within party preference polls is what the definitive attitudes of "undecided" voters will be, and the degree to which the "decided" voters will ultimately not turn out to participate in the voting at all. In simplified terms, it could be said that party preference polls suffer most from the discrepancy between declared voting decisions and real voting decisions.

Election forecasts represent an attempt at minimising these problems. Their purpose is to achieve the most precise prediction possible of the election results. Putting together election forecasts is considerably more complicated, and the way in which one agency proceeds in doing so is often maintained a secret. Generally, data from surveys are cleaned and "weighed" with the aim of eliminating any distorting factors. This methodology could be compared to a kind of "cookbook", whose recipes are based on experiences drawn from the previous elections. The authors of the forecast attempt to map the differences between previous election surveys and the actual election results. They attempt to uncover trends that they then, in the form of "weights", apply to current data. Understandably, the basis of all good forecasts (in addition to quality data) lies in experience with a longer series of elections, and, if possible, a consolidated party system. In the Czech case, the lack of a sufficiently long series of parliamentary elections, and moreover elections occurring within a similarly structured party system, is the main impediment to achieving quality forecasts.

[6]Most of the research results this article works with are drawn from the websites of: Czech Radio, http://www.rozhlas.cz/volby2002cerven/pruzkumy/ - STEM; iHNed, http://www.parlamentnivolby.cz – TNSF; Lidové noviny - STEM and press releases issued by CVVM – see references.

[7]Nonetheless, from time to time in some media it is possible to come across comparisons of incomparable results.

This is about as close a definition of the methodology of election forecasts that can be given in general. But the definition provided by one agency, TNSF, which, prior to the 2002 elections, made the most frequent election forecasts, may be useful: *An election forecast is an estimate of the actual outcome of the elections. A mathematical model that is applied to produce an estimate takes two specific aspects into consideration. First, it reflects the degree of likelihood that a potential voter will genuinely take part in the voting, and second, it draws into consideration the fact that at the real elections it is impossible to retain the option of being undecided (i.e. the answer "I don't know"), and the undecided voters are allocated among the candidate parties.* [iHNed, www.parlamentnivolby.cz, 10 June 2003].

It is clear that it is not really possible to compare the outcomes of party preference polls and election forecasts, especially since the basic group drawn on for producing the outcome (100 %) refers in each case to a different group among the public. Party preference polls also feature the percentages of "undecided", those who "don't know", or those "who would not vote". The basic group on which the results are based is the entire adult population with the right to vote. Conversely, an election forecast defines as the basic group behind its outcome only that part of the public that the authors of the forecast can assume will turn out to vote. Therefore, in the case of election forecasts the sum of the percentages for the individual parties always equals 100 %, while in the case of party preferences the final sum is usually much lower (and the remainder to 100 % is made up of those who are undecided or unwilling to participate, etc.). As a result, the percentages for individual parties in party preference polls are proportionally lower than they are in election forecasts. Nonetheless, situations do at times arise in which it is necessary to draw both types of results into a comparison. There is one solution to this, albeit not an ideal one, that renders party preferences more comparable with election forecasts. It involves changing the basic group in the party preferences from the entire adult population to comprise only decided voters (here it is necessary to accept the simplification that such respondents are those who have indicated the name of a specific party). The percentages for the individual parties are re-weighed so that their sum is equal to 100 %. This procedure is controversial because it could contribute to poor quality of outcome. But it is accepted by the research agencies.[8]

A similar procedure is used to prepare the third type of survey: **voting preference polls.** The aim of this type of survey is to provide an image of what the outcome of the elections will be. However, the results are based exclusively on survey data that is not modified in any way, unlike in the case of election forecasts. This type of survey was used primarily by the agency STEM.

The questioning methodology of each of the three agencies is different. Moreover, CVVM and STEM change their standard methodologies to abbreviated versions around election time. TNSF ascertains voting preferences using a "closed question". STEM and CVVM, outside the election period, use an "open question", but once it becomes perfectly clear which parties are going into the elections, both these agencies switch to the use of closed questions. The open question leaves the response solely up to the respondent. The respondent answers the question about which party they would vote for without any lead. In a closed question, the respondent is provided with a card containing a list of parties (or in some cases a number of cards with the names of the different parties on them). This issue came to be a focus of interest in connection with ascertaining Quadcoalition and later Coalition preferences, where open questions proved to be at the very least impractical, problematic, and evincing a tendency towards misinterpretation.[9] There are ongoing discussions surrounding the suitability and advantages of different methods of questioning, but these are not the subject of this paper.

[8]CVVM proceeded in this manner when it made public the results of the final survey carried out prior to the elections [Party Preferences Prior to the Elections… 2002].

[9]For details see Vlachová 2002.

The timing intervals applied in pre-election research at all three agencies also varied considerably. Only the "academically-based" CVVM stayed with regular surveying. The interval between the final two studies was shortened from the regular one-month interval to fourteen days. The other two agencies delivered commercial studies of preferences developed for the media. Owing to this, the frequency with which they were conducted grew considerably. From the middle of February, TNSF was producing regular election forecasts weekly. The ambitious project began almost four months before the elections. The results were intended for the exclusive use of the iHNed internet news server (the final survey for both the daily newspaper *Hospodářské noviny* and Radio Impuls). About one month before the elections STEM was providing regular surveys every day! Theirs were for the exclusive use of the daily newspaper *Lidové noviny.*

In addition to the usual countrywide polls, some novelties appeared on the scene in the year leading up to the 2002 elections. Both TNSF and STEM initiated **surveys from the electoral districts.** STEM made an ad hoc study of party preferences in all the districts in the country. The surveys were carried out in three waves: May 6-7 – in five districts, May 13-14 – again in five districts, and May 20-21 – in four districts. The results were delivered for the exclusive use of Czech Television and Czech Radio. The survey results were made public gradually, with one district presented each day. TNSF came up with an even more ambitious project. Each week, in addition to statewide forecasts, it provided 14 district election forecasts. For each district, then, there emerged a time series at one-week intervals almost three months in length. These studies were delivered for the exclusive use of the news server iHNed.

At first glance, the **methodology of data collection** used in the research does not differ much from agency to agency. All three agencies use the method of quota sampling, though the structure of quota characteristics differs somewhat at each agency.[10] The most apparent difference is found in the size of the samples. CVVM and TNSF generally interview around one thousand respondents, while STEM typically uses a sample twice that size. All three are intended to be representative for the Czech Republic. The two-thousand-person sample used by STEM is meant to ensure a narrower reliability interval, thus a slightly smaller statistical error. In the daily surveys STEM carried out for *Lidové noviny,* samples of around one thousand respondents were used. But the individual surveys in each of the electoral districts were performed in a completely different manner. In each district STEM interviewed approximately 600 respondents over the course of two days. The question is whether the agency was able to sustain the representativeness for each specific district.[11] If it were, it would be possible to consider the size of the sample gained through quota selection as still sufficiently large. In not one case did TNSF indicate in any of the districts the number of respondents in the samples. They gave only the total number of respondents for all the districts combined, which was around 4000 respondents in sum, as each week the agency surveyed a sample of around 1000 respondents (for all of the Czech Republic). That means that the average number of respondents per district was only around 70 per week! A sample of that size is of course completely insufficient. To compensate, the agency multiplied back in time, adding to the formula the last three previous surveys (each of which was also comprised of 1000 respondents), and on the basis of this combined sample the agency produced its results.

[10]For details on the quota characteristics of the individual agencies see first section by Martin Kreidl.

[11]The reports by STEM, publicised on the website of Czech Radio (one of the clients who commissioned the surveys) indicate representativeness for the population of the Czech Republic in the first five surveys of the first wave of surveys carried out, while in the remaining nine surveys in the next two waves it indicates representativeness only for the specific district.

In addition to the problem that the sample in each district was insufficiently large, the other fundamental problem was that the data was collected only for one month in total. The regularly publicised weekly surveys from the districts were created on the basis of an insufficiently large sample which was moreover rotated each time by only one quarter. TNSF was indicating representativeness only for the voters of the entire Czech Republic, not for the voters of the individual districts.[12]

An Analysis of the Results

To begin the analysis, we will first look at the most focused on and widely discussed subject – how the individual agencies managed to **predict the overall results of the elections.** A comparison between the last surveys conducted by all three agencies before the election and the election results is presented in graph G3 (see appendix at the end of the article). TNSF has two surveys in this graph, and the reason for this is that TNSF was considerably more accurate in its next to last election forecast than it was in its final forecast, publicised four days later. In the case of CVVM, its final continuous survey is included in the comparison, the data for which was collected just shortly prior to the elections (5-12 June 2002). As a consequence, the results of the research were publicised after the elections – on 27 June 2002 – i.e. two weeks later.[13] TNSF and STEM publicised their results in the form of election forecasts, while CVVM did so in the form of party preference polls re-weighted without the responses "I don't know". The outcome of each agency is thus comparable with others and with the actual results of the elections.

With regard to the actual election results, it was CVVM that came out best. Paradoxically, this particular outcome was actually only a (re-calculated) party preference poll and not an election forecast. The distortion index[14] "D", equal to the value of 3, signifies a very satisfactory result. The most significant deviation was an overestimate of the results for ODS by 2.5 %, while in the case of the other parties the deviations were around 1 %. CVVM correctly estimated the order of the parties as ČSSD, ODS, KSČM and the Coalition. This very good outcome in comparison with the others may have been positively influenced by the timing of data collection, which was completed one day before the elections. That was a four-day advantage over the other two agencies. This kind of "luxury", collecting data immediately prior to the elections and publicising them after the elections, could only be afforded by an academic agency.

[12]See the reports by TNSF publicised in the election website of iHNed – www.parlamentnivolby.cz.

[13]The head of CVVM issued the following written comments on the timing of the research: "The Centre for Public Opinion Research, as a part of its project titled Our Society 2002, dealt with research on the formation of election decisions as a key subject in this "ultra-election" year. The basic issue CVVM addressed related to the factors that influence a voter's decision to participate or not to participate in parliamentary elections, as well as when a voter reaches the decision to submit their vote and how they are influenced in making this decision. The timing of fieldwork was thus adapted to ensure that it was possible to gather opinions "at the last minute", thus, at the time when the pre-election campaign was peaking. Given that the timing of the research (the fieldwork took place two days before the elections to the Chamber of Deputies were held) made it impossible to publicise the results before the elections (by law the three-day moratorium on surveys), the results of the June survey in the Our Society 2002 research were publicised after 27 June 2002."

[14]As a distortion index I am using half of the sum of the absolute values of the differences between the gain in the elections and the predicted gain for each party. This is analogous to the Loosemore-Hanby proportionality.

index – "D" $D = \frac{1}{2} \Sigma | g_i - p_i |$ (g – gain of party, p - prediction for party), see Pennisi 1998. The values of the index start from zero (i. e. that prediction for all parties is absolutely identical to the results) and increase along with the growing prediction error.

The private agencies, STEM and TNSF, could hardly be commercially successful were they to time their research similarly. Very close to the quality of CVVM's preference polls was the next to last election forecast of TNSF, publicised 5 June 2002. At D=3.3, only slightly higher than that of CVVM, the distortion index demonstrates a very accurate forecast. In its next to last research TNSF also correctly estimated the order of the parties in each position. The ensuing and final forecast of this agency, publicised 9 June 2002, was however a failure. They were unable to determine the correct order for the third and fourth places – KSČM and the Coalition. The value of the distortion index, at D = 8.5, indicates serious inaccuracy. The results for the communists were underestimated by more than 6 %, while the total gain for the parties that did not gain entry into the Chamber of Deputies was considerably overestimated (by 7.1 %). The forecast by STEM turned out in a similarly inglorious vein, though ever so slightly better with its distortion index at D = 7. The distortions were nominally lower, around three to four percent, but with the exception of ČSSD they occurred equally among all the parties. STEM also did not manage to correctly predict the order for the third and fourth places.

It is worth mentioning that the size of the sample in no way had a determining impact on the accuracy of the research. CVVM interviewed 956 respondents, which is less than half as many as in the last forecasts by STEM and TNSF. The size of STEM's sample in the daily research surveys hovered around 1000 respondents – a minimum of 500 respondents per day, with the sample being comprised at all times of data collected over the last two days [sim. 2002]. However, the final forecast was boosted. Within a mere two days 2376 respondents were interviewed. In addition, the data gathered over the last three weeks as part of the daily polls, i.e. 12 055 respondents, was also to be taken into account [sib. 2002]. TNSF also boosted its sample in a similar manner. The standard one thousand was doubled to 2006 respondents, who, similarly, were interviewed over the course of only two days (June 7-8). The question is whether the effect of boosting the size of the sample, and collecting the data within a very short timeframe, was not perhaps counterproductive, and whether it did not in fact lead to overloading the network of interviewers and consequently also to poor quality field work. This question will be raised again further on in the article.

Also worth noting are the **time series** of the individual research studies.[15] These are presented in graphs G5, G6 and G7. These series are not mutually comparable. Graph G5 illustrates the series of daily research surveys on voting preferences conducted by STEM for *Lidové noviny,* which were carried out for a period of three weeks prior to the elections. Not once throughout the entire period, nor even in its conclusion, did STEM correctly predict the order of the third and fourth places (KSČM, the Coalition). ODS dominated first place for almost the entire period. Only during the last two days did ČSSD take over the leading position. In comparison with the other agencies, this advancement of ČSSD into first place was considerably delayed.

The TNSF series (graph G6) reflects the regular weekly forecasts prepared for the iHNed server, which ran from the middle of February 2002. Here, ČSSD first managed to move into first place at the end of April. After two weeks ODS took over the lead for fourteen days, after which ČSSD moved back into the lead for the final two weeks. The order of third and fourth places was determined correctly in the next to last research survey, but not however in the final one (see above). In the TNSF series, there are more notable "jumps" in evidence, particularly with regard to the total gain of the remaining parties.

The CVVM series (graph G7) is the "scarcest" from the perspective of survey intervals. It features party preferences drawn from regular monthly surveys during the period of one year running up to the elections. In CVVM's surveys the Social Democrats definitively settled in first place as early as October 2001.

[15]The dates in graphs G5 and G6 refer to the final day of data collection and not the date of publication.

From that time and for the next three-quarters of a year CVVM regularly declared the first and second positions identical to the results in the ensuing elections, that is, ČSSD in first place and ODS in second. CVVM managed to correctly predict the order of the third and fourth places in their final survey, which in contrast with the competition had an advantage in terms of the time of data collection. CVVM concluded data collection on June 12, while the competition had ceased on June 8 (so that the results could still be publicised before the elections). To a certain degree this fact could have helped in determining the correct order of KSČM and the Coalition. However, the CVVM time series features one basic problem. With the exception of the last two surveys the results were acquired using an open question. The way CVVM does it, the data for the Coalition are added up from the responses referring to all of its constituent parties. However, this kind of practice is questionable to say the least (see above, for details Vlachová 2002).

A separate chapter entirely are the **research surveys conducted in the individual districts.** In each district STEM carried out an ad hoc survey. TNSF gathered data in each district on a weekly basis, which gave rise to a series of eleven election forecasts with one-week time intervals. The methodology that both agencies used in the districts has already been sketched out above. Now we will take a look at the results.

The results of the surveys by STEM are presented in the series of graphs SG1. These are party preferences recalculated after eliminating the response "I don't know" so as to render them comparable with the results of the elections. The surveys were held over a period of three to five weeks running up to the elections, which in itself justifies some of the differences with regard to the election results. STEM managed to establish the correct order of all four parliamentary parties only in three districts. The victor in the polls, and the real victor in the elections, was correctly determined only six times. The polls in Southern Bohemia, Central Bohemia, the Olomouc district and the Vysočina district came out worst of all. In this case it is practically impossible to excuse the dramatic differences from the actual results as being a result of the pause in the time between the surveys and the elections. According to STEM, in Southern Bohemia, ODS was supposed to win by 11 % over ČSSD. But the reverse was true, and ČSSD overtook ODS by 4 %. They were also unsuccessful in determining the third and fourth place order. The outcome of the research in Central Bohemia was similar. In the Vysočina district, STEM indicated ODS as slightly favoured over ČSSD. That in itself was shocking for a district that is typically leftwing. The actual results of the elections shifted ODS off into third place, behind even the communists, and losing to the victorious Social Democrats by almost 13 %. Following the same pattern the research also failed in the Olomouc district. There is no point in going on to fill this paper with a list of problematic preference predictions.

The graphs provide a sufficient overview. It is, however, necessary to stress that in some districts STEM came out substantially better. It presented a very good estimate in the Karlovy Vary district, the smallest electoral district in the country. Also, an absolutely precise prediction was achieved in Prague, where the largest deviation was still only a mere six tenths of a percentage point!

It is very difficult to find the words appropriate to describe the district forecasts produced by TNSF. The very methodology of the research is barely acceptable (see above). The time series of eleven forecasts often lacked a continuous tendency. If we take into consideration the fact that in each survey the sample was altered by only one-quarter of respondents, some of the dramatic week-to-week fluctuations seem almost incomprehensible.

Some of the results publicised by TNSF were an absolute failure. Here are some random examples. According to the final forecast (one week before the elections) the Coalition was supposed to win with 25.6 % of the vote in the Plzeň district. Instead, the party finished in fourth place, with less than half the predicted votes, i.e. 11.8 %. In the Southern Moravian district the communists were predicted in the final forecast to gain 8.5 % of the votes. In reality they fared twice as well with 19.8 %. In

the Olomouc district, according to TNSF, ODS was to gain victory with 31.2 % of the votes. Instead it finished third with 20.3 %. Similar fluctuations were also registered in some of the district surveys conducted by STEM. The fundamental difference though is the time when the surveys were conducted, as STEM carried out their surveys approximately one month before the elections, while the fluctuations described in the surveys of TNSF stem from data collection concluded a week before the elections! The unparalleled peak of "professionalism" is the time series for the Karlovy Vary district. The forecasts that TNSF permitted itself to publish on the Karlovy Vary district are simply shocking (see graph G8). It should be recalled that in this district ČSSD won (29.3 %), followed by ODS (24.7 %), KSČM (22 %) and the Coalition (8.9 %). Non-parliamentary parties gained 15.2 % of the vote. However, in the next to last forecast the non-parliamentary parties were predicted together to gain 42.3 % of votes; the Green Party, for example, was to gain 16.4 %. According to TNSF at the end of March, ODS was to win with more than half of all the votes at 52.8 % (ODS does not even get this much in Prague). In the elections in the Karlovy Vary district, ODS obtained less than half (24.7 %). The agency was not worried even about publicising the deathly forecasts relating to the Coalition. One week before the elections it was predicted the Coalition would gain 1.1 % of the vote, and one week later absolutely zero. It is better not even to contemplate the consequences that news of this kind may have. The Karlovy Vary district was the only one in which the Coalition did not obtain a single mandate even though it obtained almost 9 % of the votes. If TNSF indicated 0 % for the Coalition, this means that in four weeks of surveys not a single respondent was found in any of the samples that expressed support for the Coalition. Such an occurrence needs no further comment. The Coalition was not, however, the only victim in the "inventive approach" maintained by this agency. According to TNSF in the first half of April, ČSSD was to gain only 2 % of the vote!

In the SG2 series of graphs, a comparison is presented between the final district forecasts prepared by TNSF, composed one week before the elections, and the election results in the individual districts. It should be added that the final forecasts often considerably improved in comparison with some of the at times unbelievable values produced in the previous ten surveys in the series. Overall, however, the forecasts by TNSF achieved poorer results than the district preferences produced by STEM, even despite the clear advantage they had owing to the time of data collection. However, among the final district forecasts by TNSF it is possible to find some very successful ones. The forecasts for the Vysočina district (a separate graph for the time series - G7), and the Pardubice and Ústí nad Labem districts came out best of all. Even Prague came out relatively well. But that does little to make up for the problems that have been described above.

The district surveys featured one more point of interest. In the case of both agencies there were clearly discernible trends that could be traced either systematically strengthening or weakening certain parties. Even more interesting is the fact that these trends were at both agencies very similar. The polls by STEM systematically overestimated the gains of the rightwing parties of ODS and the Coalition and systematically underestimated the gains of the leftwing parties of ČSSD and KSČM. This was also the case in the last pre-election forecasts by TNSF. Table 1 presents the average deviation values, the highest overestimation values and the highest underestimation values, and the numbers of such cases. How can these visible trends be explained? It is not possible to provide any certain answer to this question. Most likely, however, it appears that the greatest source of the deviation lies again in the interviewer network, and specifically in its social composition and integrity, and in inconsistent supervision on the part of the agency.

Table 1. Overestimation and underestimation of trends in district surveys

	ODS		ČSSD		Koalice		KSČM	
	STEM	TNSF	STEM	TNSF	STEM	TNSF	STEM	TNSF
average deviation	+3,3%	+3,8%	-4,7%	-2,2%	+3,2%	+1,2%	-3,6%	-2,8%
no. of overestimation	12x	12x	2x	3x	12x	8x	0x	1x
highest overestimation	+7,3%	+10,9%	+2,7%	+5,3%	+5,6%	+13,8%	0%	+7,6%
no. of underestimation	2x	2x	11x	11x	2x	6x	14x	12x
highest underestimation	-1,4%	-1,9%	-9,1%	-9,8%	-0,6%	-7,8%	-6,9%	-11,3%

Conclusions

There were whole series of pre-election polls conducted and the quality of these surveys evidently varied. Even the polls within a single agency differed qualitatively. In any evaluation it is necessary to distinguish between the typical statewide surveys for the entire Czech Republic and the surveys aimed only at the individual districts.

It appears that the quality of the research surveys described here is connected to some degree with their quantity, that is, with the number of surveys conducted. CVVM alone among the agencies remained relatively faithful to the standardly followed time intervals in their research surveys. The field collection of data was conducted in a typical manner and the interviewer network was not burdened more than usual. The private agencies carried out series of surveys for different bodies in the media. In the history of surveys on voting preferences the volume of this research was probably the largest such research to date in this country. Over the course of four months TNSF carried out regular surveys of one thousand respondents every week. The interviewer network was heavily burdened and we can only speculate as to whether perhaps it was in fact overloaded. In the case of such frequent interviews it is considerably more difficult to inspect the quality of the data collected. The shorter questionnaires, containing fewer questions (typical for this kind of survey), are themselves a source of difficulty in uncovering potential negligence or even deception in the work of the interviewers. It is indicative that in the final survey in the entire series conducted by TNSF the deviation from the actual results was particularly striking. The final forecast that was developed on the basis of a sample twice the size of all the previous samples (with double the burden put on the network of interviewers) arrived at a considerably worse outcome than the very accurate next to last survey based on 952 respondents. Of course, it is possible that the reasons for this failure lie elsewhere, and the large demands put on the interviewers may have played only a marginal role.

In the case of STEM it appears more likely that the interviewer network was indeed overloaded. For a period of three weeks STEM was producing surveys on a daily basis. The agency declared that it was managing to conduct an interview with at least five hundred people daily [sim. 2002]. In comparison with standard conditions this is a disproportionately greater burden.

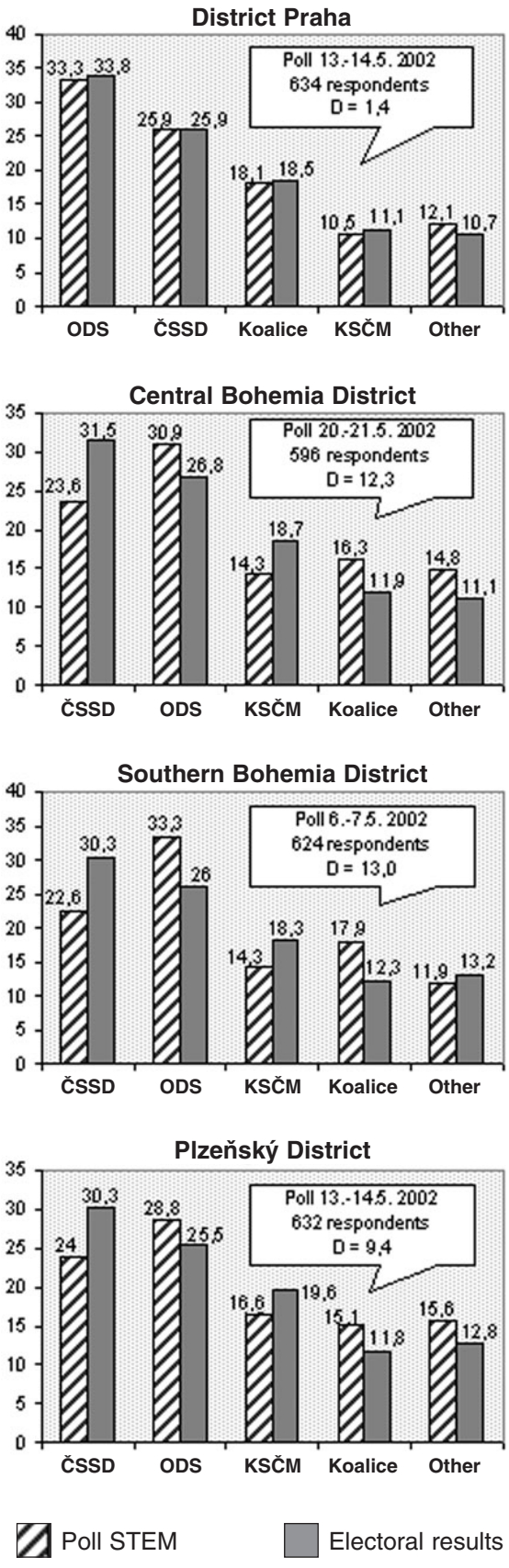
For the interviewer, this task represented a practically non-stop three-week marathon. In standard circumstances the interviewers do not work so continuously. Moreover, it must be remembered that just before this series was initiated the STEM interviewers had absolved what in terms of work volume were the similarly demanding research surveys in the individual districts. Understandably, the interviewer network reacted to the disproportionately high quantitative demands by modifying the quality and methods of their research in order to be capable of fulfilling the intensified demands. The conclusion that the poor quality of the work produced by the interviewer network, and by extension the poor quality work of the agency itself, is supported by the conclusions reached by Martin Kreidl, which are presented in the first part of this study.

The classic surveys conducted for the entire Czech Republic were more or less accurate. In this connection, it cannot be said that any one of the agencies failed completely or was absolutely off the mark. That however is not true in the case of the surveys carried out within the individual districts. The dubious quality of these surveys was outlined above. The suspicion remains then that there was a problem with overloading the network of interviewers in the case of both agencies carrying out district surveys (STEM and TNSF). STEM interviewed around 600 respondents in each district over the course of a mere two days. Each interviewer was required to carry out a several times higher than usual number of interviews, and from the methodological perspective in particular this becomes unsustainable. There also exists the unproven hypothesis that the quality of the district surveys was ignobly influenced by the system whereby interviewers were remunerated.[16] In the case of TNSF the interviewer network was likely less burdened. However, the problem is augmented by the more serious problem of their insufficiently large samples, the likely lack of representativeness, and the weekly overlays to produce a monthly outcome (see above).

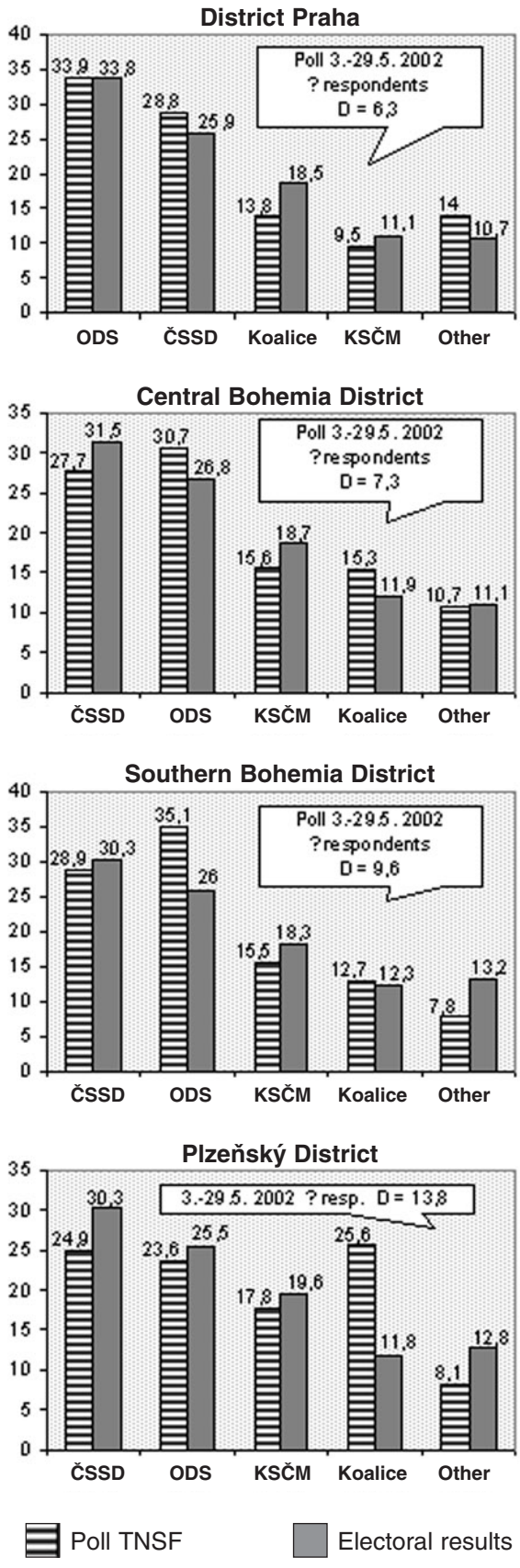
The district surveys conducted by both agencies, but especially those by TNSF, come across as a bad experiment. However, the experiment did not only affect the survey methodology and the interviewer network, etc. It was much more dangerous an experiment that affected the influence of surveys on voting behaviour and voting results. Last but not least, it was an experiment involving public confidence in public opinion research itself. Unfortunately, the results of these (we believe) unintentional experiments are not yet known. Clearly, the private sociometric agencies are moving into the market, where like any other business they are striving to survive. Minimising costs, maximising gain, and fulfilling the unfulfillable – that's the market logic. But the basic purpose of conducting surveys must not be forgotten. Their aim is to provide the most accurate possible portrait of society, its moods and its attitudes. No research should allow itself under market pressure to veer from its basic purpose. No research so broadly publicised should be perceived by its authors as a secondary product, warranting less attention and integrity than any other research work. Every serious agency should keep at the top of its mind the professional and ethical criteria that the field demands and should be prepared to refuse any task that cannot be carried out in harmony with these requirements.

[16] According to a reliable source the interviewers were remunerated with only small sums. These may have been in proportion with the smaller number of questions in the questionnaire per respondent, but they were not in proportion with the fact that it was necessary to seek each respondent in accordance with the requirements of the quota characteristics, in the same manner as in surveys containing dozens of questions.

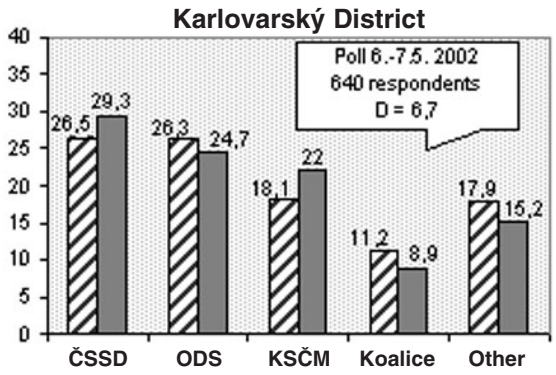
SG 1. Party preference polls STEM (in %) re-weighed to electoral preferences

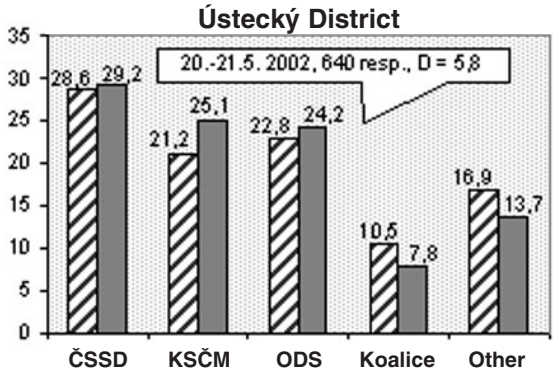


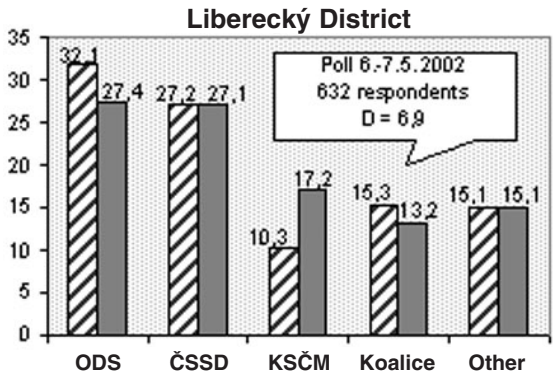
SG 2. Election forecasts TNSF (in %)

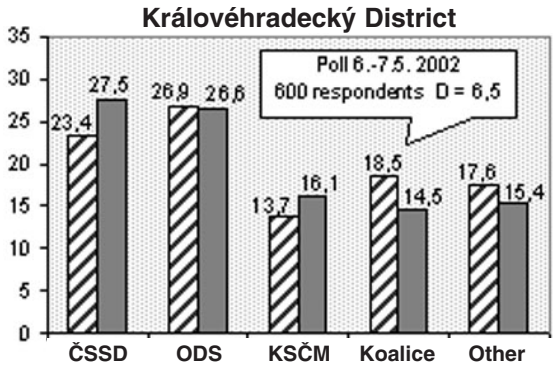


SG 1. Party preference polls STEM (in %) re-weighed to electoral preferences

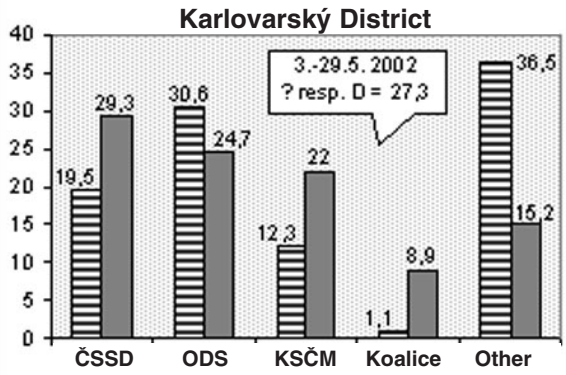


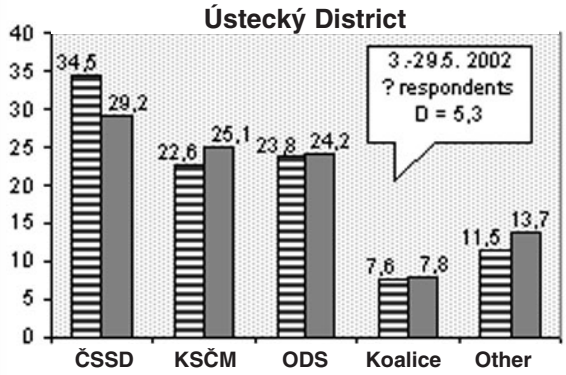


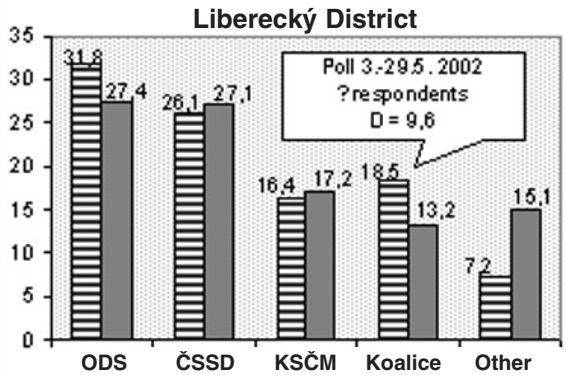


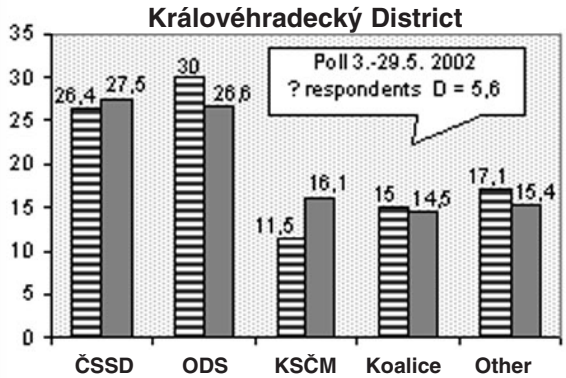


SG 2. Election forecasts TNSF (in %)



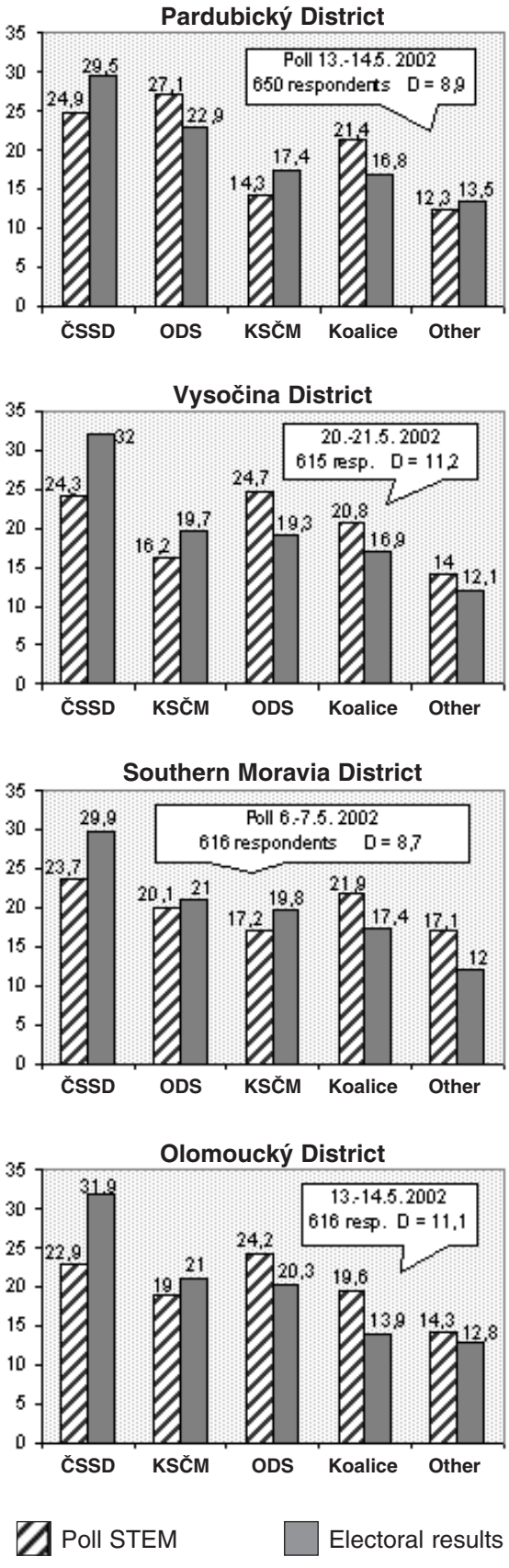




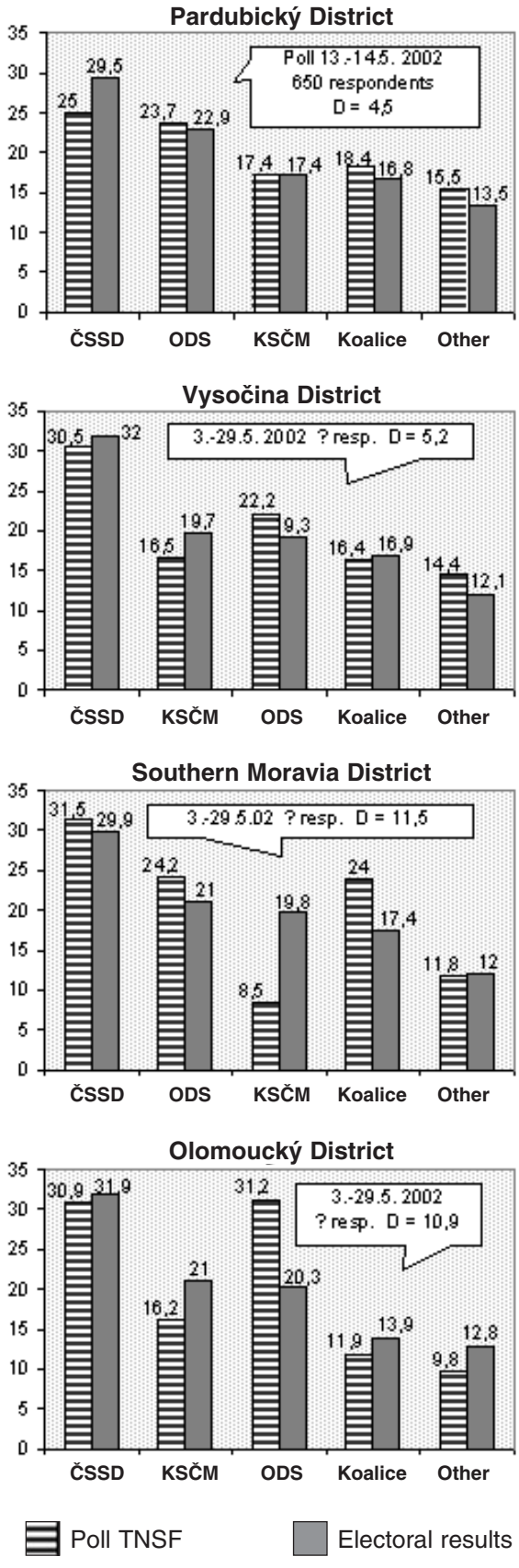


Poll STEM | Electoral results | Poll TNSF | Electoral results

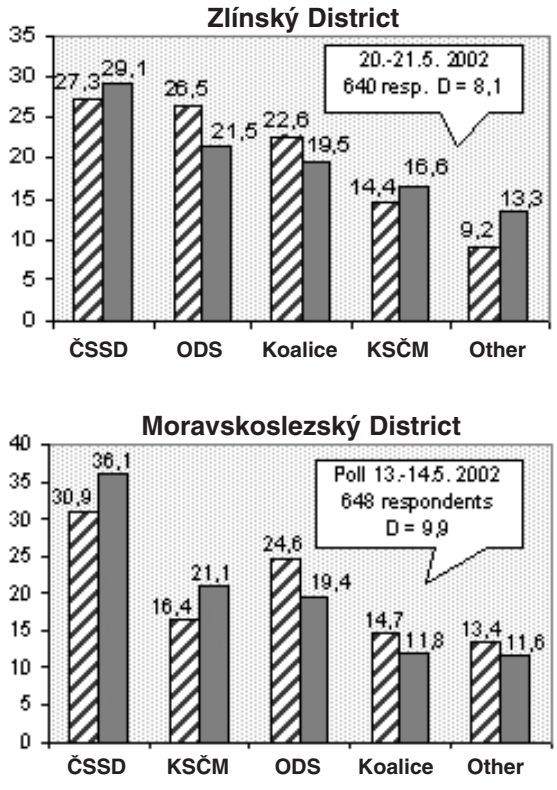
SG 1. Party preference polls STEM (in %) re-weighed to electoral preferences



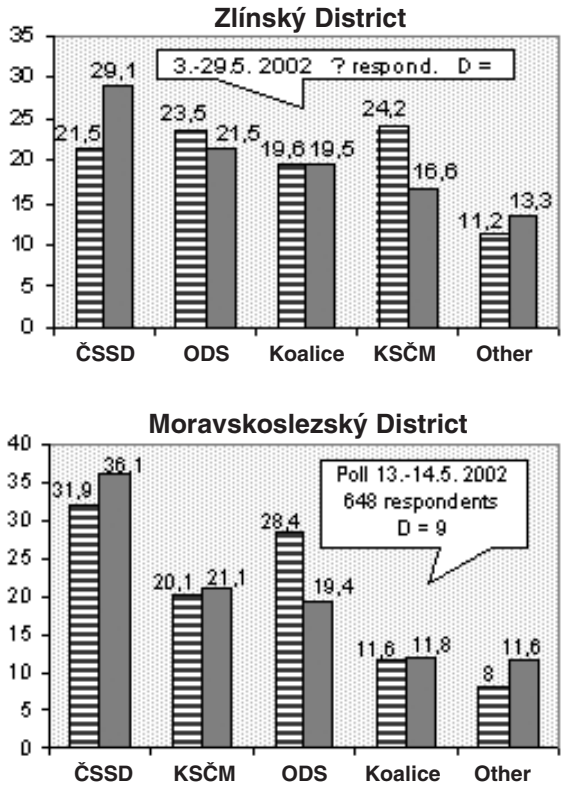
SG 2. Election forecasts TNSF (in %)



SG 1. Party preference polls STEM (in %) re-weighed to electoral preferences

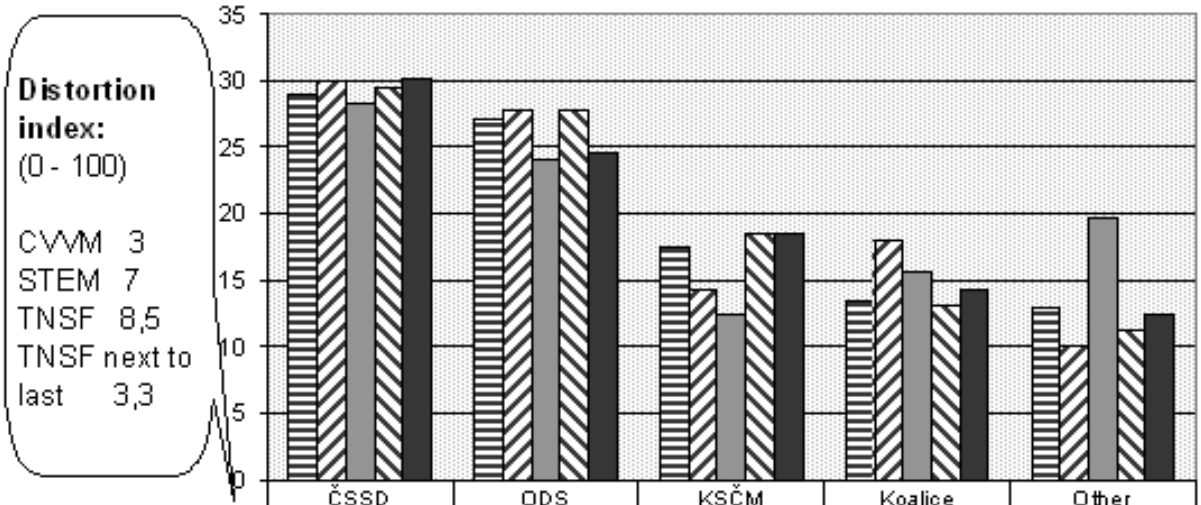


SG 2. Election forecasts TNSF (in %)



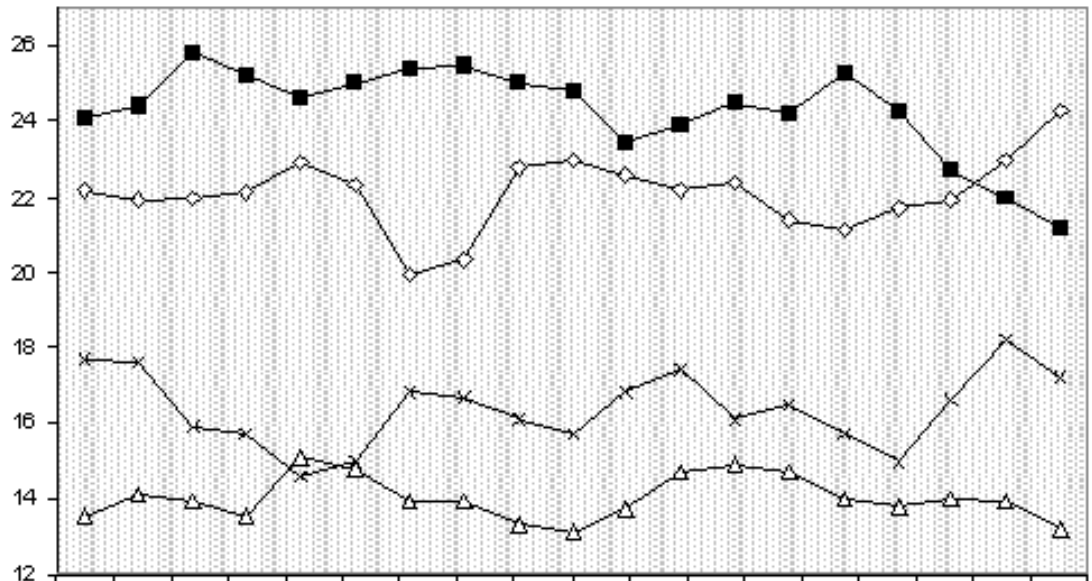
Poll STEM | Electoral results | Poll TNSF | Electoral results

G3. The last polls before elections



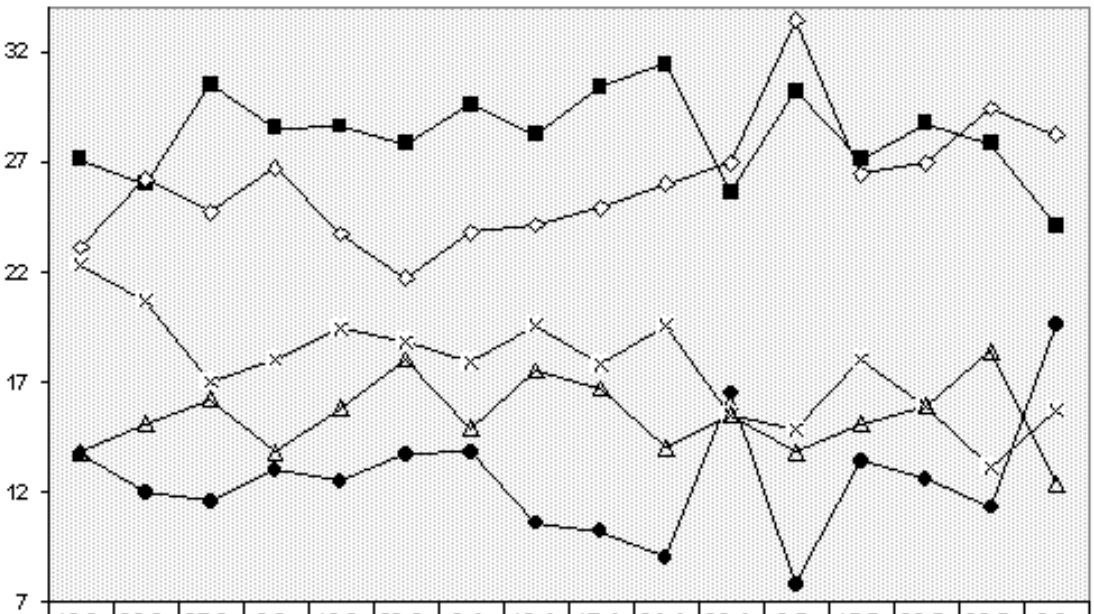
	ČSSD	ODS	KSČM	Koalice	Other
CVVM poll	29	27	17,5	13,5	13
STEM poll	29,9	27,8	14,3	18	10
TNSF poll	28,2	24,1	12,4	15,7	19,6
TNSF next to last poll	29,4	27,8	18,4	13,1	11,3
Electoral results	30,2	24,5	18,5	14,3	12,5

G4. Daily party preference polls trend of STEM (in %)



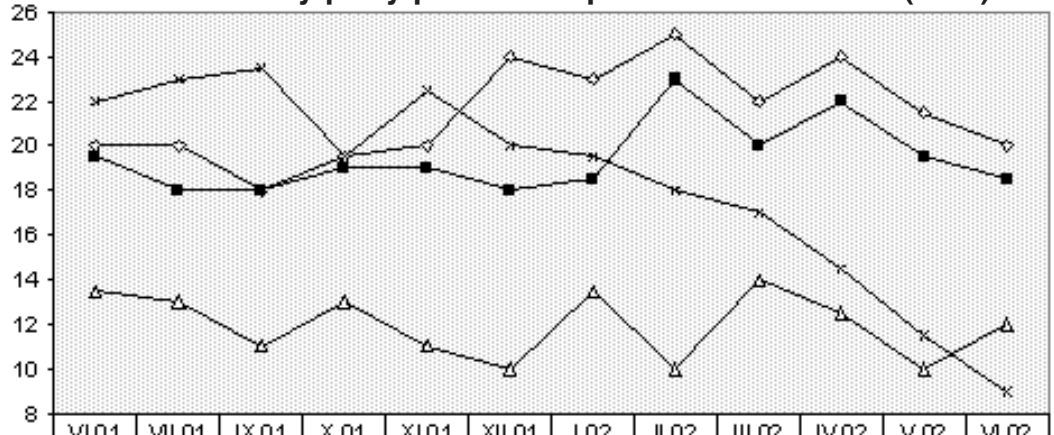
	18.	20.	21.	22.	23.	24.	25.	27.	28.	30.	31.	1.	3.	4.	5.	6.	7.	8.
ČSSD	22,2	21,9	22	22,1	22,9	22,3	19,9	20,3	22,8	22,6	22,2	22,4	21,4	21,1	21,7	21,9	23	24,3
ODS	24,1	24,4	25,8	25,2	24,6	25	25,4	25,5	25	23,4	23,9	24,5	24,2	25,3	24,3	22,7	22	21,2
KSČM	13,5	14,1	13,9	13,5	15,1	14,8	13,9	13,9	13,3	13,7	14,7	14,9	14,7	14	13,8	14	13,9	13,2
Koalice	17,7	17,6	15,9	15,7	14,6	15	16,8	16,7	16,1	16,8	17,4	16,1	16,5	15,7	15	16,6	18,2	17,2

G5. Weekly electoral forecasts trend of TNSF (in %)



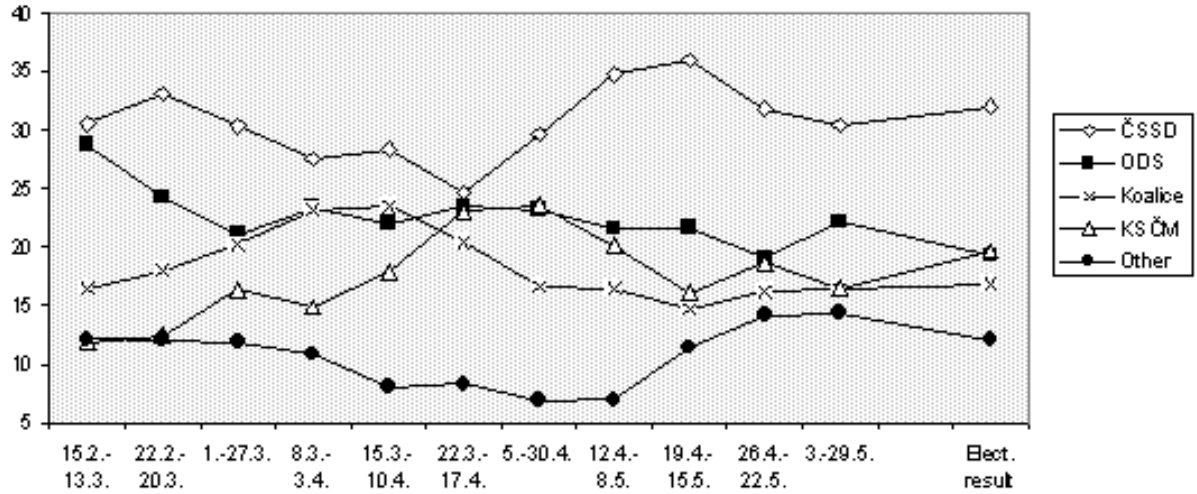
	13.2.	20.2.	27.2.	6.3.	13.3.	20.3.	3.4.	10.4.	17.4.	24.4.	30.4.	8.5.	15.5.	22.5.	29.5.	8.6.
ODS	27,1	26	30,5	28,5	28,6	27,8	29,6	28,2	30,4	31,4	25,6	30,2	27,1	28,7	27,8	24,1
ČSSD	23,1	26,2	24,7	26,7	23,7	21,7	23,8	24,1	24,9	26	26,9	33,4	26,4	26,9	29,4	28,2
Koalice	22,3	20,7	17	18	19,4	18,8	17,9	19,6	17,8	19,6	15,5	14,8	18	15,9	13,1	15,7
KSČM	13,8	15,1	16,2	13,8	15,8	18	14,9	17,5	16,7	14	15,5	13,8	15,1	15,9	18,4	12,4
Ostatní	13,7	12	11,6	13	12,5	13,7	13,8	10,6	10,2	9	16,5	7,8	13,4	12,6	11,3	19,6

G6. Monthly party preference polls trend of CVVM (in %)

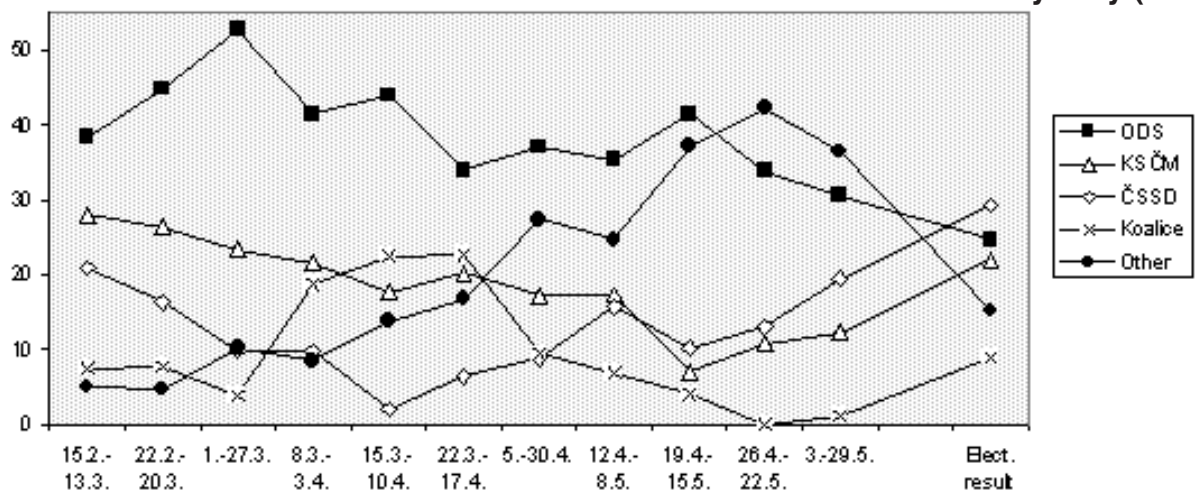


	VI.01	VII.01	IX.01	X.01	XI.01	XII.01	I.02	II.02	III.02	IV.02	V.02	VI.02
ČSSD	20	20	18	19,5	20	24	23	25	22	24	21,5	20
ODS	19,5	18	18	19	19	18	18,5	23	20	22	19,5	18,5
KSČM	13,5	13	11	13	11	10	13,5	10	14	12,5	10	12
Koalice (4K)	22	23	23,5	19,5	22,5	20	19,5	18	17	14,5	11,5	9

G7. The most successful electoral forecast of TNSF - District Vysočina (in %)



G8. The less successful electoral forecast of TNSF - District Karlovy Vary (in %)



References

„Agentury se rozcházejí v odhadech volebních výsledků“ 2002. *Hospodářské noviny*, 19. 4. 2002.

Bílek, O., S. Blechová 2002. „Exkluzivní volební prognóza agentury STEM.“ *Lidové noviny,* 10. 6. 2002.

Černý, Jiří. 2002. „Proč průzkumy selhaly.“ *Lidové noviny, příloha Orientace*, 20. 6. 2002.

„Jak vznikla prognóza agentury STEM.“ 2002a. *Lidové noviny,* 10. 6. 2003.

„Jak vzniká průzkum STEM pro LN.“ 2002. *Lidové noviny.* 20. 5. 2002.

Průzkumy agentury STEM pro ČT a ČRO. 2002. http://www.rozhlas.cz/volby2002cerven/pruzkumy/. Staženo 11. 6. 2002.

Průzkumy agentury TNSF pro iHNed - Volební informační server. 2002. http://www.parlamentnivolby.cz (2002). Staženo 11. 6. 2002.

Lipold, Jan 2002. „Proč se prognózy pokaždé aspoň trochu mýlí“. *Hospodářské noviny,* 10. 6.: 2.

Pečinka, Bohumil 2002. „Čtyřkoaliční preference nebyly správné. Rozhovor se sociolož kou Klárou Vlachovou.“ *Reflex,* 2. 5. 2002.

Pennisi, Aline 1998. „Disproportionality Indexes and Robustness of Proportional Allocation *Methods.“ Electoral Studies,* č. 1, str. 3-19.

„Poslední průzkum: Koalice se vzdaluje komunistům.“ 2002b. *Lidové noviny,* 10.6.2002.

Stranické preference na konci května 2002. Tisková zpráva CVVM. 6. 6. 2002.

Stranické preference před volbami 2002. Tisková zpráva CVVM. 27. 6. 2002.

Šídlo, Jindřich. 2002. „Volební preference: hra bez záruky.“ *MF Dnes,* 29. 4. 2002.

Vlachová, Klára. 2002. „Výzkumy volebních preferencí (Trnitá cesta od otázky k datům)“. *Proglas,* 3.

„Volební průzkumy bývají divoké.“ 2002 *Právo,* 1. 6. 2002.

Shrnutí

Potenciál každého, a tedy i sociálního, empirického výzkumu závisí kritickým způsobem na kvalitě a přesnosti měření stěžejních konceptů. Příkladem sociologického měření s masivními veřejnými a politickými konsekvencemi jsou výzkumy veřejného mínění a volební prognózy prováděné před každými volbami. V tomto textu se zaměřujeme na výzkumy provedené před volbami v roce 2002 v České republice a na validitu jejich výsledků. Rovněž zkoumáme validitu měření a měřící techniky u jednotlivých výzkumných agentur.

V textu kombinujeme dvě diametrálně odlišné techniky. V první části textu používáme pokročilé statistické techniky původně určené k měření konstruktové validity škál k identifikaci rozdílů v dosažené validitě mezi jednotlivými agenturami. V této analýze používáme sadu standardizovaných měřících nástrojů a argumentujeme, že veškeré zjištěné rozdíly ve validitě musejí být připsány terénní práci a dalším faktorům, které jsou specifické pro každou agenturu. Tento přístup identifikuje rozdíly v kvalitě měření a terénní práce mezi agenturami, ale není možné jej použít k seřazení agentur na pomyslném žebříčku kvality měření, neboť neznáme skutečnou validitu škály v populaci a nemáme k dispozici externí kritéria k jejímu zjištění.

Druhá sekce textu popisuje a zhodnocuje kriteriální validitu instrumentů používaných jednotlivými agenturami a výsledků, kterých je těmito instrumenty dosaženo. Tato část textu se zaměřuje na volební předpovědi a srovnává je se skutečnými volebními výsledky a spekuluje o možných zdrojích zjištěných odchylek, srovnává jejich velikost a směr, aby tak odhalila jejich systematickou komponentu. Protože v tomto případě máme k dispozici jasné externí kritérium ke zhodnocení volebních predikcí, získáváme tak silnou evidenci, že ne všechny volební prognózy zveřejněné před volbami v roce 2002 měly stejnou kvalitu.

Summary

The strength of any empirical research, and of social science research in particular, is entirely dependent on the quality and precision with which it is able to measure its concepts. One example of social measurement with massive public and political consequences are public opinion polls and election predictions made before each election day. In this text the authors focus on how polls were conducted prior to the 2002 elections in the Czech Republic and how valid their results were. They also seek to compare measures and techniques used by individual research agencies.

In order to assess the quality of measurement of the main research agencies the authors combine two rather different approaches. In the first section of the text, they use advanced statistical instruments originally intended to measure construct validity of scales, to look at differences in achieved validity between agencies. In this exercise they use a standardised set of measurements and argue that all other sources of measurement errors and differences in validity must be attributed to the fieldwork and other procedures that are agency-specific. While this approach enables them to identify differences in the quality of work that individual agencies supply, it is impossible to use it to rank the agencies performance, because there are no external criteria showing the actual validity of the scale we used.

The second section of the text describes and measures the criterion validity of instruments employed and results produced by research agencies. More specifically, it looks at election predictions publicised by agencies and compares them with actual election results. It also speculates about the sources of deviations, compares their sizes and directions to see if they appear to have a systematic component. Because there exists a clear criterion against which election predictions can be contrasted, there is rather strong evidence that indeed not all results that were circulated in the public sphere before the 2002 election were of the same quality.

Zusammenfassung

Das Potential sowohl jeder empirischen, also auch sozialen Forschung, hängt in kritischer Weise von der Qualität und der Genauigkeit der Messung der grundlegenden Konzepte ab. Ein Beispiel soziologischer Messungen mit großen öffentlichen und politischen Konsequenzen sind die Erfassung aktueller Meinungstrends und Wahlvorhersagen, die vor jeder Wahl durchgeführt werden. Im Text konzentrieren wir uns auf Erhebungen, die vor den Wahlen 2002 in der Tschechischen Republik durchgeführt wurden, sowie auf die Validität ihrer Ergebnisse. Die Validität der Messungen und Messtechniken der einzelnen Meinungsforschungsinstitute werden ebenfalls untersucht.

Wir kombinieren dabei zwei diametral verschiedene Techniken. Im ersten Teil verwenden wir fortgeschrittene statistische Techniken, die ursprünglich zur Messung der konstruktiven Validität von Skalen gedacht war, für die Identifizierung der Unterschiede in der erreichten Validität zwischen den Instituten. In dieser Analyse verwenden wir eine Reihe standardisierter Messwerkzeuge und argumentieren damit, dass alle gefundenen Unterschiede in der Validität, erstens der Arbeit im Terrain zugeschrieben werden müssen und zweitens Faktoren, die für die verschiedenen Institute spezifisch sind. Diese Vorgehensweise macht Unterschiede in der Messqualität und der Arbeit im Terrain der verschiedenen Institute deutlich, kann aber nicht dazu dienen, die Institute in einer vorstellbaren Qualitätsskala bzgl. ihrer Methoden aufzureihen, da wir die Validität der Skalen der Bevölkerung nicht wirklich kennen und keine externen Kriterien zu ihrer Feststellung besitzen.

Der zweite Teil des Textes bewertet die Kriterien-Validität der verwendeten Instrumente der jeweiligen Institute und ihrer damit erzielten Ergebnisse. Dieser Abschnitt konzentriert sich auf die Wahlvorhersagen, vergleicht sie mit den Wahlergebnissen und spekuliert über die möglichen Quellen für die gefundenen Abweichungen (deren Größe und deren Richtung), um so deren systematische Komponente zu enthüllen. Weil wir in diesem Falle ein klares externes Kriterium zur Verfügung haben, um die Wahlvorhersagen zu bewerten, erhalten wir einen überzeugenden Nachweis dessen, dass nicht alle Wahlvorhersagen, die vor den Wahlen 2002 veröffentlicht wurden, dieselbe Qualität besaßen.

Publikace edice Sociologické texty/Sociological Papers:

SP 03:7 J. Krejčí: Výzkumy stranických preferencí, jejich uplatnění ve společnosti a jejich kvalita. 63 s.

SP 03:6 Z. R. Nešpor, P. Holub, M. Skovajsa: Proměny českých socioekonomických hodnot na přelomu 20. a 21. století. 67 s.

SP 03:5 P. Sunega: Objektivní a subjektivní hodnocení finanční dostupnosti bydlení v ČR v průběhu 90. let. 100 s.

SP 03:4 D. Hamplová: Vstup do manželství a nesezdaného soužití v České republice po roce 1989 v souvislosti se vzděláním. 50 s., 67 Kč

SP 03:3 J. Večerník: Work and Job Values in CEE and EU Countries. 50 s., 134 Kč

SP 03:2 F. Zich, O. Roubal, B. Spalová: Mezigenerační biografická konfigurace obyvatel české části Euroregionu Nisa. 120 s., 134 Kč

SP 03:1 P. Matějů, B. Řeháková, N. Simonová: Strukturální determinace růstu nerovností. 49 s., 80 Kč

SP 02:13 E. Rendlová, T. Lebeda: Výzkumy veřejného mínění – teoretické souvislosti a praktická aplikace. 52 s., 100 Kč

SP 02:12 M. Havelka, M. Tuček, J. Černý, J. Česal, M. Hudema: Skupinové mentality. 85 s., 154 Kč

SP 02:11 M. Hájek, T. Holeček, J. Kabele, J. Kandert, P. Kohútek, Z. Vajdová: The World of Hierarchies and Real Socialism. 108 s., 190 Kč

SP 02:10 A. Křížková, L. Václavíková-Helšusová: Sociální kontext života žen pracujících v řídících pozicích. 48 s., 107 Kč

SP 02:9 L. Linek, P. Rakušanová: Parties in the Parliament. Why, When and How do Parties Act in Unity?. 73 s., 81 Kč

SP 02:8 A. Křížková: Životní strategie manažerek: případová studie. 67 s., 129 Kč

SP 02:7 T. Kostelecký, J. Stachová, D. Čermák: Region a politika. 76 s., 116 Kč

SP 02:6 M. Hájek, T. Holeček, J. Kabele, J. Kandert, P. Kohútek, Z. Vajdová: Svět hierarchií a reálný socialismus. 97 s., 169 Kč

SP 02:5 P. Sunega, D. Čermák, Z. Vajdová: Dráhy bydlení v ČR 1960 - 2001. 97 s., 177 Kč

SP 02:4 Z. R. Nešpor: Reemigranti a sociálně sdílené hodnoty. Prolegomena k sociologickému studiu českých emigračních procesů 20. století se zvláštním zřetelem k západní reemigraci 90. let. 85 s., 148 Kč

SP 02:3 M. Lux: Spokojenost českých občanů s bydlením. 56 s., 93 Kč

SP 02:2 N. Simonová: The Influence of Family Origin on the Evolution of Educational Inequalities in the Czech Republic after 1989, 36 s., 73 Kč

SP 02:1 P. Machonin, M. Tuček: Zrod a další vývoj nových elit v České republice (od konce osmdesátých let 20. století do jara 2002), 64 s., 97 Kč

SP 01:12 M. Hájek, T. Holeček, J. Kabele, P. Kohútek, Z. Vajdová: Kdo se bojí hierarchií? Dědictví komunistické vlády; 99 s., 133 Kč

SP 01:11 H. Jeřábek a E. Veisová: 11 September. Mezinárodní internetový komunikační výzkum International On-line Communication Research; 60 s., 95 Kč

SP 01:10 J. Rychtaříková, S. Pikálková, D. Hamplová: Diferenciace reprodukčního a rodinného chování v evropských populacích; 83 s., 117 Kč

SP 01:9 T. Kostelecký: Vzestup nebo pád politického regionalismu? Změny na politické mapě v letech 1992 až 1998 – srovnání České a Slovenské republiky; 96 s., 133 Kč

SP 01:8 M. Novák a K. Vlachová: Linie štěpení v České republice. Komparace národní úrovně s příkladem konkrétní lokality; 32 s., 68 Kč

SP 01:7 D. Kozlíková: Romská otázka – překážka vstupu České republiky do Evropské unie?; 64 s.,100 Kč

SP 01:6 P. Soukup: ISSP – Životní prostředí; 74 s., 105 Kč

SP 01:5 J. Večerník: Mzdová a příjmová diferenciace v České republice v transformačním období; 66 s., 111 Kč

SP 01:4 F. Zich: The Bearers of Development of the Cross-Border Community on Czech-German Border; 54 s., 88 Kč

SP 01:3 P. Sunega: Adresný příspěvek na nájemné v prostředí České republiky: komparace vybraných modelů; 96 s., 138 Kč

SP 01:2 M. Kreidl: The Role of Political, Social and Cultural Capital in Secondary School Selection in Socialist Czechoslovakia, 1948-1989; 48 pp., 89 Kč

SP 01:1 P. Štěpánková: Income Maintenance Policies, Household Characteristics and Work Incentives in the Czech Republic; 40 pp., 77 Kč

SP 00:7 L. Simerská, I. Smetáčková: Pracovní a rodinná praxe mladých lékařek; 70 s., 95 Kč

SP 00:6 P. Machonin, L. Gatnar, M. Tuček: Vývoj sociální struktury v české společnosti 1988-1999; 70 s., 101 Kč

SP 00:5 K. Vlachová: Stranická identifikace v České republice; 38 s., 62 Kč

SP 00:4 M. Kreidl: What makes inequalities legitimate? An international comparison; 54 pp., 80 Kč

SP 00:3 D. Hamplová: Náboženství a nadpřirozeno ve společnosti (mezinárodní srovnání na základě jednoho empirického výzkumu); 64 s., 90 Kč

SP 00:2 K. Müller, V. Štědronský: Transformace a modernizace společnosti na příkladech vybraných institucí. První část případových studií: střední průmyslový podnik, softwarová firma, banka, různá zdravotnická zařízení, vysokoškolský institut; 116 s.,109 Kč

SP 00:1 M. Lux: The housing policy changes and housing expenditures in the Czech Republic; 64 pp., 82 Kč

Publikace edice „Working papers“

WP 99:11 M. Jeřábek (ed.): Geografická analýza pohraničí České republiky; 184 s.

WP 99:10 M. Kreidl, K. Vlachová: Rise and Decline of Right-Wing Extremism in the Czech Republic in the 1990s; 40 p.

WP 99:9 B. Řeháková: Vnímané a spravedlivé nerovnosti: vývoj v devadesátých letech a další souvislosti; 46 s.

WP 99:8 L. Brokl, A. Seidlová, J. Bečvář, P. Rakušanová: Postoje československých občanů k demokracii v roce 1968; 84 s.

WP 99:7 H. Jeřábek, R. Gabriel, M. Kříž, H. Malečková, M. Novák, E. Pilíková, K. Plecitá, J. Remr, A. Vlachová: Utváření postojů obyvatel českého města I. Lidé s vlivem a osobní mezilidská komunikace při utváření politických postojů v lokální komunitě; 136 s.

WP 99:6 H. Maříková: Muž v rodině: demokratizace sféry soukromé; 110 s.

WP 99:5 M. Musilová: Vývoj politiky rovných příležitostí mužů a žen v České republice v kontextu evropské integrace; 60 s.

WP 99:4 F. Zich: Nositelé přeshraniční spolupráce na česko-německé hranici; 108 s

WP 99:3 L. Buštíková: Známosti osobností lokální politiky; 68 s.

WP 99:2 A. Nedomová (editor), L. Buštíková, E. Heřmanová, T. Kostelecký, Z. Vajdová, P. Vojtěchovská: Trh bydlení, jeho regionální diferenciace a sociální souvislosti; 82 s.

WP 99:1 M. Tuček, E. Rendlová, M. Rezková, A. Glasová, J. Černý: Odraz společenských změn ve veřejném mínění 1990-1998 (analýza dat IVVM); 104 s.

WP 98:6 K. Müller: Modernizační kontext transformace, strukturní a institucionální aspekty; 82 s.

WP 98:5 L. Brokl, Z. Mansfeldová, A. Kroupa: Poslanci prvního českého parlamentu (1992-1996); 94 s

WP 98:4 M. Lux: Konzervatismus a liberalismus na pozadí percepce sociálního státu; 56 s.

WP 98:3 M. Tuček a kol.: Česká rodina v transformaci – Stratifikace, dělba rolí a hodnotové orientace; 162 s.

Pre-election Polls, Election Results, and Validity of Measurement before the 2002 Elections

Martin Kreidl & Tomáš Lebeda

Edice Sociologické texty/Sociological Papers SP 03:8
Řídí: Marie Čermáková
Překlady: Robin Cassling, Daniel Meier
Redakce: Anna Čermáková
Sazba: WOW Praha
Vydal: Sociologický ústav AV ČR
Jilská 1, 110 00 Praha 1
Náklad 170 ks
1. vydání

Prodej zajišťuje Tiskové a ediční oddělení
Sociologický ústav AV ČR
tel.: 222 221 761, fax: 221 183 250
e-mail: prodej@soc.cas.cz