The Evolution of Tax Evasion in the Czech Republic: A Markov Chain Analysis

December, 2004

by Jan Hanousek, CERGE-EI, and Filip Palda, ENAP*

Abstract

This study departs from standard analyses of tax evasion, which ask why people evade. We seek instead to measure the average individual's transitions in both directions between evading and not evading and use this to predict the evolution of tax evasion for the Czech Republic with the help of surveys taken in 2000, 2002, and 2004. We asked each respondent whether he evaded taxes in 1995, 1999, 2000, 2002 and 2004. Answers to the questions allowed us to calculate probabilities that the average individual will move between being a non-evader and being an evader. This "transition" probability leads us to predict a rising tide of tax evasion in the next decade. We estimate the structural and reduced form parameters, which determine evasion and suggest how government might influence these parameters to prevent the Czech Republic from bogging down in a permanent mire of tax evasion.
(JEL Codes : H26, H43, K42, O17)

1. Motivation

Most research into tax evasion has focused either on measuring the size of the sector that evades taxes, or in explaining why people evade taxes. Put differently, most studies of evasion are concerned with structural equations that predict the partial equilibrium response of an individual to a change in preferences or incentives. Few have sought to model how tax evasion evolves. The man who today does not evade taxes may next year decide to deduct from taxable income the car he uses to take his children to school. Emboldened by his first dodge he may in later years graduate to the corps of hardened evaders who do not declare their incomes and so do not need to fuss with receipts for false deductions. Another hardened

* Jan Hanousek is Professor at CERGE-EI, a joint workplace of Charles University and the Academy of Sciences of the Czech Republic. Filip Palda is Professor at Ecole Nationale d'Administration Publique, Montreal, Canada and visiting professor at CERGE-EI. Address for correspondence: CERGE-EI, P.O.Box 882, Politickych veznu 7, 111 21, Prague, Czech Republic. Emails: Jan.Hanousek@cerge-ei.cz, Filip_Palda@enap.uquebec.ca.
We thank Randy Filer and Stepan Jurajda for their comments on an earlier draft of this paper. This research was supported by by Grant Agency of the Czech Republic (402/04/0166), the Phare ACE Programme of the European Union (P97-8119-R) and the Volkswagen Foundation (II/75831).

evader may decide it is time to quit his stealthy ways and return to the society of taxpaying citizens. Each year government revenues will vary as millions of individuals slip in and out of the shadows of evasion. If we could know how likely an individual is to slip between tax-paying and tax-evading status we could draw a line to the future that traces the size and composition of the community of tax evaders. To know this likelihood we need know nothing about why he evades. We can rely on simple time series of his evasive behavior and use these time-series to predict how he will behave in years to come.

To model and track transitions an individual makes between evading and paying taxes we use non-panel surveys we conducted in 2000, 2002, and 2004. We asked respondents whether they evaded never, sometimes, or frequently and compacted this information to classify individuals as evaders or non-evaders. We asked 2000 respondents these questions for 1995, 1999, and 2000; 2002 respondents these questions for 1997, 2000, 2002; and 2004 respondents these questions for 1999, 2002, and 2004. By asking these questions for different years we could form an idea of how people drift between evading and not evading. Our estimate of the drift allows us to surmise how people might move between the categories of evading and not evading over the next five years. Our results suggest that for the moment evasion in the Czech Republic has stabilized and that the next five years will see a small decrease in evasion.

The main focus of our paper is on the methodology and meaning of estimating transition probabilities between the state of evading and the state of not evading. We make the Markov assumption that transition probabilities are independent of the path individuals follow in their evasion. We go to some length to justify this assumption and provide empirical support for the stability of transition probabilities by means of a Chi-squared test. Given the novel nature of our analysis we believe close attention must be given to methodology. The payoff from a sound methodology will yield a method that gives a rough idea of how tax evasion will evolve in a stable policy and demographic climate.

Our methodology is novel in several respects. First, we use Chi-square tests of significance to see whether each of our three surveys is consistent with each other. Consistency means that answers to certain questions of importance given retrospectively in later surveys cannot be distinguished statistically from answers given to the same questions contemporaneously in previous surveys. For example, we asked respondents in 2004 about their evasion in 2002. This is a retrospective question. The retrospective answers respondents of the 2004 survey gave were not statistically distinguishable from answers respondents in the 2002 survey gave about their evasion in 2002 (contemporaneous answers). We take such a result to mean that even though we are not using panel data, the memory of respondents is good enough that random samples taken in different years produce groups of people who can be merged and analyzed as a whole. The ability to merge surveys is important because we want to analyze the determinants of evasion using regression methods, and merging provides us with a sufficient number of observations to seek out statistically meaningful relations between evasion and the variables that determine evasion.

The second novel aspect of our study is that we test for the stability of our Markov probabilities. *Ceteris paribus*, Markov probabilities should not change. Yet few things in a transition economy are, of course, *ceteris paribus*. Demographics, institutions, and government policy towards evasion are in rapid change. We find some change in our Markov probabilities and attribute this to the aforementioned changes in the parameters that underlie tax evasion. We then go on to estimate the structural and reduced form determinants of our Markov transition probabilities so that, given certain assumptions about the evolution of these determinants, we can form a picture of how tax evasion will evolve. Again we emphasize the novel nature of our tests for changes in transition probabilities and we believe we have laid the groundwork for forecasting the evolution of the underground economy; a hitherto poorly charted field.

The plan of the present paper is first to discuss the dataset on which we base our analysis, and then to see how Markov transition probabilities can be estimated from our dataset. Using our data we forecast how tax evasion will evolve if government does nothing to change the parameters which influence an individual's choice to evade. Finally, we estimate the parameters of the individual's choice to evade and use these to perform simulations that show how government might stem the rising tide of evasion that our data predict. Throughout we caution that our efforts at prediction are subject to many critiques and that the main objective of this paper is to lay the groundwork for prediction.

2. Data on the Czech Republic

Our data come from three surveys of Czechs we carried out in 2000, 2002, and 2004. Summary statistics of all variables used in the survey as well as the survey questions are available in Hanousek and Palda (2002).[1] Our surveys are similar to that of Fortin et al. (2000). The technique they used was to conduct interviews (in our case face-to-face interviews) to gather information about how much tax people evade and why they evade. The Fortin et al. survey differed from ours in that it did not ask questions that would allow a researcher to infer the dynamics of tax evasion. Fortin et al. were interested in the link between buying goods and services on which taxes were not declared, and evasion. We present a detailed analysis of our dataset elsewhere (Hanousek and Palda 2002), but the main features of tax evasion to note are that it is primarily a function of the life cycle (rising until

[1] A detailed description of the surveys including questionnaires, summary tables and results explicitly mentioned in the text are available from the authors upon request or at http://home.cerge-ei.cz/hanousek/evasion

late middle age, then falling), is primarily a male activity, and is highly associated with part-time work and unemployment. Tables A1a-c in the appendix give overall descriptive statistics. These tables include both reduced form (age, sex) variables and structural variables (perceived quality of government services, perceived probability of being caught evading) that we shall discuss in greater depth in the final part of our paper where these variables take their place in reduced-form and structural-form regressions of the determinants of moving between categories of evasion.

Trends in evasion

The main questions of interest for the present paper were those that asked people how often they evaded taxes. We gave them the option of answering never, sometimes, or often. We found that the incomes of those who answered they sometimes and never evaded were statistically indistinguishable from each other, as were most of their other demographic features; in addition there was small difference for those two categories in evaded amounts for respondents who were willing to disclose it. With little to distinguish these two groups of evaders we chose to merge them into one category that we call "evaders." We have also decided to do away with the distinction between frequent and sometime evaders because we have found it puts excessive demands on the data, with dubious gains in distinguishing why people shift between the subjective categories of frequent and sometime evasion.[2] Table 1, parts A, B and C are derived from our 2000, 2002, and 2004 surveys, respectively and show the frequency with which people answered they evaded taxes for varying years. We calculated 95% confidence intervals for each category of evasion, (the details of which may be had in the technical appendix).

---

[2] An earlier version of this paper which analyzed all three categories of evasion is available at http://home.cerge-ei.cz/hanousek/evasion

Table 1. Values and 95% confidence intervals for relative frequencies of tax evasion.

A. Results based on 2000 survey: Years 1995, 1999, 2000

Year	Evade	Never evade
1995	15.8% (13.5%, 18.1%)	84.2% (81.9%, 86.5%)
1999	20.3% (17.8%, 22.9%)	79.7% (77.1%, 82.2%)
2000	25.1% (22.4%, 27.9%)	74.9% (72.1%, 77.6%)

B. Results based on 2002 survey: Years 1997, 2000, 2002

Year	Evade	Never evade
1997	23.1 (20.4%, 25.8%)	76.9% (74.2%, 79.6%)
2000	25.9% (23.1%, 28.7%)	74.1% (71.3%, 76.9%)
2002	23.9% (21.2%, 26.6%)	76.1% (73.4%, 78.9%)

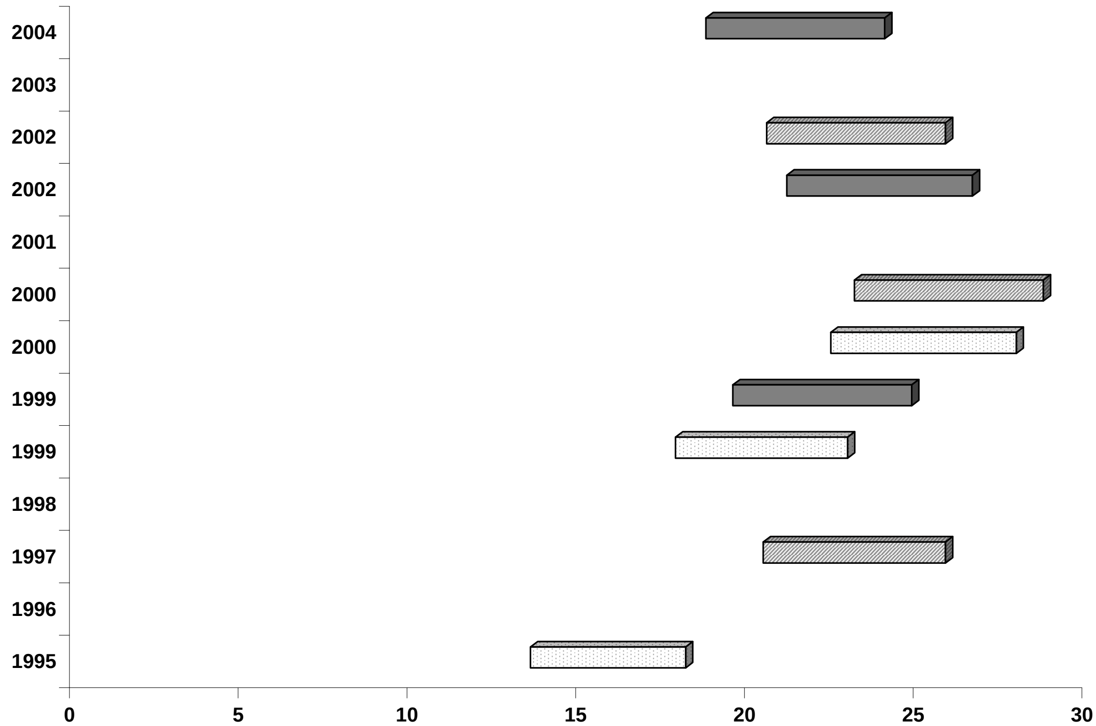
C. Results based on 2004 survey: Years 1999, 2002, 2004

Year	Evade	Never evade
1999	22.2% (19.5%, 24.8%)	77.8% (75.2%, 70.5%)
2002	23.2% (20.5%, 25.8%)	76.8% (74.2%, 79.5%)
2004	21.4% (18.7%, 24.0%)	78.7% (76.0%, 81.3%)

Source: Authors' computation using 2000-2004 surveys.
Confidence interval formulas are given in the appendix.

What can we say about trends in evasion?

Figure 1. Trends in estimated confidence intervals for percentage of tax evaders.



Note: The patterns in bars denoting estimated confidence intervals of tax evasion correspond to surveys done in 2000 (lightest), 2002 (medium), 2004(darkest).

In Figure 1 we have merged all three surveys to provide a time series of the above point and interval estimates of evasion. The graphic illustration above shows an upward tendency in the middle 1990s in the number of those who say they evaded; after that evasion seems to level off. We have conducted Chi-square tests indicating that between 1995 and 1997 tax evasion, as measured by the number of evaders, was rising but that after this evasion fell and leveled off until 2004.

Consistency of surveys

One of the objectives of this paper is to merge our three surveys in order to conduct regressions that tell us something about the determinants of the probabilities of moving between different categories of tax evasion. But as our surveys are not panel data but rather independent surveys taken at two year intervals, the question arises whether such a merging is legitimate. One of the main variables of interest in our surveys is the individual's answer to whether or not he evaded taxes. In each survey we asked people about their current and past evasion. If we could find that answers about evasion in 2002 given to questions in the 2004 survey are statistically indistinguishable from answers about evasion in 2002 given by respondents in the 2002 survey we might conclude that memory is good and that surveys in 2002 and 2004 are consistent with each other in the sense that we could merge the surveys to form a time series of data on tax evasion. The economy of merging in this way comes from the fact that in each survey we garner retrospective data on evasion and hence in each survey capture data for three periods. If memory is good, and our numbers suggest it is, then with confidence we can build a time series on evasion that for every survey produces three data points.

Tables 1a-c appear consistent in the sense described in the preceding paragraph, but a formal test is needed to confirm this guess. In general, consider two independent sample surveys of *n* and *m* observations respectively $\mathbf{x_1}=(x_{11}, x_{12}, \ldots, x_{1n})$ and $\mathbf{x_2}=(x_{21}, x_{22}, \ldots, x_{2m})$, where x_{ij} denotes the j^{th} observation of the i^{th} survey. Survey 1 is taken in the year 2000 and survey 2 is taken in the year 2002. The x's in the 2000 survey are the answers of each respondent to whether he evaded in 2000 and the x's in the 2002 survey are the answers to whether a respondent remembered evading in 2000. « Yes » answers are coded as ones, no answers as zeroes. The data are non-panel. Our variables of interest are the proportions of evaders in each sample $p_1 = \frac{1}{n}\sum_{i=1}^{n} x_{1i}$, and $p_2 = \frac{1}{m}\sum_{i=1}^{m} x_{2i}$ and we wish to test the hypothesis

$$H_0: p_1 = p_2,$$

i.e., that the proportion tax evaders in both samples is the same. Consider the following test statistic

$$u = \frac{p_1 - p_2}{\sqrt{\overline{p}(1-\overline{p})\left(\frac{1}{n}+\frac{1}{m}\right)}} \qquad (1)$$

where $\overline{p} = \frac{1}{(n+m)}\left(np_1 + mp_2\right)$, Under the null hypothesis, the test statistic *u* has a standard normal distribution. The above is a test statistic that allows us to distinguish whether certain variables have been drawn from different distributions. Our results are summarized in the following table.

Table 2. Tests of consistency of surveys: Comparison of retrospective estimates of evasion.

A. Tax evasion in 2000 (test of consistency 2000&2002)

Survey	Evaders	Non-evaders	Total
2000	268	794	1062
2002	268	766	1034
Test statistics	**-0.359**	p-value:	**0.360**

B. Tax evasion in 2002 (test of consistency 2002&2004)

Survey	Evaders	Non-evaders	Total
2002	247	788	1035
2004	245	813	1058
Test statistics	**0.382**	p-value:	**0.649**

C. Tax evasion in 1999 (test of consistency 2000&2004)

Survey	Evaders	Non-evaders	Total
2000	219	843	1062
2004	234	822	1056
Test statistics	**-0.863**	p-value:	**0.194**

In sum, these tables lead us to conclude that:

1) No difference can be found for the 2000 survey estimate of evasion in 2000 and the 2002 survey retrospective estimate of evasion in 2000 because the U-statistic of U = -0.359 is not significant.
2) The same can be said of the 2002 survey estimate of evasion in 2002 and the 2004 survey estimate of evasion in 2002 (U = 0.382, not significant).
3) The same can be said of the 2002 survey estimate of evasion in 1999 and the 2000 survey and its estimate of evasion in 1999 (U = -0.863, not significant).

By showing a strong consistency between surveys we have not only given some justification for merging surveys, but have also uncovered the result that answers to questions about past evasion in a survey taken in one year are statistically indistinguishable from answers to questions about contemporary evasion given in a survey two years earlier. Even though the two surveys are independently drawn, we are tempted to say that *people remember*. We are not quite sure where this result fits in the context of the present analysis, but we find it an important footnote to our research that might be worth pursuing on its own. Questions will, of course, remain about how well respondents can recall their tax evasion five years earlier. Our only check on the memories of respondents lies in the consistency of answers about tax evasion in our 2000, 2002, and 2004 surveys. A sharp-eyed critic might mention that if one assumes that the upward trend in evasion evident from the 2000 survey was monotonic then the evasion rate for 1997 drawn from the 2002 survey should be somewhere between its 1995 and 1999 levels, which according to the 2000 survey means that it should be between 15.8% and 20.3% but that according to the 2002 survey evasion in 1997 was on average 23.1%. This seems the only blip on our surveys that would suggest some gross lack of memory on the part of respondents. Our own reasoning for this blip is that retrospective estimates of evasion may be more honest than current estimates. A respondent cannot be punished for evasion in the distant past. His incentive to lie about his past evasion is lesser than his incentive to lie about current evasion. We should then expect the retrospective 2002 answers about evasion in 1997 to be highly relative to what respondents of the 2000 survey were claiming for their not-so-distant-past. This is perhaps why the point estimates of evasion in 1997 using the 2002 survey exceed those one might interpolate between 1995 and 1999 from the 2000 survey.

3. How Tax Evasion Evolves

To see how tax evasion will evolve in the Czech Republic we focus on the probability of changing between states of evasion. A proportion of the new labor force arriving on the market will not evade and others will jump to evading. Those who are not new to the labor force and do not evade can make similar jumps. Those who evade may jump to not evading. These flows in and out of tax evasion can be summarized by the following 2x2 "stage/transition matrix" for each individual for the 2000 survey:

Table 3. Markov transition matrix between years 1995 and 2000.

<table>
<tr><td colspan="2" rowspan="2">Tax evasion</td><td colspan="2">2000</td></tr>
<tr><td>Evade</td><td>Never evade</td></tr>
<tr><td rowspan="2">1995</td><td>Evade</td><td>P_{ee}</td><td>P_{en}</td></tr>
<tr><td>Never evade</td><td>P_{ne}</td><td>P_{nn}</td></tr>
</table>

Each cell gives for an individual the probability he will go from one state in 1995 to another state in 2000. For example, P_{en} gives the probability an individual who evaded in 1995 will never evade in 2000.

We assume that the transitions from evader to non-evader (and vice-versa) satisfy the Markov property; that is, the best forecast of future transitions depends on the current behavior of the taxpayer. Stated differently, the Markov assumption says that the path by which one has arrived at one's current state has no influence on the probability of acceding to a future state. Evading in three previous states then has no bearing on evading in a subsequent state. To those who believe in learning-by-doing our Markov assumption will seem objectionable, but we do not model current transition probabilities as depending on past behavior in small part for practical reasons. To forecast with path dependency means that one must know the contributions of the path and the contributions of individual characteristics that determine jumps from one state to another. We do not have a long-enough time series to isolate both effects nor do we know how this omission in our modeling might bias or not bias our forecasts. Given the novelty of the present research we wish to lay bare what will need to be done in future research and also to state that if the Markov assumption seems a strong one, we have not made this assumption gratuitously.

Each individual's transition probability will differ from the transition probability of other individuals. To precisely estimate how total evasion will evolve we would need to calculate a stage-transition matrix for each individual and then see what "percentage" of that individual (singular) moves from cell to cell. We would then add all these percentages in each year to arrive at the total number of evaders in each of the two categories. A simpler, though slightly less precise way of arriving at the same calculation is to calculate the aggregate transition probabilities. This is easily done by calculating the percentage of people who moved from cell to cell between 1995 and 2000. The aggregate probabilities are slightly less accurate than if we used a transition matrix for each individual, but given the large numbers we surveyed, the central limit theorem suggests that the variance of our calculations around the true mean (provided that individual transition probabilities are uncorrelated with each other) will not be far off their true values.

Our technique for predicting the evolution of tax evasion is new and needs to be considered in the context of past research on evasion. The work of Allingham and Sandmo (1972), Watson (1985), Jung et al. (1994), Yaniv (1994), and others hold that tax evasion is seen as a risky decision. Agents weigh the risk of detection against the gains from evasion. These models are mainly concerned with optimal audit and detection policy as in the literature on the economics of crime and do not model tax evasion over extended periods. Engle and Hines (1999) have built on these previous models to simulate and test a model of long-term evasion dynamics in the US using aggregate data. Those outside the US Internal Revenue Service do not know the basis upon which that service decides to audit taxpayers, but surmise that a taxpayer's probability of being audited is an increasing function of his current evasion. Engle and Hines (1999) build this conjecture into their model in which a taxpayer's current evasion is a decreasing function of prior evasion, "since if audited and caught for evading this year, the taxpayer may incur penalties for past evasions." Aggregate evasion shows cycles if a sufficiently large number of individual taxpayers cycle together, as happens under the influence of aggregate shocks which tend to influence all in the same direction. In the absence of such shocks Engle and Hines find the interesting result that the cross-section of evasion rates converges to a steady state and aggregate tax evasion approaches a limit even though individual rates cycle. The distinction between aggregate and individual cycles arises because an individual's steady state is conditional on not being audited, while the economy's steady state is conditional on a distribution of individual audits across taxpayers with differing evasion histories. The distinction between aggregate and individual cycles in tax evasion is similar to the distinction between family and societal sex ratios.

We can use Engle and Hines' (1999) insight that tax evasion converges to a steady state to draw conclusions about the evolution of tax evasion in the Czech Republic. Engle and Hines used their model to examine continuous aggregate data on tax evasion. Our data is on individuals, is discrete, and spans five years during which we can see how the respondent jumped between the categories of evading and not evading. As far as we know, such a dataset is unique.

To understand the salient features of our dataset and how it allows us to predict future evasion, consider that we are able to estimate both short-term and long-term Markov probabilities of moving between evasive and non-evasive states. A short-term probability can be understood from the following matrix

Table 4. Short-term transition matrix for 1999 and 2000.

1999/2000		**2000**		
		Evaders	**Non-evaders**	**Total**
1999	**Evaders**	**216**	**0**	**216**
		100.0%	0	
		20.3%	0.0%	20.3%
	Non-evaders	**51**	**795**	**846**
		6.0%	94.0%	
		4.8%	74.9%	79.7%
	Total	**267**	**795**	**1062**
		25.1%	74.9%	

Note: remaining short-term transition matrices are available at http://home.cerge-ei.cz/hanousek/evasion, or upon request.

To read this table consider the cell in the first column, which gives a value of 100%. This cell indicates that of those who said they evaded in 1999, 100% said they were evading in 2000, and that this percentage represents 216 observations in our sample of those who evaded in 1999. These 216 observations represent 20.3% of our sample. This is what we call a short-term transition matrix because it calculates Markov transition probabilities by looking at jumps between the first anterior period for which a question was posed and the current period. A long-term transition matrix would look at the jump between the most distant period about which evasion questions were posed and the current period. The following table shows such long-term transition probabilities for our 2000 survey for jumps between 1995 and 2000.

Table 5. Long-term transition matrix for 1995 and 2000.

1995/2000		**2000**		
		Evaders	**Non-evaders**	**Total**
1995	**Evaders**	**168**	**0**	**168**
		100.0%	0.0%	
		15.8%	0.0%	15.8%
	Non-evaders	**100**	**795**	**895**
		11.2%	88.8%	
		9.4%	74.8%	84.2%
	Total	**268**	**795**	**1063**
		25.2%	74.8%	

Note: remaining long-term transition matrices are available at http://home.cerge-ei.cz/hanousek/evasion, or upon request.

Tables A2a-A2c and A3a-A3c in the appendix show six Markov transition matrices: two for each survey of which one is a short-term transition matrix and the other is a long-term transition matrix. We can use these transition matrices to predict evasion out to any indefinite period; of course, the accuracy of predictions falls with increasing time horizon. Table 6 shows our projections using long-term transition probabilities.

Table 6. Predictions using fixed Markov (long-term) transition matrices.

Year	Evaders			Non-evaders		
	2000	**2002**	**2004**	**2000**	**2002**	**2004**
2000	**25.1%**			**74.9%**		
2001	*26.9%*			*73.2%*		
2002	*28.6%*	**25.9%**		*71.5%*	**74.1%**	
2003	*30.2%*	*26.3%*		*69.8%*	*73.7%*	
2004	*31.9%*	*26.7%*	**21.4%**	*68.1%*	*73.3%*	**78.6%**
2005	**33.6%**	*27.1%*	*21.5%*	**66.4%**	*72.9%*	*78.5%*
2006	*35.1%*	*27.5%*	*21.7%*	*64.9%*	*72.5%*	*78.3%*
2007	*36.5%*	**27.9%**	*21.8%*	*63.5%*	**72.1%**	*78.2%*
2008	*38.0%*	*28.2%*	*22.0%*	*62.0%*	*71.8%*	*78.0%*
2009	*39.5%*	*28.5%*	**22.2%**	*60.5%*	*71.5%*	**77.8%**
2010	**41.0%**	*28.8%*	*22.3%*	**59.0%**	*71.2%*	*77.7%*

Note: Bold-face numbers are values estimated from the survey; the estimated probabilities were based on a five-year transition matrix. Numbers in italics correspond to linear interpolation between bold face values.

The first three columns under the heading Evaders shows projections using the 2000, 2002, and 2004 surveys of the evolution of the number of tax evaders in the Czech Republic out to 2010.[3] Projections using the 2000 survey, which uses retrospective data back to 1995, show a near doubling of the number of evaders while the 2002 and 2004 surveys show either a decrease or a leveling off of evaders. The difference can be attributed to the Czech government's 2002 crackdown on evasion. Even though this crackdown changes the transition probabilities for 2002 and 2004 only slightly from what they were predicted to be using the 2000 survey, such a small change leads to large changes in future evasion. Such a result might be best posed to a politician as being similar to what happens to money when interest compounds upon it: small changes now have large consequences later.

[3] Let us denote $\Pr\{e(t)\}$ and $\Pr\{n(t)\}$ probability of tax evasion and not-evading taxes in time *t*, respectively. One can easily use transition probabilities $\Pr(e(t), e(t+5))$ and $\Pr(n(t), e(t+5))$ to estimate "future" probabilities of tax evasion using the following formula.

$$\Pr\{e(t+5)\} = \Pr\{e(t)\}.\Pr\{e(t), e(t+5)\} + \Pr\{n(t)\}.\Pr\{n(t), e(t+5)\}$$

Obviously, if we use a long-term transition matrix we obtain predicted probabilities of tax evasion on a five- year grid. The level of tax evasion in between those periods could be constructed via linear interpolation.

We find that long-term and short-term transition probabilities make projections which go in the same direction, but predictions presented in Table 6 are less radical than those using the short-term stage-transition matrix. That the results using the long-term stage-transition matrix are less pronounced than those using the short-term matrix may be due to chance. Errors may cumulate or cancel each other over the long term. Nevertheless, short and long term transition matrices set the bounds on what will be actual tendencies.

Stability of Markov Probabilities

In calculations not shown here we found that whatever survey we used, the percent of evaders and non-evaders each converge to some constant percentage. Such convergence is similar to Engle and Hines' (1999) discovery of a steady state in evasion. However, our work goes a step further than theirs in that we use transition probabilities estimated from individual data to make our projections. We must of course be sensitive to the criticism that our convergence to a steady state is an artifact of the computation method we use. The essence of this computation lies in the assumption of stable transition probabilities. In a transition country it may be reasonable to argue that these transition probabilities will not be constant and that as a result evasion may cycle rather than stabilize. The matter cannot be resolved theoretically and appeal must be made to the facts.

We have no conception of whether transition probabilities are stable or unstable. We would be surprised to find these probabilities to be stable given the changing policy climate and demographic evolution of the Czech Republic over the period which we study. Stable transition probabilities would be desirable for forecasting the evolution of tax evasion, but are not essential if we can estimate the determinants of transition probabilities and forecast how these probabilities will change based on different scenarios for changes in the variable that determine them.

Before jumping to the conclusion that some structural estimates of the determinants of transition probabilities are needed to forecast these probabilities we want to first check for constancy of the transition probabilities. To check for the constancy of these probabilities we can resort to a Chi-square test which asks whether the probabilities calculated from each of our three surveys can be said to have been drawn from the same distribution of variables that underlie the calculation of our probabilities (general setup of the test is described in the appendix).

For the test of stability (homogeneity) of the Markov transition matrices, both short and long-term, let us consider the four following categories: [E->E] (someone who evaded in an earlier period continues to evade in the current period), [E->N] (someone who evaded in an earlier period no longer evades in the current period), [N->E] (someone who did not evade in an earlier period begins to evade in the current period) and [N->N] (someone who did not evade in an earlier period continues to not evade in the current period). The numbers *n* transiting from one state to another are the basis of the calculation of our transition probabilities. A non-parametric test that the underlying transition matrices are the same across all surveys can be carried out by using a standard test of homogeneity of distributions. Let us summarize all outcomes in the following contingency table:

Table 7. Contingency table for test of stability of long-term transition matrices.

Sample (survey)	Change in evasion categories				Total
	1 [E->E]	2 [E->N]	3 [N->E]	4 [N->N]	
1 (2000)	n_{11} *168*	n_{12} *0*	n_{13} *100*	n_{14} *795*	$n_{1.}$ *1063*
2 (2002)	n_{21} *194*	n_{22} *44*	n_{23} *73*	n_{24} *720*	$n_{2.}$ *1031*
3 (2004)	n_{31} *148*	n_{32} *78*	n_{33} *86*	n_{34} *741*	$n_{3.}$ *1053*
Total	$n_{.1}$ *510*	$n_{.2}$ *122*	$n_{.3}$ *259*	$n_{.4}$ *2256*	N *3147*

Under the null hypothesis, the test statistic is

$$\chi^2 = \sum_{i=1}^{3} \sum_{j}^{4} \frac{\left(n_{ij} - \frac{1}{n} n_{i.} n_{.j}\right)^2}{\frac{1}{n} n_{i.} n_{.j}} = n. \sum_{i=1}^{3} \sum_{j}^{4} \frac{n_{ij}^{\ 2}}{n_{i.} n_{.j}} - n \qquad (2)$$

a chi-square distribution with 6 degrees of freedom.

The test statistic of 88.54 indicates that jointly the three surveys cannot be said to be drawn from the same distribution as far as transition probabilities are concerned.[4] The fact that transition probabilities should change from survey to survey might at first seem to cast a shadow of doubt over our endeavor to build a framework for forecasting tax evasion. Yet we tend to see this result in a cheery light. In a transition economy we would be disappointed not to find changing transition probabilities.[5] The trick for a forecaster is to find out what factors influence transition probabilities, to guess at how these factors will change, and then to use this guess within the Markov prediction framework to make forecasts. Our allies in this pursuit are regressions that will allow us to identify the force with which causal variables

[4] Similarly, test statistics for stability of short-term transition matrices of 149.4 shows that short-term transition probabilities could not be considered stable, either.

impinge upon transition probabilities. These forces are regression coefficients. Once we have these coefficients in hand we can then simulate how different values of our independent variables will influence our Markov probabilities. We can then shift our estimates of these probabilities in the future, depending on how causal variables change.

4. How to Moderate Tax Evasion

The challenge to providing credible estimates of the evolution of tax evasion lies in dealing with changing demographics and policy. As a population grows rich it will change its evasion practices. As government cracks down on evasion or changes the quality of services it provides to its citizens, people will make new decisions about whether to evade or not evade. Such changes are the woof and warp of the Lucas critique and the bane of forecasters. When the parameters that underlie the decision to evade change in aggregate, so must the aggregate Markov probabilities we have calculated. We showed earlier that we cannot believe these probabilities to be stable for the Czech Republic, though unstable parameters of evasion throw into doubt the accuracy of our forecasts of evasion. The best we can do to restore belief in our forecasts is to modify them by guessing how the parameters of evasion will change and using these guesses to modify our Markov transition probabilities. Put technically, we wish to use regression to estimate the impacts of the determinants of the *transition* between evasion and non-evasion on evasion and non-evasion. With these estimates in hand we can say that if demographics or policy take a certain path, Markov transition probabilities will also take a certain path. With the path of Markov transition probabilities in hand we can modify our forecasts to span over a changing future.

[5] Let us note that if we do similar tests of stability for variables which could be linked to change in evasion status like household income (compared to one year and five years ago, respectively), satisfaction with government services, etc. we also reject the stability of these variables.

It is all well and good to say we wish to estimate the importance certain variables have on transition probabilities, but what sorts of variables should we be looking at? The question strikes at the heart of deficiencies in current approaches to tax evasion. Forecasters like to look at reduced-form estimates of the coefficients attached to variables that do not enter into simultaneous relations with the dependent variable. We can estimate reduced-form regressions of the determinants of Markov transition probabilities, but such estimates will be mute on what we believe are important policy variables. We cannot include policy variables in our transition probability regressions because we have no objective measures of policy change that would not require a long time series and call on event-study methods. At best we can ask people what they perceive government policy to be, but perceptions are slippery quantities to include in reduced form regressions because we do not know if people state their perceptions to justify their evasive behavior. Ask me if I evade and I say yes. Then ask me if it is moral to evade and I may say yes to make me look respectable in the eyes of the interviewer (our surveys were face-to-face). My answers will foil the researcher running reduced form regressions and force him to estimate a recursive or simultaneous model of evasion. There is as yet no such generally accepted full-equilibrium model of evasion.

In an earlier study Hanousek and Palda (2004) developed a trick for partially bypassing the need for elaborate modeling of structural parameters while still including variables in their reduced-form regression such as perceptions of government policy. They ran a regression of the evasive behavior of individuals on each individual's perception of the quality of government services. Their notion, drawn from Downsian voting theory, is that people evade not just for instrumental reasons (putting more money in their pockets) but also for moral reasons ("if I do not get good quality government services I will protest by withholding my taxes"). To ensure that people would not justify their evasive behavior by answering that they believed they were getting poor government services, interviewers told subjects that the survey was about the quality of (government) services. Interviewers posed questions of quality at the start of the survey. Much later in the survey came questions about whether the respondent evaded taxes. We believe the order in which the two questions were posed reduced spurious correlation between answers to the two questions; the reverse order of questions gives respondents opportunity to "justify" by claiming that they evaded taxes because they believed government services to be of low quality (For more discussion and results related to this particular phenomenon see Hanousek and Palda, 2004).

Other policy variables that both theory and empirical literature suggest are important are the perceived probability of being caught evading and the perceived penalty for evasion. Clearly such variables belong in a structural regression. Our regressions should thus be thought of as quasi-reduced form regressions, integrating clearly exogenous variables such as demographics, and perceived policy variables over whose erogeneity some cloud of doubt may hang.

There are four stage-transition probabilities (E->N, E->E, N->E, N->N) and this suggests we estimate a reduced-form regression for each of the four possible transition probabilities. Once we estimate the parameters associated with the variables driving tax evasion, we can then simulate how Markov transition probabilities will change should the independent variables in the regressions change[6].

Table 8 shows the reduced form regression of one transition probability; that going from never evading to evading. There are many possible candidates for variables that might influence transition probabilities. We must choose only the most likely candidates for

[6] Readers will wonder how new entrants to the labor force figure in our calculations. Our data give us no way of knowing who is a new entrant. If we assume that entry and exit from the labor force bear a stable relation to each other and that entry and exit from the labor force is uniformly distributed over evasion categories, we need not consider explicitly the rates of entry and exit from the labor force in our calculations of how tax evasion will evolve. Some indirect evidence in support of this conjecture comes from our survey, which shows that those who evade often and those who evade occasionally have statistically indistinguishable average incomes.

inclusion in our equations because maximum likelihood is a technique whose appetite for data rises exponentially as we add parameters to be estimated. Demographic variables such as age and sex are standard proxies for a vector of individual characteristics. We also include a number of regional variables such as town size, and finally what the individual perceives to be the morality of evading and the probability of being caught, as well as whether his economic status is deteriorating. Table 8 should be viewed as one that seeks the factors determining evasion. Prominent among the determinants is the change in the economic status of the individual (going from good to bad increased the tendency to evade taxes), an individual’s experience of buying goods on which taxes have been evaded, and the perceived probability of being caught evading taxes.

What does Table 8 tell us about the stability of the transition probabilities we use to forecast the evolution of evasion? As the population ages, we can expect the transition probability of going from not evading to evading to fall. Space limitation preclude us from doing so, but it is a simple exercise to imagine different rates of increase in the number of elderly, plug them into Table 8 and add or subtract the change in the transition probability to that transition probability we used in earlier forecasts of the evolution of tax evasion. Table 8 also tells us that if the government can make people think their chances of being caught increase or that tax evasion is immoral then evasion will also fall. By themselves these findings are unremarkable, if respectable, additions to empirical work on tax evasion, but in the context of forecasting evasion these findings give us a precise way of modifying Markov transition probabilities to hone our forecasts of the evolution of evasion.

A further step in predicting changes in Markov transition probabilities would be to estimate multinomial logits, which treat all four transition probabilities as simultaneously determined. Table 9 presents these estimates for three of the four transition probabilities (we need not estimate the fourth regression because the three, less one, by definition give us the fourth equation).

Table 8. Logit regression results for P_{ne} (transition from never to a tax evasion stage) in the Czech Republic, Marginal effects for combined surveys

Variables	**Derivative dP/dX going from never to a tax evading category**	
	Long-term	Short-term
Demographics		
Age	-0.006*	-0.002**
Age squared	-4E-05	1E-05
Female	-0.035**	-0.013
Education		
Primary school education	0.060**	0.023
Apprenticeship (2 years)	0.032	0.005
Apprenticeship (3-4 years) w/t diploma	0.028	0.007
Secondary vocational w/t diploma	0.015	-0.002
Income		
< 10.000	0.014	0.032
10.001 to 15.000	0.039	0.034
15.001 to 20.000	0.023	0.035
20.001 to 25.000	0.064*	0.067
25.001 to 30.000	0.025	0.022**
Income relative to the past		
much worst compared to 5 years ago	0.071**	0.039
much better compared to 5 years ago	-0.036	-0.006
much better compared to a year ago	-0.001	0.004
Demographical dummies		
Big town	-0.021	0.003
Village	-0.009	-0.019
Prague	0.051**	-0.003
Middle Bohemia	0.004	0.004
Southern Bohemia	0.021	-0.008
Western Bohemia	0.047*	0.018
Northern Bohemia	0.024	0.007
Eastern Bohemia	0.041	0.032**
Southern Moravia	0.030**	0.012
Factors linked to tax evasion status		
bought goods from the underground economy	0.036**	0.005
Tax evasion is moral	0.006	-0.011
Tax evasion is very immoral	-0.028*	-0.017*
Probability of being caught	0.001**	-9E-05
Scaled R^2 (2859 observations)	0.08	0.03

Marks * and ** denote cases when underlying coefficients were significant on 5% and 1% significance level, respectively

Table 9. Results of multinomial logit regressions for all transition probabilities: Marginal effects (derivatives of transition probabilities w.r.t. the right hand side variables)

Variables	Long-term			Short-term		
	E=>E	**N=>E**	**N=>N**	**E=>E**	**N=>E**	**N=>N**
Demographics						
Age	.012	-.007**	-.010**	.006	-.002	-.005
Age squared	.000	.000**	.000**	.000	.000	.000
Female	-.045	-.033	.098*	-.066	-.012	.084
Education						
Apprenticeship (2 years)	.065	.002	-.104**	.075	-.007	-.090*
Apprenticeship (3-4 y) w/t diploma	.052	.001	-.074*	.056	-.004	-.071*
Secondary vocational w/t diploma	.002	-.012	-.006	.003	-.012	-.002
Income						
< 1.000	.037	.012	-.055	.015	.033	-.053
1.001 to 15.000	.071	.033	-.130	.071	.033	-.108
15.001 to 2.000	.079	.015	-.134	.062	.033	-.102
2.001 to 25.000	.114	.057	-.166	.101	.066	-.169
25.001 to 3.000	.139	.020	-.166	.137	.022	-.168
Income relative to the past						
much worst than 5 years ago	-.052	.071	-.038	-.019	.039	-.031
much better than 5 years ago	.055	-.029	-.032	.030	-.003	-.031
much better than a year ago	.010	.001	.009	.006	.003	-.012
Regional and Size Town dummies						
Big town	.017	-.020	.000	-.006	.003	.009
Village	-.006*	-.007	.049*	.004*	-.018	.045*
Prague	-.025	.051*	-.016	.028	-.003	-.028
Middle Bohemia	-.021	.006	.007	-.023	.005	.028
Southern Bohemia	.004	.025	-.040	.034	-.007	-.035
Western Bohemia	.024	.049	-.078	.052	.020	-.069
Northern Bohemia	-.057	.026	.027	-.041	.007	.039
Eastern Bohemia	.023	.041	-.074	.026	.032	-.056
Southern Moravia	.012	.032	-.024	.034**	.013**	-.010*
Factors linked to tax evasion status						
bought goods from the underground economy	.100	.028**	-.186**	.121**	.000**	-.185**
Tax evasion is moral	.066	.017	-.087	.086	-.005	-.084
Tax evasion is very immoral	-.149**	-.019	.162	-.149**	-.014	.171**
Probability of being caught	-.002**	-.001	.002**	-.002	.000	.003**
Scaled R^2 (2859 observations)	0.29			0.28		

Marks * and ** denote cases when underlying coefficients were significant on 5% and 1% significance level, respectively.

Many variables loose their individual significance in the multinomial logits but this method of estimating coefficients should give us forecasts of greater accuracy than if we simply estimated each transition probability logit separately from the other. We shall not dwell on the above estimates but produce them as an example of the “next step” in predicting changes in Markov transition probabilities.

5. Conclusion

In the present paper we have offered up a new method for predicting how tax evasion may evolve. With the help of several surveys we asked individuals about how they moved between four possible states: evading to not evading, evading to evading, not evading to not evading, not evading to evading. The answers to these questions allowed us to predict how the number of evaders (if not the value of taxes evaded) would evolve for several years out. Sensitive to the critique that the transition probabilities of evading might not be constant we presented a formal methodology for testing the constancy of evasion transition probabilities. In the worst-case scenario where these transition probabilities might not be constant we showed how multinomial logit regressions might be used in simulations. If the transition probability is a function of demographic and government policy variables then a multinomial logit can pinpoint the coefficients attached to the independent variables determining the transition probabilities. Forecasts can then me modified based on different projections of the evolution of variables that influences transition probabilities.

This paper is short on normative content and has focused on the methodology of forecasting the evolution of tax evasion using surveys.

References

Allingham, Michael.G. and Sandmo, Agnar.1972. "Income Tax Evasion: A Theoretical Perspective." *Journal of Public Economics,* 1, pp. 323-338.

Boadway, Robin; Marchand, Maurice and Pestieau, Pierre. 1994. "Toward a Theory of the Direct-Indirect Tax Mix." *Journal of Public Economics*, 55, pp. 71-88.

Cochran, William G. 1963. *Sampling techniques*. 2nd edition, New York: Willey.

Cremer, Helmuth and Gahvari, Firouz. 1996. "Tax Evasion and the Optimum General Income Tax." *Journal of Public Economics,* 60, pp. 235-249.

Deaton and Muellbauer

Engel, Eduardo M. R. A. and Hines, James R., Jr. 1999. "Understanding Tax Evasion Dynamics." Working Paper: 6903, National Bureau of Economic Research, January 1999.

Feige, Edgar L. and McGee, Robert T. 1983. "Sweden's Laffer Curve: Taxation and the Unobserved Economy." *Scandinavian Journal of Economics,* 85, pp. 499-519.

Fortin, Bernard; Lacroix, Guy and Montmarquette, Claude. 2000 "Are Underground Workers More Likely to be Underground Consumers?" *Economic Journal*, October 2000, 110 (466), pp. 838-60.

Hanousek, J. and Palda, F. 2003., "Why People Evade Taxes in the Czech and Slovak Republic: A Tale of Twins," in: Belev, B. (ed.) *The Informal Economy in the EU Accession Countries: Size, Scope, Trends and Challenges in the Process of EU Enlargement*, Center for the Study of Democracy, Sofia, Bulgaria, 139-174.

Hanousek, Jan and Palda, Filip. 2002. "Why People Evade Taxes in the Czech and Slovak Republics: A Tale of Twins." Discussion Paper, CERGE-EI, 2002.

Jung, Young H.; Snow, Arthur, and Trandel, Gregory. 1994. "Tax Evasion and the Size of the Underground Economy." *Journal of Public Economics*, 54, pp. 391-402.

Lucas, Robert, E. 1976, “Econometric Policy Evaluation: A Critique”, *Carnegie-Rochester Conference Series on Public Policy*, 1, 19-46

McFadden, Daniel. 1987, “Regression-Based Specification Tests for the Multinomial Logit Model”, *Journal of Econometrics*, 34(1/2): 63-82.

Reinganum, Jennifer E. and Wilde, Louis L. 1991. "Equilibrium Enforcement and Compliance in the Presence of Tax Practitioners." *Journal of Law, Economics, and Organization*, 7, pp. 163-168.

Watson, Harry. 1985. "Tax Evasion and Labor Markets." *Journal of Public Economics*, 27, pp. 231-246.

Yaniv, Gideon. 1994. "Tax Evasion and the Income Tax Rate: A Theoretical Perspective." *Public Finance,* 49, pp. 107-112.

Appendix:

Table A1.a. Descriptive statistics for 2000 survey, Czech Republic.

2000 survey	***Total sample***	***Purchase of informal goods/ services***		***Active engagement in informal activities***	
		Yes	*Never*	*Evaders*	*Non-evaders*
Total	**1062**	**507**	**555**	**268**	**794**
Sex					
Male	518	50.0%	50.0%	34.6%	65.4%
Female	544	45.6%	54.4%	16.4%	83.6%
Age					
18 to 25 years	188	46.8%	53.2%	30.3%	69.7%
26 to 35 years	247	52.6%	47.4%	25.9%	74.1%
36 to 45 years	229	52.8%	47.2%	31.4%	68.6%
46 to 55 years	248	50.8%	49.2%	22.2%	77.8%
56 to 65 years	150	28.0%	72.0%	13.3%	86.7%
Level of education					
Primary	252	47.6%	52.4%	29.0%	71.0%
Without GCE	427	51.5%	48.5%	32.6%	67.4%
With GCE	295	45.1%	54.9%	14.6%	85.4%
Higher	88	38.6%	61.4%	14.8%	85.2%
Level of income[CZK]					
< 10,000	566	45.1%	54.9%	22.1%	77.9%
10,001 to 15,000	266	54.9%	45.1%	33.1%	66.9%
15,001 to 20,000	91	54.9%	45.1%	23.1%	76.9%
20,001 to 25,000	22	59.1%	40.9%	50.0%	50.0%
25,001 to 30,000	9	55.6%	44.4%	55.6%	44.4%
30,001 to 40,000	1	100.0%	0.0%	100.0%	0.0%
Rejected	107	34.6%	65.4%	15.9%	84.1%

Source: 2000 survey, authors' computations

Table A1.b. Descriptive statistics for 2002 survey, Czech Republic.

2002 survey	*Total sample*	*Purchase of informal goods/ services*		*Active engagement in informal activities*	
		Yes	*Never*	*Evaders*	*Non-evaders*
Total	**1041**	**573**	**464**	**247**	**788**
Sex					
Male	513	58.9%	41.1%	29.8%	70.0%
Female	528	51.3%	47.9%	17.8%	81.3%
Age					
18 to 25 years	196	56.6%	42.9%	29.6%	69.4%
26 to 35 years	241	58.1%	41.5%	26.1%	73.4%
36 to 45 years	214	55.6%	44.4%	26.6%	72.4%
46 to 55 years	237	54.0%	45.1%	18.6%	81.4%
56 to 65 years	153	49.0%	51.0%	16.3%	83.0%
Level of education					
Primary	195	54.9%	44.6%	25.1%	74.4%
Without GCE	404	58.4%	41.3%	26.0%	73.5%
With GCE	339	53.1%	46.6%	23.3%	75.8%
Higher	103	48.5%	50.5%	13.6%	86.4%
Level of income[CZK]					
< 10,000	487	55.4%	43.9%	19.5%	79.9%
10,001 to 15,000	335	51.0%	49.0%	24.5%	74.9%
15.001 to 20,000	114	59.6%	40.4%	30.7%	68.4%
20,001 to 25.000	35	71.4%	28.6%	42.9%	57.1%
25,001 to 30,000	18	72.2%	27.8%	50.0%	50.0%
30,001 to 40,000	4	75.0%	25.0%	25.0%	75.0%
40,001 to 50,000	3	100.0%	0.0%	33.3%	66.7%
>= 50,001	1				
Rejected	44	45.5%	52.3%	20.5%	79.5%

Source: 2002 survey, authors' computations

Table A1.c. Descriptive statistics for 2004 survey, Czech Republic.

2004 survey	*Total sample*	*Purchase of informal goods/ services*		*Active engagement in informal activities*	
		Yes	*Never*	*Evaders*	*Non-evaders*
Total	**1066**	**568**	**494**	**227**	**836**
Sex					
Male	535	58.7%	40.7%	29.9%	69.7%
Female	531	47.8%	52.0%	12.6%	87.2%
Age					
18 to 25 years	183	58.5%	41.0%	25.1%	74.9%
26 to 35 years	270	55.6%	44.4%	23.3%	76.3%
36 to 45 years	237	51.9%	46.8%	16.5%	83.5%
46 to 55 years	269	50.6%	49.4%	23.0%	76.6%
56 to 65 years	181	47.0%	53.0%	18.2%	81.2%
Level of education					
Primary	159	59.7%	39.6%	28.3%	71.7%
Without GCE	445	60.0%	39.6%	24.5%	75.5%
With GCE	347	46.7%	54.5%	16.4%	83.0%
Higher	115	38.3%	61.7%	13.9%	85.2%
Level of income[CZK]					
< 10,000	448	54.2%	45.5%	17.6%	82.4%
10,001 to 15,000	331	54.7%	44.7%	23.9%	75.2%
15,001 to 20,000	139	48.2%	51.8%	23.7%	76.3%
20,001 to 25,000	41	58.5%	41.5%	39.0%	61.0%
25,001 to 30,000	28	64.3%	35.7%	28.6%	71.4%
30,001 to 40,000	2	0.0%	100.0%	0.0%	100.0%
40,001 to 50,000	1	0.0%	100.0%	0.0%	100.0%
>= 50,001	0				
Rejected	76	46.1%	52.6%	15.8%	84.2%

Source: 2004 survey, authors' computations

Table A2.a. Short-term transition matrix for 1999/2000 based on 2000 survey.

1999/2000		2000		
		Evaders	**Non-evaders**	**Total**
1999	**Evaders**	**216**	**0**	**216**
		100.0%	0	
		20.3%	0.0%	20.3%
	Non-evaders	**51**	**795**	**846**
		6.0%	94.0%	
		4.8%	74.9%	79.7%
	Total	**267**	**795**	
				1062
		25.1%	74.9%	

Table A2.b. Short-term transition matrix for 2000/2002 based on 2002 survey.

2000/2002		2002		
		Evaders	**Non-evaders**	**Total**
2000	**Evaders**	**218**	**50**	**268**
		81.3%	18.7%	
		21.1%	4.8%	25.9%
	Non-evaders	**29**	**737**	**766**
		3.8%	96.2%	
		2.8%	71.3%	74.1%
	Total	**247**	**787**	
				1034
		23.9%	76.1%	

Table A2.c. Short-term transition matrix for 2002/2004 based on 2002 survey.

2002/2004		2004		
		Evaders	**Non-evaders**	**Total**
2002	**Evaders**	**195**	**32**	**227**
		85.9%	14.1%	
		18.5%	3.0%	21.5%
	Non-evaders	**50**	**778**	**828**
		6.0%	94.0%	
		4.7%	73.7%	78.5%
	Total	**247**	**787**	
				1055
		23.4%	74.6%	

Table A3.a. Long-term transition matrix for 1995/2000 based on 2000 survey.

1995/2000		2000		
		Evaders	**Non-evaders**	**Total**
1995	**Evaders**	**168**	**0**	**168**
		100.0%	0.0%	
		15.8%	0.0%	15.8%
	Non-evaders	**100**	**795**	**895**
		11.2%	88.8%	
		9.4%	74.8%	84.2%
	Total	**268**	**795**	**1063**
		25.2%	74.8%	

Table A3.b. Long-term transition matrix for 1997/2002 based on 2002 survey.

1997/2002		2002		
		Evaders	**Non-evaders**	**Total**
1997	**Evaders**	**194**	**44**	**238**
		81.5%	18.5%	
		18.8%	4.3%	23.1%
	Non-evaders	**73**	**720**	**793**
		9.2%	90.8%	
		7.1%	69.8%	76.9%
	Total	**267**	**764**	**1031**
		25.9%	74.1%	

Table A3.c. Long-term transition matrix for 1999/2004 based on 2004 survey.

1999/2004		2004		
		Evaders	**Non-evaders**	**Total**
1999	**Evaders**	**148**	**78**	**226**
		65.5%	34.5%	
		14.1%	7.4%	21.5%
	Non-evaders	**86**	**741**	**827**
		10.4%	89.6%	
		8.2%	70.4%	78.5%
	Total	**267**	**764**	**1053**
		25.4%	72.6%	

Calculating confidence intervals for estimated probabilities of evasion

The sample relative frequencies *pe*, *p a*nd p_n allow us to construct confidence intervals for underlying probabilities P_e, and P_n and for a whole transition matrix Π_{ij} . Since we analyze a random sample from the Czech population, population size, *N*, will refer to several million, and therefore we can use the well-known normal approximation (see for example Cochran, 1963) to show that

$$\frac{p-P}{\sqrt{\frac{p(1-p)}{n-1}\frac{N-n}{N}}} \approx \frac{p-P}{\sqrt{\frac{p(1-p)}{n-1}}} \sim N(0,1)\,, \qquad \text{(A.1)}$$

where *p* and *P* can refer to p_e, and p_n and P_e, and P_n , respectively. Hence, a (1- α)% confidence interval is simply determined by (we use *1/2n* correction for non-continuous random variables)

$$p-\frac{1}{2n}-u_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n-1}} < P < p+\frac{1}{2n}+u_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n-1}}\,, \qquad \text{(A.2)}$$

where *u* denotes quintile of the standard normal distribution.

Test of homogeneity of sample distributions applied to Markov transition matrices

Let us assume we have $I \geq 2$ independent random samples (surveys) of size n_1, n_2, .. n_I. Assume that our variable of interest takes on values which fit into *J* disjoint categories. For example, if we have two categories of tax evasion (evade, not evade) then *J* would equal two. Let us denote n_{ij} the total number of variable values in *i*-th sample which fall into category *j*, where *i*=1,…,*I*, and *j*=1,..,*J*. For example, if we call the 2002 survey i=2 and we call the evasion category j=1 then n_{21} would be 247, the number of male and female evaders as gleaned from Table 1b in the appendix. To see whether our Markov transition probabilities are stable we want to test the hypothesis H_0: all samples *i*=1,..*I* are from the same distribution.

This situation is known as test of homogeneity of underling distributions i=1,…,I and it is typically summarized as in the following contingency table:

Sample	Category						Total
	1	2	…	J	…	J	
1	n_{11}	n_{12}	…	n_{1j}	…	N_{1J}	$n_{1.}$
2	n_{21}	n_{22}	…	n_{2j}	…	N_{2J}	$n_{2.}$
…	…	…	…	…	…	…	…
i	n_{i1}	n_{i2}	…	n_{ij}	…	n_{iJ}	$n_{i.}$
…	…	…	…	…	…	…	…
I	n_{I1}	n_{I2}	…	n_{Ii}	…	n_{IJ}	$n_{I.}$
Total	$n_{.1}$	$n_{.2}$	…	$n_{.i}$	…	$n_{.J}$	N

Under the null hypothesis, the test statistic

$$\chi^2 = \sum_{i=1}^{I} \sum_{j}^{J} \frac{\left(n_{ij} - \frac{1}{n} n_{i.} n_{.j}\right)^2}{\frac{1}{n} n_{i.} n_{.j}} = n.\sum_{i=1}^{I} \sum_{j}^{J} \frac{{n_{ij}}^2}{n_{i.} n_{.j}} - n \qquad \text{(A.3)}$$

has a chi-square distribution with (I-1)(J-1) degrees of freedom.