Financial 3/2016
Theory & Practice

Institute of
Public Finance

Financial 3/2016
Theory & Practice

Reviewed scientific journal

Institute of
Public Finance

hr
HR EXCELLENCE IN RESEARCH

Financial Theory and Practice publishes scientific articles in the field of public sector economics, and also welcomes submissions of applied and theoretical research papers on a broader set of economic topics such as economic growth and development, the role of institutions, transition to the market economy and EU integration. Empirical analysis is preferably related, but not limited, to the experience of countries in Central and Eastern Europe and Southeast Europe.

Submission of papers

Submission of a paper will be held to imply that it contains original unpublished work and is not being submitted for publication elsewhere. All papers have to pass through a double blind peer-reviewing process.

Guidelines for authors can be found at www.fintp.hr.

Financial Theory and Practice is published in accordance with the highest level of professional ethics and standards. *Ethical guidelines for journal publication* can be found at www.fintp.hr.

Financial 3/2016
Theory & Practice

TABLE OF CONTENTS

The evolution of law under communism and post-communism: a system-theory analysis in the spirit of Luhmann

BRUNO SCHÖNFELDER, dr. rer. pol.*

_____
Bruno SCHÖNFELDER
Technical University Bergakademie Freiberg, Faculty 6, Schloßplatz 1, 09599 Freiberg, Germany
e-mail: schoenfe@vwl.tu-freiberg.de

Abstract

The paper suggests that Luhmann's theory of social systems and evolution offers a powerful approach for analyzing law and the economy. It considers Yugoslavia and some Central European countries before and after 1990. Luhmann's concept of evolution stands in stark contrast to mainstream economic theory. It enables us to clarify the concept of transformation. A transformation became necessary because communism was an evolutionary dead end. According to this view, in post-communism the primacy of functional differentiation needs to be reestablished because it was partially reversed under communist rule. In these circumstances, the popular call for "sufficient" public control over the market is asking for the impossible. Post-communist law is bound to fall behind the evolution of markets. This causes economic problems and retards financial deepening.

Keywords: evolution, function systems, re-stabilization, operation closure, judge-made law

1 INTRODUCTION

When speaking about Central Europe and the Baltic states, most observers agree that the post-communist transformation[1] has been successful. The reverse is widely accepted for Russia and the other CIS countries. On the Balkan Peninsula matters are more mixed. Recent developments in Ukraine have reminded us that spelling out the ingredients of successful transformation continues to be more than just an academic concern. Some economists claim that an adequate theory of transformation has already been provided. However, their treatises[2] suffer from a major defect: they gloss over the obvious fact that in post-communism the political and legal systems were in need of fundamental changes and that this transformation was by no means less challenging than "the economic transition". In addition, they tend to play down the intricate interdependencies between these three agendas of transformation. The assertion that economic theory by itself can provide an adequate understanding of transformation is tantamount to the preposterous proposition that political science and sociology lack substance and may without further ado be replaced by economic analysis. The poor results of economists' attempts to comprehend post-communism thus reveal the folly of economic imperialism. Gambetta's quip "We know much and understand little"[3] aptly describes the failure.

This problem can be solved only by means of an overarching theory, i.e. a theory of society. Presumably, no such theory was available in 1990.[4] No doubt, economists were ill-equipped to fill the gap. As will be argued in this paper, their candidate for the job, which was the theory of economic systems and Hayek's theory of social order, was not up to it either. Matters changed in 1997, when Luhmann's magnum opus was published. However, few economists took note of the event.

---

[1] In the course of this paper it will become clear why we prefer this term over the more popular word transition.
[2] See, e.g. Roland (2000), Drahokoupil and Myant (2011).
[3] See Gambetta (1988:127).
[4] See the discussion of the crisis of sociological theory in Luhmann (2012, 2013).

Unfortunately, all of his key works were written in German. English translations became available only with great delay and even as of now some central pieces of Luhmann's titanic work have not yet been translated[5]. Since his texts are about as a readable as Kant's Critique of Pure Reason and Hegel's Phenomenology of Spirit taken together[6] this has greatly hampered diffusion.

The paper starts with a critical review of the traditional approach towards transition. Subsequently, Luhmannian concepts such as function systems, differentiation of systems, operation closure and evolutionary dead ends are introduced. Their impact is revolutionary and changes our whole way of thinking.

2 CONCEPTS OF EVOLUTION

2.1 THE TRADITIONAL VIEW

As this paper argues, the failure to understand post-communism is due to an underdeveloped theory of societal evolution. Throughout the history of economic thought, economists have occasionally referred to evolution but this has been little more than a biological metaphor and a makeshift device grasped at to avoid embarrassment. Usually this happened when mainstream theorists were confronted with a question for which they had no good answer. If economists are pressured hard to explain the meaning of evolution, they tend to quote Hayek[7]. Hayek reminded them of the legacy of the Scottish enlightenment, of David Hume, Adam Ferguson and Lord Kames. According to these writers, evolution is unpredictable and uncontrollable. It cannot be planned. It keeps surprising us. It is about the unintended consequences of human action. According to Ferguson, it creates institutions that are the "result of human action but not the execution of human design".[8]

Clearly, this is only a negative concept of evolution. It tells us what evolution is not. It does not even ask the question whether all social change is by necessity evolutionary and whether other modes of social change are pursuable as well. If all social change occurs through evolution and if these Scotsmen have captured its essence, bold political action is undesirable. If one knows so little about outcomes, wisdom suggests a conservative and risk-averse approach towards policymaking. Unfortunately, this policy-stance is unlikely to win democratic elections. Democracy tends towards the welfare state, which is tantamount to political activism.[9] As far as the post-communist transformation is concerned, the Scottish view, presumably, suggests gradualism. Wiles' (1992:392) advice against haste ("… things

---

[5] Luhmann (1990, 2000) to this date can be only read in German.

[6] This is not an approving statement!

[7] Or they refer to Nelson and Winter (1982) who however did not answer the questions posed in this paper. They confined themselves to an attempt to apply a particular concept of evolution to a free enterprise economy. For a refutation of this concept see Luhmann (2012:337).

[8] See oll.libertyfund.org/groups/104 with reference to Ferguson: An Essay on the History of Civil Society. London: Cadell 1767.

[9] This has been deplored by (neo)liberals who have been calling for a return to limited government. See, e.g. Epstein (1995). According to Luhmann (2000) this is unlikely to happen. He presents a theory of "systemic" hyperactivity.

done in haste are done badly") probably captures its essence. Indeed, Hayek himself used the word "gradual" in ways that suggest that evolution is more or less tantamount to gradual change.[10]

The experience of Central Europe disproves this view. In little more than a decade Central European countries went through changes that had taken centuries elsewhere. This was social change on the superfast track. If gradualists had been right all of this should have ended up in total disaster, but it did not. This false prognosis is not the only defect of the gradualist conception. It also fails to pay due respect to the degree to which the communist economic order depended on the coercive powers of government. When communist power was fading, this indispensable building block was irretrievably lost. The decline of communist power was revealed by two economic ills, inflation and foreign indebtedness. The rising tide of inflation, whether open or repressed, as well as the mounting foreign debt and increasing shortage of foreign exchange, indicated that the regime was no longer able to resolve distributional conflicts, and, as a result, lost control over the money supply and imports.[11] In the late eighties at least one of these two symptoms of crisis, if not both, could be found in all communist countries except Czechoslovakia[12] and China.

Presumably, the post-communist transition would have been a lot easier, if East European communists had undertaken decisive steps in the direction of the market at a time when they still wielded enough power to carry this project out. By the 1980s, they had wasted the chances history had offered to them. In retrospect, we know that their rule had become fragile and that most communist leaders were aware of this. Even if Gorbachev had wanted to implement bold market-oriented reforms, he could not have succeeded. The attempt would only have resulted in his fall, irrespective of the vast powers the formal organization of government bestowed upon him. Historians proved this convincingly.[13] Among the European communist countries, only Hungary and Yugoslavia undertook market-oriented reforms at a time when communist power was still solid. As a result, transition without shock may have been conceivable in these countries. In Yugoslavia, this opportunity was lost when inflation got out of control and the Yugoslav project headed towards bankruptcy, rendering political disintegration along ethnic lines unavoidable.[14]

---

[10] See, e.g. Hayek (1989:30ff).

[11] For a general discussion see Madžar (1990:358f). For a more specific view on the Soviet Union see Gajdar (2007), for the GDR Schürer (1998), for Bulgaria Christov (2007).

[12] See Turek (1995). For the reasons why Czechoslovakia was different see Možný (2009). Obviously, this made post-communist reforms easier.

[13] Durman (1998) and Gajdar (2007) offer impressive accounts.

[14] The peculiar window of opportunity that Yugoslav socialism offered presumably closed in the 1970s, when Tito sided with the conservatives against the liberal wing of the SKJ (League of Communists of Yugoslavia). Luhmann's theory suggests various speculations about the Yugoslav project. Presumably it got on a wrong start as early as 1919, when it was decided that the capital of the Kingdom should be Belgrade. If the intention had been to build a modern state rather than an empire, Zagreb would have been the better choice. One may wonder whether Italy would still exist, if Naples or Palermo had been chosen for the capital after Italian unification. In Yugoslavia, matters did not get any better after 1944. Socialist Yugoslavia was a contradiction

The history of market-oriented reforms undertaken under communist rule casts further doubts on the possibility of transition without shock. As a rule, market-oriented reforms were successful only if they were undertaken at a relatively early stage of communist rule and if they liberalized a sector of considerable importance in which no more than a partial return to pre-communist patterns was needed to unleash entrepreneurial spirits and generate a surge of output. The prime candidate was agriculture.[15] This happened e.g. in China, Hungary and Yugoslavia. The Chinese and Hungarian reforms of industrial management were predated by an agricultural reform that created prosperity because it offered entrepreneurial opportunities to farmers.[16] According to McKinnon (1992) the Chinese path towards market-oriented reforms was in the spirit of Hayek. He suggested that economists should think about Hayekian approaches for Eastern Europe. However, nobody ever managed to meet this challenge. This indicates that the Hayekian concept of evolution could not tackle the issue of post-communist transformation.

2.2 A NEW APPROACH

In this dilemma, Luhmann's concepts come in handy.[17] In the spirit of Luhmann, communism may be conceptualized as an evolutionary dead end. If, by mistake, you are running down a dead end road, turning around early saves time and trouble. The failure to do this in time explains the misfortunes of later attempts at reform. The abstract concept behind the metaphor of a dead end is "a normalization of the improbable that no longer suffices for further evolution"[18]. In the course of evolution certain operations, which hitherto have appeared far-fetched and unlikely, become perfectly normal and a matter of everyday experience. However, not everything can be normalized. Evolution keeps trying. Most, if not all, attempts fail. A second piece of evidence supporting the view that communism was a dead end is the amazing ease of its collapse. Have we ever seen an empire of this enormous size vanishing with so little ado? To be sure, the ideological conviction which had built the empire had long been fading, but such problems usually can be solved and the fading ideas substituted for by fresh ones. The medieval and early modern nobility of Western Europe managed to do this. It overhauled its

---

in terms. The socialist emphasis on redistribution was bound to alienate the wealthier regions of the country. At a deeper level of analysis it is worth pointing out, that, with few exceptions, "nations" came into existence as a side-effect of transition towards the primarily functionally differentiated society, i.e. the very opposite of socialism. See Luhmann (2013:283).

[15] For more on this see Gajdar (2012:468). His ultimate source of inspiration was an unpublished paper of Berliner quoted in Sachs and Woo (1994:121). Berliner wrote: "The Chinese transformation began with millions of peasants and others virtually beating at the gates of government to dismantle the restraints of the past and to let them work and thrive. When the gates were let down, they rushed in, and produced that remarkable surge of output. Soviet farmers, however, were not beating …"

[16] See Berend (1990:93f). In 1958, Hungarian communists started to revise their concept of socialist agriculture and created opportunities for private entrepreneurship. As a result, agriculture thrived. This was widely noticed. GDR-economists called it a miracle. The Hungarian party leadership took care to hide the reasons for this outcome. It did not want to be castigated for "Titoism". Polish agriculture was not fully socialized either. Unlike the aforementioned countries, this was not a prelude to market socialism, which was tried only in the 1980s and turned into an outright disaster for the regime. See Hardy (2009:23-26). For this reason Gajdar classifies Poland and the Soviet Union as prime examples of the dead end-proposition.

[17] A complete review of Luhmann's theory and its applications is beyond the scope of this paper. Moeller (2006) provides a user-friendly introduction.

[18] Luhmann (2013:29).

self-image and world-outlook several times without jeopardizing its supremacy in society.

Moreover, Luhmann proposes that the term evolution should be used only if variation and selection can be distinguished and have become random events for each other. This can occur only in complex systems. Such systems can generate large numbers of variations, most of which are discarded and forgotten without leaving a trace.[19] At some moment, unexpectedly, a variation is selected. This may result in great changes. As an example, think of the innumerable instances of dissent and opposition that occurred throughout the history of communist rule and were quickly crushed by the regime.[20] In 1989, all of a sudden, dissent caused momentous and unpredicted consequences. This conceptualization implies that evolution is not tantamount to a slow-moving process. The opposite may be true. Long times of stagnation may be followed by a catastrophic avalanche of change, by an accumulation of effects and after-effects. In Luhmann's words (2012:253): "Evolution is, as it were, a theory of waiting for useful chances."

For further analysis, we need to take a closer look at the system that is subject to evolution. At this point, the term post-communist *transformation* turns out to be superior to the term transition, because it captures the idea that post-communism is about a change of the primary *form* characterizing society, its primary mode of internal differentiation. Throughout the world, the nineteenth century witnessed the rise of functional differentiation and its establishment as the primary form of world society. This means that the economy, law, politics and some other areas became "operationally closed", autonomous function systems, each of which is specialized in fulfilling a particular function for society and employs its own self-generated structures and "memory" for this purpose. The idea of self-generated structures distinguishes this concept from older teachings about the division of labor. The prevalence of this form on a world scale, however, does not imply that all of its function systems operate properly in all territories of the world society. They do not! In most countries of the world, law works badly. Its autonomy is jeopardized and courts are subject to political interference if a case is considered politically important. The resulting underdevelopment of the legal system causes severe problems for both business and politics. By way of illustration, consider the division of powers and the institution of a democratically elected legislative assembly. This idea loses much of its appeal if the laws given by the legislature have little or no effect. This is the likely outcome if the administration of justice is less than "tolerable"[21].

---

[19] This definition implies that the technical progress generated by the research and development departments of companies is not evolution in the Luhmann sense. In contrast, the basic research done at universities (such as the research that resulted in this paper) is evolutionary. Proof: most of the research papers written by professors rapidly fade into oblivion and do not even get a single quote. Like biological evolution, social evolution tends to produce tremendous abundance.

[20] Havel's (2012) essay conveys the feeling of hopelessness that the apparent invincibility of the regime caused among dissenters.

[21] In the sense of Adam Smith, who considered "tolerable administration of justice" as a prerequisite for increasing the wealth of a nation beyond a certain level. See Irwin (2014).

The mutual relations between the three function systems for the economy, law and politics have often been described as harmony and support. However, in reality they frequently disturb and obstruct each other. As an example of sabotage, consider a political move towards the overregulation of important industries. This weakens their economic performance and backfires on politics because public revenue declines. In addition, corruption spreads because it enables businessmen to bypass burdensome regulations. In emerging and developing countries, this may result in a sort of vicious circle[22]. Presumably, this poses less of a danger in wealthy countries. They wouldn't be wealthy if their function systems hadn't achieved high levels of performance. If performance declines after destructive interventions, this is strongly felt in many walks of life. Important players may then change their mind and seek to correct the error. In post-communism, persistent mutual obstruction is a more likely outcome because function systems have not yet reached high standards of performance and the population has not got used to them.

Communist ideology, in essence, rejected the primacy of functional differentiation.[23] This held in particular for the economy, law and politics. After coming to power, communists did their utmost to reverse it, striving for dedifferentiation. In the confines of the territory they controlled, they largely succeeded. In the rest of the world, however, the primacy of functional differentiation has been strengthened. Rejecting the form of functional differentiation did not enable communists to create a formless society. There is no such thing. Moreover, complex societies need to make their primary form somehow observable for themselves, even though such self-descriptions are of necessity no more than crude simplifications.[24] Availability of a plausible self-description is a prerequisite for a steady and reliable reproduction of the form. The primary form of the society created by communists was hierarchical differentiation. The doctrine that made this observable inside the system was Lenin's teachings about the leading role of the Communist Party.[25] Post-communist transformation, similarly, was about a change of form, this time from hierarchical to functional differentiation. This is why transformation is an apt term.

This change of form implies an evolution of evolution, i.e. a change of the style of evolution. Under the primacy of hierarchical differentiation, variations had a chance to be selected if they appeared to be helpful for the stabilization of the hierarchy. This sort of stability was a primary criterion for selection. Applying it could be difficult and mistakes did happen, most notably in the stormy years between

---

[22] In Greece, for example, reciprocal obstruction of function systems became mutually reinforcing. The decline of manufacturing and agriculture resulted in political pressure to create even more public sector jobs.

[23] Concerning the law this was implied by the subordination doctrine which viewed the law as "concentrated politics". See Marković and Vuković (1978:550). The source illustrates that this doctrine was upheld even in the most liberal of all communist regimes. When Uzelac (2012) expounds the "overarching principle of instrumentalism" in socialist law, he points to an aspect of dedifferentiation. As far as the economy is concerned, the doctrine of planning found in innumerable textbooks of the political economy of socialism similarly implied dedifferentiation.

[24] See Luhmann (1995:298f) Functional differentiation was first observed as (enhanced) division of labor. This description became available in the 18th century.

[25] For a "classic" restatement see Stalin (1947:85-98).

1953 and 1957. Under the primacy of functional differentiation, every function system selects variations according to its own criteria and without regard to the stability of other systems. Moreover: "functional systems switch their mode of selection to essentially unstable criteria" (Luhmann, 2012:297). The functionally differentiated society does not have a central authority applying and enforcing a selection criterion that makes overall stability likely. Politicians may be under the illusion that this is up to them but they are wrong. If the selections performed by some function system oversupply other function systems with disturbances, corrections can be made only ex post. "Overall, society switches its stabilization efforts to reactive procedures. Society has become too complex and too opaque to set stability as an attainable goal" (ibid, 2012:295). An example of a reactive procedure was given above. If politics obstructs important industries, this will backfire on politics. After a while, politics will start to recognize and try to correct its mistake. This means that evolution routinely consists of three different components. It is not just about variation and selection, but also about re-stabilization, because selections, be they positive or negative (acceptance and rejection of a variation), inadvertently cause stability problems that will need to be treated subsequently.

2.3 SYSTEM DIFFERENTIATION IN POST-COMMUNISM: THE LAW AND POLITICS
In section 1, the Russian and the Ukrainian transformation were assessed as unsuccessful and incomplete. Section 2.2 revealed the theoretical foundations of this proposition. In Russia and the other CIS-countries, the "outdifferentiation" of function systems has remained incomplete. At least for some of the major function systems "operation closure" has not yet occurred. This can be shown for numerous function systems[26], but this paper is primarily concerned about the economy, law and politics. Corruption is a case in point. Of course, corruption cannot be extinguished. However, there is a difference between countries in which corruption is pervasive and well-entrenched, and others, where it is more of an occasional affair and largely limited to a few interfaces between different function systems that tend to be particularly corruption-prone like public procurement and socialized medicine. If corruption is pervasive, the distinction between the economy and politics is blurred, while occasional corruption honors it in the breach. Pervasive corruption means that politics and public administration generate and sell business opportunities. Businessmen are turned into subtenants of politics and are kept in a precarious position because the rental fee may change suddenly. In contrast to Eastern Europe, in Central Europe corruption is not pervasive.

System differentiation between the law and politics requires that politics is in actual fact constrained by the constitution of the country.[27] If this constraint is bind-

---

[26] The peculiar role of the Orthodox Church in contemporary Russia is noteworthy in this context and so is the lack of academic freedom. It indicates that operation closure has remained incomplete not only in the function systems discussed in this paper, but in others as well.

[27] See Luhmann (2004:410) for his concept of constitution. In the functionally differentiated society the constitution is a structural coupling between politics and the law. For more on such couplings see below.

ing, politicians take a serious political risk whenever they violate clear-cut constitutional provisions. Countries with incomplete differentiation usually have some sort of constitution as well, but its primary role is "to protect hierarchy's need for latency."[28] Generally speaking, a merely symbolic role of the constitution is incompatible with the full differentiation of function systems. Full functional differentiation between the law and politics implies that the law embarks on its own independent intra-societal evolution. What matters is that this evolution actually occurs, whether jurisprudence and legal theory have taken notice or not. Jurisprudence may need time to discover this evolution. To be sure, in England the discovery was made as early as the 17th century. It happened in England, because in Common Law evolutionary features are easier discernible.[29] On the Continent, the historic school recognized evolution, but its impact remained limited. As a result, Eugen Ehrlich, professor legum (of laws) at Czernowitz (Bukovina)[30], managed to make a stir, when he proposed in 1913: "We shall have to get used to the thought … that the intent of the author of a statute is a matter of absolute indifference so far as its effects are concerned. Once in force, it goes its own way".[31] The stir was symptomatic of the prevailing spirit, since Ehrlich's dictum was little else than a law-specific restatement of the most widely quoted Scottish definition of evolution: The result of human action but not of human design. The traditional neglect of evolutionary concepts in much of Continental legal theory explains, why in 1990 even the boldest post-communist reformers tended to think about law as a product of legislation rather than evolution.[32]

The distinction between law and politics has deep roots in Europe. The ancient idea of a legal right to resist tyrannical rule exemplifies this tradition.[33] Soviet rulers were confronted with it when they extended their empire towards the Balkan Peninsula and Central Europe. Historians have managed to trace some of the effects. They showed that even at the climax of Stalinism Czech courts were not always obedient servants of the secret service and communist officials. Seeking refuge in the courts sometimes helped citizens if their alleged political offences

---

[28] See Luhmann (1995:337): "The more starkly a system is hierarchized, the more clearly do forms whose latent function is to protect hierarchy's need for latency stand out."

[29] This was noticed already by Maine (1906:4) who wrote: "… that an Englishman should be better able than a foreigner to appreciate the historical fact…"

[30] Nowadays this is Černivci.

[31] Ehrlich (1962:375). For a much-quoted American statement of the evolutionary view, see Holmes (1991:1): "The life of the law has not been logic: it has been experience. The felt necessities of the time, the prevalent moral and political theories, intuitions of public policy, avowed or unconscious, even the prejudices which judges share with their fellow-men, have had a good deal more to do than the syllogism in determining the rules by which men should be governed. The law embodies the story of a nation's development through many centuries … In order to know what it is, we must know what it has been, and what it tends to become. We must alternately consult history and existing theories of legislation. But the most difficult labor will be to understand the combination of the two into new products at every stage…"

[32] Notice, that the evolutionary approach doesn't negate the role of legislation. It "only" insists that legislation cannot determine the development of law. And this is what Ehrlich said.

[33] In actual fact, this was primarily a right of the nobility who used it quite actively. Outside Europe, conflicts between kings and the nobility were common-place as well, but even if they were couched as matter of right the underlying notion of right was not clearly distinct from custom. As Berman (1983) points out, the establishment and amplification of this distinction constituted the unique advance made in Western Europe during the papal revolution of the 11th century.

FINANCIAL THEORY AND PRACTICE
40 (3) 293-318 (2016)

BRUNO SCHÖNFELDER:
THE EVOLUTION OF LAW UNDER COMMUNISM AND POST-COMMUNISM:
A SYSTEM-THEORY ANALYSIS IN THE SPIRIT OF LUHMANN

were considered minor. This was shown in Czech research about the expropriation and resettlement of so-called kulaks (large farmers)[34]. Notwithstanding these obstacles, communist governments deformed the law radically. All fields of law were thoroughly politicized. Law lost much and sometimes all of its autonomy. In this (and only this) regard, a number of communist countries, among them the GDR and Czechoslovakia, returned to the Dark Ages, since as Berman (1983) showed, the "outdifferentiation" of the law started as early as the 11th century[35]. A major difference between Central Europe and the Soviet Union was that a large part of the Soviet population failed to notice this return because it had never been much in contact with the law. To be sure, after the reforms of Alexander II, Russian law had started to evolve. However, in the remaining time before the communist take-over, only a small elite group became accustomed to this new development. What remained of the elite after the Civil War was largely liquidated during the Cultural Revolution (1928-31)[36]. Proceeding likewise was not feasible in Central Europe. The law had put down much deeper roots and a large part of the population had become used to it. As a result, communist propaganda found it difficult to convince people that a merger of law and politics was a progressive move. It was presumably for this reason that in late socialism Central European courts acquired more independence than Soviet courts. This can be demonstrated from the example of *telefonskoe pravo*.[37]

This analysis leads to the conclusion that a widely held view about post-communist legal reform is mistaken, because it underestimates the scale of the required change. After 1989, it was commonly thought that this reform was primarily about an effective protection of courts from political interference and about amending statute law. In actual fact, what was required was a radical change of the whole operating method of the law. There was a need to regain a nearly forgotten culture of refined legal argumentation, not to be confounded with political rhetoric. The preconditions for this change were much worse than e.g. in Austria or West Germany in 1945. Austrian and German judges and lawyers had not yet forgotten what they had learnt before 1933 or 1938, respectively. In Yugoslavia more of this culture had been preserved than in many other parts of Central and Eastern Europe, but this was mostly an academic affair. Moreover, even among Yugoslav professors there was a vocal minority group that cared little about legal argumentation proper and kept mixing up law and politics. Differences between the newly established and the traditional faculties of law tended to be large.[38] The old facul-

---

[34] See Petráš and Svoboda (2014:409f).

[35] The reader of Berman's work is likely to be struck by the distinctiveness which legal communication acquired – even vis-à-vis theology! The use of Latin was helpful in differentiating legal from everyday communication.

[36] See Fitzpatrick (1999).

[37] This Russian word means that the decision of the court is determined by a telephone call. The power of *telefonskoe pravo* [telephone law] in various communist countries is discussed in Schönfelder (2012:413-417). As of today, Russian *telefonskoe pravo* is alive and well.

[38] At times, some of the new faculties turned into a hotbed of dedifferentiating activism. Examples can be found e.g. in the *Zbornik Radova* of the Novi Sad faculty of law. For a textbook summarizing the "progressive" view see Popović (1981). He was a professor of the newly established law faculty of Niš. For a review of Yugoslav attempts to push the dedifferentiation of law ahead, see Pokrovac (1999).

ties kept their tradition of scholarship alive, but not much of this rubbed off on students, except a few unusually talented ones. These talented students rarely decided to become judges or solicitors and if they did they often changed their profession after a while. The judiciary was considered an unattractive employer.[39]

The limited role of legal argumentation in the daily operations of the judiciary is revealed by the texts it used and produced. In Czechoslovakia and the GDR, civil procedure essentially turned into some type of social work.[40] The writings of lawyers became indistinguishable from those of laymen. Most litigants were self-represented. In the GDR, legal representation was considered too unimportant to be mentioned in the otherwise meticulous protocol. Civil cases usually ended with a settlement. If litigants were intransigent and refused attempts at reconciliation, the court more often than not imposed a compromise. Judges educated litigants and treated them as immature creatures in need of parental guidance. If the judge did not impose a compromise but decided the case, he usually wrote a very brief opinion. These "socialist" opinions rarely delved into the legal questions raised by the case. If a citizen wanted to get legal information and consulted academic textbooks, academic journals or commentaries written by legal scholars etc., he usually did not feel illuminated either. Much of this literature was shallow and routinely avoided the in-depth discussion of controversial legal issues. In contrast, in Yugoslavia more and better texts were available. Yugoslav civil procedure never degenerated into social work. However, the skill of writing elaborate opinions was lost as well, and needed to be relearnt after 1990. This was a drawn-out process. An analysis[41] of Croatian opinions written in the 1990s found that an attempt to build a legal argument was undertaken only in 30 per cent of these texts. Usually it was discontinued after a few lines. Only 3 per cent cited at least one precedent. Citations of legal literature like commentaries, law journals or monographs were found even less frequently. Opinions written by appellate courts tended to be even less elaborate than opinions written by trial courts. This is disturbing because appellate courts should play a key role in establishing precedents. Research about opinions written by Czech judges leads to somewhat different conclusions. In contrast to some of their Croatian colleagues, they were diligent and hard-working persons. After 1990, they started to write long opinions that contained a detailed description of the facts established in court and the course the procedure had taken, the judicial advice given to litigants and so on. However, like the Croatian opinions, these Czech texts avoided discussing the legal merits of the case.[42] In either case, it would be wrong to blame judges for a failure that was primarily due to the loss of the tradition of legal argumentation.

---

[39] See, e.g. Uzelac (1992:582).
[40] See Markovits (1995, 2000) on the GDR. Markovits was primarily interested in the sociology of communist law. In most communist countries, the available research primarily takes a legal-history and legal-theory approach. According to Kühn (2005:XVI) Markovits' findings hold for Czechoslovakia as well. Mańko (2013b) similarly emphasizes the limited role of legal argumentation even in countries like Poland that managed to preserve more of their legal tradition.
[41] See Harašić (2006).
[42] For further details see Schönfelder (2012:408-412, 417-420, 506-521, 936-943).

FINANCIAL THEORY AND PRACTICE
40 (3) 293-318 (2016)

BRUNO SCHÖNFELDER:
THE EVOLUTION OF LAW UNDER COMMUNISM AND POST-COMMUNISM:
A SYSTEM-THEORY ANALYSIS IN THE SPIRIT OF LUHMANN

Luhmann (2004:330) argues that the social functions of legal reasoning have not been fully elaborated by the available theories of legal argumentation. Legal reason is "artificial reason"[43]. Its key contribution is "the creation of sufficient redundancy".[44] "Reasons are symbols for redundancy" (ibid:331) The long-run result of reasoning is the provision of "a web of points of view … we call it legal doctrine … The outcome of all these checks, which also serve to show what can go wrong, is a tradition of principles, rules, and doctrines but also of rejected alternative instructions which form the reservoir from which legislation and above all judicial lawmaking take their materials" (ibid:326-327). A developed culture of legal reasoning provides guidance to judicial decision-making, but it doesn't determine its results. It leaves room for controversy by limiting controversy. It serves as a "shield … against a constant and ultimately limitless political 'questioning', that is, as a rule for limitation for reasoning in its quest for reasons" (ibid:342). This is the ultimate defense against political pressures on judicial decision-making. If this defense has been put in place, "all efforts to steer courts onto a politically desirable course must confront the internal workings of courts. Most of the time they founder on the internal culture, the mode of argument operating within the legal system" (ibid:365). This is the reason why communists wanted to abandon tradition and rebuild the legal system from scratch. This rendered law defenseless against political interference. The missing defense cannot be adequately substituted for by formal fire-walls, such as the institutional independence of the judiciary that Italian-style self-administration offers.[45] The autonomy of the law should not be confused with institutional independence. Neither is it institutional independence plus something else. The judiciary is a large organization and large organizations are always infiltrated by politics. Italian-style institutional independence cannot prevent it. What really matters is what happens afterwards. Italian-style judicial independence makes it difficult to prevent long-term damage from infiltration. In contrast, in Germany, the (state or federal) minister of justice appoints judges. These decisions are indisputably political, but for this very reason the minister takes care to avoid the shameless politicization of judicial appointments that the political factions represented in the Italian High Judicial Council pursue. In German appointment processes, considerable emphasis is placed on provable professional expertise and competence. As a result, levels of professional competence tend to be high.

I have emphasized the role of routine and redundancy in legal reasoning and the discipline that it provides. The belief that this discipline can be replaced by organizational arrangements amounts to a confusion of two different types of social

---

[43] This is the famous wording of Chief Justice Coke who "rejected the authority that James I claimed to have over his own reasoning. Reasoning, in Coke's view, had to be 'artificial reason', that is, to be professionally induced through experience and competence." (Luhmann 2004:311).

[44] Luhmann (2004:319), who continues in a footnote: "Or, to use the words coming from legal practice, 'to keepe as neare as may be to the Certainty of the Law and to the Consonance of it to it Selfe', namely those of Sir Matthew Hale in an objection to Hobbes in the seventeenth century." Redundancy means routine, i.e. repetitive communication. Redundancy renders it possible to recognize the Self of the Law.

[45] See Wittreck (2006) for a critical analysis of Italy and Germany.

system. It fails to recognize the difference between the function (sub)systems of modern society and organizations.[46] To be sure, all function systems depend on organization – the law cannot operate without courts, the economy cannot operate without firms – but what they need is a great multitude of different organizations. Some play the role of centers while others are located in the periphery of their respective function system. In the case of the economy, banks and central banks form the center while real sector firms and other entities populate the periphery. In the case of the law, the courts form the center, while law firms, parliament and so on are in the periphery. This positioning of parliament demonstrates the radical divergence of Luhmann's approach from more traditional political theory.

2.4 DIVERGING VELOCITIES AND THE LOSS OF PUBLIC CONTROL OVER THE MARKET

In post-communism, the economic system evolved much faster than the legal system. When profit-maximizing behavior was no longer ostracized, the economy spurted. After the removal of administrative obstacles, it quickly reintegrated into the world economy, provided that some basic requisites were met such as a reasonably developed transportation network and a manufacturing sector enabling the country to generate export revenues outside the former socialist camp. This reintegration was tantamount to emancipation from political tutelage and sharply reduced political control over the economy. The speed at which this materialized stood in striking contrast to the vacillations of the law.[47] In Central Europe, the lagging of the law was not as obvious at first, because the enactment of new statutes seemed to proceed at a fast pace. Economists tend to be unaware of the difference between law and legislation and to overrate statute law. In Yugoslavia, the erroneous belief that social reality can be determined by statute production used to be referred to as *zakonomanija* (law mania) and *dekretomanija* (ordinance mania).[48] Economists (and socialists) are prone to these manias. They were over-impressed by the legislative activity of Central European parliaments. It took them years to detect "implementation problems". When the problems could no longer be overlooked, economists once more blamed them on insufficient statute-production. *Zakonomanija* has long become a sort of reflex action.

This fallacy can be detected in the writings of American economists as well, and the Nobel prize winner Joseph Stiglitz has distinguished himself as its most prominent proponent.[49] The wording differs. Instead of statute-production or delegated

---

[46] This distinction is elaborated in Luhmann (2013) and is one of his major contributions.

[47] Luhmann suggests a theoretical approach to measuring velocity. It employs the concept of eigenvalue, which in system theory means relatively stable parts of system structures. Stability is needed to forecast system operations. System-theory proposes that a system is unlikely to survive for long if it does not manage to create eigenvalues. In the legal system, eigenvalues are crucial for the predictability of judicial decisions. In the economic system, the key eigenvalue is the future value of money. In either case, it is intuitive that both social support for the law and the acceptance of money will decline if such forecasts tend to be extremely unreliable. In most post-communist countries, money was stabilized in the course of the 1990s. In contrast, forecasting the outcome of litigation remained very difficult. According to Rychetský (2004:92), in the Czech Republic, post-communist administration of justice resembled a "dark jungle". Its decisions were totally unpredictable.

[48] See Pašić (1981:225).

[49] See, e.g. Stiglitz (2003).

legislation he writes about "regulation". This vague term suggests the need for public control over the market. In Stiglitz's writings, it presumably means both the common law method of public control and direct public regulation[50]. He has not addressed the issue of whether direct public regulation should be subject to judicial review. Presumably, he meant some mixture of both methods and ignored the concern that judicial review was unlikely to constrain government interference in post-communism. Because of Stiglitz' authoritative intervention, the fallacy was "reconfirmed" as a putative truth and once more committed to the collective memory of the profession.

Among legal scholars, the fallacy is less popular. The author of this paper remembers how discussions with legal scholars went in 1991 and 1992. They argued that, in the short run, only liberalizing measures would produce an effect. The other provisions would become effective much later, because courts would need at least ten years to clarify their meaning. Only after this clarification could citizens be expected to abide by the law. By way of illustration, the decriminalization of private enterprise and "speculation", a liberalizing measure, had an immediate effect, and it was of considerable importance that it occurred in Czechoslovakia in 1990 while in most CIS-countries this came much later. In contrast, the provisions of company law that protect small stock-holders against block-holders and managers remained irrelevant for at least a decade. Economists often ignored this insight because it didn't fit in with their mindset. Legal scholars weren't quite right, either. They were too optimistic about the length of the lags. They should have said 30 years instead of ten, or maybe 50. They assumed that clarifying the meaning of the law would take post-communist courts no longer than West German courts. This was utterly unrealistic. The assessment of legal scholars was based on experience, not on theory.

After 1989, a rebirth of legal reasoning was needed. This was the key to the operative closure of the function system of the law, since "in the normal process of decision-making, the system … observes itself – as an accumulation of legal texts that refer to each other … The crucial point is that the system can 'recall' internal contexts from the past and through that reduce the scope of possible operations in the present. Finding the relevant texts … requires professional competence and thus it represents a crucial (and frequently overlooked) instance of legal skill" Luhmann (2004:305-306). Only after this skill is regained will solutions for doctrinal controversies be sought and found inside the function system of the law. When communist judges considered a case, they usually did not consult a variety of relevant legal texts[51] and even if they did, they usually had no intention of subjecting them to critical review, which could have resulted in an innovative, but system-compatible solution of the case at hand. In some Central and Eastern European countries as in the GDR, judges received detailed instructions concerning

---

[50] For the distinction see Posner (2011:487).
[51] Kühn (2011) suggests that Hungarian and Polish judges may have been exceptions to this rule.

how they were to construe the law. The instructors who performed this task were selected and appointed by the department for law and state that worked for the party leadership. Instructors used mimeographed texts, which, however, were either not handed out to judges or had to be returned. During their lectures, judges took notes. Guidelines changed frequently and their rationale was, as a rule, political, not legal. Critical discussion of these materials was not encouraged. If a judge had his own opinion about some question of law, he was required to seek prior authorization before he could decide a case accordingly. In 1989/90 all of this came to an end. All of a sudden, judges had to take decisions independently and on their own responsibility. Moreover, most post-communist countries witnessed a great shortage of legal texts written in the judge's mother tongue that could claim relevance in the new situation. As the Luhmann-quote indicates, this was a serious dilemma. The normal mode of system operations was blocked. Even in the Czech Republic which had a more lucrative market for legal texts than most other post-communist countries[52], it took more than a decade for a substantial body of relevant texts to be printed or made available electronically.

If post-communist legislation revitalized pre-communist legacies or copied foreign models judges could, as a matter of principle, consult pre-communist and foreign literature. In Czechoslovakia, the GDR and Yugoslavia, the pre-communist literature had been removed from the shelves of most libraries and much of it destroyed. Foreign literature was accessible only to the few who had learnt foreign languages well enough to read it. In Croatia, the old *Opći građanski zakonik* [General Civil Code] (OGZ, Austrian ABGB) was considered a second-string source of law after 1945. The judge was required to consult it if he was confronted with a legal question for which the new socialist law held no answer. In doing this, he was supposed to read the OGZ through socialist lenses. Austrian commentaries of the OGZ were not helpful for this task, and, consequently, they were rarely used by judges and lawyers. Moreover, knowledge of German has been much in decline among judges and lawyers. After 1990, change was slow. Likewise, in the Czech Republic the pre-communist heritage became widely used only after a monumental (5000 pages!) Czech commentary of the OGZ was reprinted. It had been written in the 1930s by two of the most prominent legal scholars of the country[53]. Following this example would not have been very helpful in Croatia because Croatian pre-communist discussions of civil law were less profound than the Czech.

Thus, in 1990 law started nearly from scratch and it is in its very nature that this was a slow beginning. The former British premier Gordon Brown put it like this: "In establishing the rule of law, the first five centuries are always the hardest."[54] As a result, scholars who believed that markets must be subject to public control ei-

---

[52] Except Poland, of course. The Czech market was lucrative, because the population was comparatively affluent. Moreover, Czech texts continue to be used in Slovakia.
[53] This is Rouček and Sedláček (1998).
[54] See *The Economist* September 19, 2015, p. 10.

FINANCIAL THEORY AND PRACTICE 40 (3) 293-318 (2016)

BRUNO SCHÖNFELDER: THE EVOLUTION OF LAW UNDER COMMUNISM AND POST-COMMUNISM: A SYSTEM-THEORY ANALYSIS IN THE SPIRIT OF LUHMANN

ther had to abandon the request that this control should be under the law, or consider post-communism a fatality that should not have happened. Actually, few were ready to face the challenge. Instead they complained. Stiglitz was the loudest complainant and was praised for the eloquence of his lament.[55] However, by 2005 one could have noticed that in Central Europe matters were not at all as bad as the academic tales of woe suggested. To be sure, the economy often had to struggle along without an adequate performance of the law.[56] This caused numerous problems and defects. However, most of the time, temporary expedients were found. No doubt, conditions were far from ideal, but liberalization nevertheless had a positive impact. Central Europe has fared better than CIS countries, which liberalized less and later.

The proposition that the law should have the lead and "regulate" the economy and other walks of life universalizes the course which history took in Europe (except in the Ottoman and the Russian empire), but nowhere else, neither in China nor in India, nor in Japan. In Europe, law acquired great significance as early as the 12th century. The "outdifferentiation" of the law was thus ahead of the "outdifferentiation" of the political system which gained momentum only in the 16th century. Canonical law was a key factor in these early advances. The early "outdifferentiation" of law strengthened, and was made possible by, a very pronounced stratification of society. As a result of this early evolution of the law, the privileges of the nobility were fully juridified in early modern times. This juridification was supported by theories that emphasized the natural inequality of the estates. Such theories were accepted in the 16th century, but in the 20th this mindset had long lost its appeal. Presumably, its restoration was inconceivable under communist rule. To be sure, in common parlance the privileges of communist officials were compared with the former nobility, and officials were referred to as a new nobility, but this did not restore the idea of natural inequality, although an attempt at revival was undertaken[57]. The failure to juridify the hierarchy meant that the upper class essentially remained outside the law. This division of society was a serious obstacle to the growth of the law and the expansion of its sphere of influence. Yugoslav history holds lessons about the limits of law under communist rule. Presumably, socialist Yugoslavia never came closer to the operational closure of the law than during the Croatian Spring (1969-72). Moreover, there was more economic and political liberty than at any other time during the Tito era. This resulted in civil

---

[55] For a critical review of various contributions to this debate see Schönfelder (2012).
[56] For the concept of performance see Luhmann (2013:96). It means the services the law provides to other function systems.
[57] For a review of Stalin's efforts see Fitzpatrick (1999:82). After the conquest of Belgrade, Tito did not waste time either and moved into the king's premises. Even though he probably never doubted that this was appropriate housing, he found it difficult to explain this view to the intellectuals of the country. According to Pirjevec (2013a:297, "Nažalost, naši su ljudi uglavnom seljaci, a ti znaš…") he argued that he was ruling a nation of peasants who were much impressed by pomp and rank and wouldn't respect a ruler who dispensed with it. However, this was at best a temporary excuse. According to the party-platform, peasants were to be reformed and assimilated to the condition of workers. This observation as well as the ultimate failure of Stalin's attempts to lay the ground-work for a juridification of privilege, support the proposition that under communist rule the normalization of the improbable remained insufficient for further evolution.

unrest which was thought of as a threat to the Yugoslav federation and communist rule and provoked a harsh backlash.[58]

2.5 FREEDOM OF CONTRACT AND FREE USE OF INDIVIDUALLY ATTRIBUTED PROPERTY AS STRUCTURAL COUPLINGS BETWEEN THE LAW AND THE ECONOMY

Above it was argued that in post-communism reforms could not be sequenced in the way that the public-control view suggests. In this section it will be shown that even if this problem had been solved somehow, the appearance of suitable, market-friendly controls would have been unlikely. Regulations need to be created by somebody. Who should be their creator? Presumably, the political system is the only available candidate. The legal system is too immature and too slow to make significant contributions in time. However, may we hope that the political system will perform reasonably well? Not really, if Luhmann is right. Its capacity is limited and the likely standards of achievement and performance are low. As far as the economic system is concerned, low standards imply backwardness and poverty; in the legal system they stand for the short reach and powerlessness of the law. If one wants more, the operational closure of function systems cannot be avoided. This means that function systems are released into autonomy and launch their own system-specific evolution. One may object that this may generate clashes between function systems. Indeed, such clashes cannot be ruled out. If the economic system embarks on a journey of its own making, nobody can guarantee that the results are bearable from the viewpoint of the law. The autonomous evolution of the function system of the law causes similar concerns. If left to itself, the law may evolve in ways that are incompatible with economic growth. This would force entrepreneurs to avoid the sphere of the law and conduct their affairs in lawless ways. In contrast to 19th century concepts of evolution, Luhmann's theory does not believe in natural selection and survival of the fittest. It rather asks how a certain minimum degree of adjustment can be maintained. It considers maladjustment as a likely outcome of evolution. The real question is then how readjustments can be initiated.

The economic analysis of law has often taken a more optimistic view. Posner (2011:714) claims that "judge-made rules tend to be efficiency-promoting while those made by legislatures, other than those rules that codify common law principles tend to be efficiency-reducing." According to him, adjustment problems between the law and the economy will disappear in the course of time, if the legislature doesn't interfere. An optimal fit will be reached because a hitherto unknown sort of invisible hand guides judicial decision-making. This is definitely not the invisible hand referred to by Adam Smith. In Posner's view, the only problem with judicial law-making is the slow speed at which common law rules change. "Judge-made law … will not do when new law has to be made in a hurry, or when a big

---

[58] In retrospect, it seems clear that Tito saved the federation by means which rendered it non-viable in the long run. For one of the best accounts of the Croatian Spring see Pirjevec (2013b).

change in law is desired." German experience with judge-made law does not seem to confirm this optimistic stance. Important parts of German labor law are judge-made. It is efficiency-reducing rather than efficiency-promoting although it is not as bad as the French legislature-made labor law.[59] Before we are ready to trust the new invisible hand invoked by Posner, its nature and working should be explained in some detail. The economic analysis of law has tried to meet this challenge, but has failed to come up with a convincing story. All of its attempts seem to have in common, that the outcome crucially depends on the intellectual abilities of judges. They assume that the judge's mind can somehow grasp economic rationality and decide accordingly. As far as simple controversies are concerned, this is perhaps not utterly unrealistic. Nevertheless, according to Luhmann, it is unwise to rely on this mechanism because it is unlikely to ensure a sufficient adjustment of the law to the needs of the economic system.

In Central Europe, we have a particular type of empirical evidence suggesting that we should not trust Posner's invisible hand. This is the teaching effectiveness of economics courses taught in law school. In Germany as well as in the former Habsburg Empire and its successor states, students of law have long been required to take some economics courses. In Germany, most law schools abolished this requirement in the 1960s. Ever since, nearly all German law students have refused to study economics. In Austria, Croatia and some other countries the old requirement is still in force. As a result Austrian and Croatian judges are much better educated about economics than German judges, and one may hypothesize that this enlightens them about the economic consequences of judge-made rules and enables them to take more efficiency-promoting decisions than are commonly found in Germany. Unfortunately, there exists, to my best knowledge, no evidence supporting this hypothesis. Being a professor of economics himself, the author of this paper would be immensely pleased to learn that the efforts of his Austrian (and Croatian) colleagues have not been in vain. Unfortunately, he is still waiting for the good news. This observation suggests that Luhmann's skepticism is warranted.

If this is taken for granted, we need other mechanisms to ensure a certain degree of compatibility between the economy and the law. Luhmann refers to these mechanisms as structural couplings. According to him, individually attributed property, contract and competition are the three crucial couplings, but they can carry out their tasks only after most legal controls on contract and property have been removed. They can be removed only if there is sufficient competition, because in the absence of competition the economic and social consequences of freedom of contract and free use of property are unbearable. Removal of most controls is, of course, not tantamount to removal of all controls. These structural couplings require a considerable degree of economic liberty, but not "total" freedom. These familiar principles of free enterprise reappear in Luhmann's theory, but the positioning differs from classical liberalism. They reappear as structural

---

[59] See Hardege and Schmitz (2008) and Valentin (2002:230).

couplings between two subsystems of society and not, as classical liberalism had it, as the very "foundations of law and society per se"[60]. Freedom of contract, of course, does not mean that all contracts will be enforced by courts. Courts may and should refuse to enforce certain types of contracts (like contracts for criminal purposes or contracts in restraint of trade). Moreover, before a contract can be enforced, courts need to form an opinion about the proper "interpretation of the contracting parties' expression of intent based on their presumed interests" (Luhmann, 2004:397). Neither does the proposed need for structural couplings imply that all industries should be turned over to free enterprise as soon as possible. What matters instead is that its realm is large enough, that it prospers and provides a good living for droves of lawyers who give legal advice to business people and represent them in court. If this condition is met, litigation will in the course of time produce a body of judge-made rules that will enable entrepreneurs to forecast which contracts are likely to be upheld and enforced by the court. Freedom of contract enables businesses to contract around clauses deemed unacceptable by judges. To be sure, more often than not the rulings of courts will perplex litigants and other observers and their rationale will be difficult to defend from the viewpoint of economic rationality, but, as a rule, such failures will be found tolerable under freedom of contract and individually attributed property, because they provide sufficient opportunities for these pitfalls to be avoided by writing different contracts. Freedom of contract, thus, is a safety net for the failures of courts. Operational closure of law creates hazards, but the safety net makes them bearable.

Yugoslav civil law illustrates what may happen if the safety net is removed. Freedom of contract and free use of property were rejected by Yugoslav communists.[61] Yugoslav law tried to regulate the use of private property down to the smallest detail. This petty tutelage kept private business away from the courts. Yugoslavia always (except 1949-1951) had a large private sector, which employed a sizable part of the labor force and generated a much larger share of GDP than official statistics show.[62] The judicial caseload statistics (civil cases commenced) reveal that its share in the civil caseload remained very small (CBS, 1991). The private sector preferred to conduct its business in ways that hindered investigations and discovery. Most agreements were oral and not written, which renders it difficult to identify breach by failure of performance. Law of contract, property and tort was codified only in 1978 or in 1980 and, in spite of the long gestation lag, these codifications left many controversial issues open. Before codification these fields of law were, according to official Yugoslav doctrine, primarily regulated by judge-made law. According to Alinčić (1994:148 and 2005:179) this doctrine made the socialist rejection of freedom of contract and free use of property somewhat more

---

[60] Luhmann (2012:112). This marks the contrast to Hayek (1989) who appeals to the approach of classical liberalism.
[61] For this and the following statements see Alinčić (1994:54f; 2005:83) who points out that throughout much of the history of socialist Yugoslavia, the public administration and judiciary tended to be more hostile to entrepreneurship and private property than the wording of party documents suggested. For examples see also Gams (1998:435f). This changed only in the 1980s and only in parts of the country.
[62] See Lydall (1984:267f).

FINANCIAL THEORY AND
PRACTICE
40 (3) 293-318 (2016)

BRUNO SCHÖNFELDER:
THE EVOLUTION OF LAW UNDER COMMUNISM AND POST-COMMUNISM:
A SYSTEM-THEORY ANALYSIS IN THE SPIRIT OF LUHMANN

bearable, but that did not help much.[63] Socialist civil law hindered rather than assisted trade. As an example, consider the law of real property. Communists were preoccupied with private property. Since in Yugoslavia most real estate always remained in private ownership, this preoccupation gave rise to numerous efforts at reducing its marketability. Seen from a historical perspective, the communist transformation of property law resembled a return to the 18th, if not the 12th century.[64] Observers were struck by this urge. One of them noted accordingly "that Yugoslav law … is in the process of developing new forms of almost feudal complexity. This is not surprising because, after all, the refinements of feudalism represented different types of control of land, and it is this development of different types of control and enjoyment of land which is typical of contemporary Yugoslav land law".[65] Medieval land law often restrained the alienation of land. Divided ownership and estates in land made transfer difficult even if there was no formal limitation. Transferability is decisive for operational closure of the economic system. "The history of English land law is a history of efforts to make land more easily transferable" (Posner, 2011:95). The unitization of divided ownership meets powerful resistance even it clearly creates value. Most tenancy holders oppose change and, even if they are ready to cooperate in principle, they have every reason to drive a hard bargain. Overcoming such resistance has taken centuries in Western Europe. In post-communism, the road back from the 12th century was long and slow. There were no short cuts. In many post-communist countries, as a result, much real estate will not become easily transferable for a long time.[66]

3 CONCLUDING REMARKS AND SUGGESTIONS FOR FURTHER RESEARCH

In post-communism, the evolution of law proceeded at a comparatively slow pace and fell behind the evolution of the economy. This difference of speed did not end up in the catastrophe predicted by some observers. Nevertheless, it seriously impaired the operations of the economic system and in particular of the financial sector, which depends more on the performance of the law than most other industries.

At this point, our analysis blends in with several strands of economic literature. One of them is the literature on sovereign debt. If debt collection is difficult or impossible, the debtor resembles the prototypical sovereign debtor[67]. Non-enforceability of contracts with sovereign debtors has long been considered a major reason why international capital markets are relatively small and limited compared with the national capital markets of countries which are under the rule of law. Another strand of literature has focused on the persistent adverse effects of temporary financial crises. A "widening of the agency friction" caused by declin-

---

[63] See also Schönfelder (2012:442ff and 481f).

[64] It is the 12th century, if English land law is the yardstick, and the 18th if French land law is referred to.

[65] See Chloros (1970:147) who also notes that this was a reversal of the French revolution. "On the Continent, as a result of the French Revolution, the absolute right of ownership was restored and thus it became an expression of individualism, of the rights of man … On the eve of 1789, land law in France was far removed from the law of Justinian. 'It is the Code civil which has given to property its individualistic character'."

[66] Poland has been an exception to this rule. I owe this information to an anonymous referee.

[67] See Obstfeld and Rogoff (1996:350f).

ing collateral value was identified as a key generator of these effects.[68] It is well known that an underdeveloped law of property and security interest widens this friction as well.

A third strand evaluates the efficiency of debt collection[69], but fails to appreciate the sociological implications of its findings. Luhmann (2004:401) suggests that the enforcement of civil judgement is a useful indicator for the "outdifferentia-tion" of the law. Civil law is about legal obligations that are created outside political control. A credit contract is an obvious example: the creditor and the debtor do not ask politics for permission before they conclude it. Enforcement of such claims means that the police, i.e. the use of physical power is at the disposal of the creditor, if this is needed to enforce his claim, and that his use of this power is limited only by the law and not by political considerations. Communists wanted to abolish the enforcement of civil judgements, but realized that this would take some time. So they took intermediate steps and changed civil procedure in ways that weakened and marginalized enforcement.[70] Although these ideological beliefs faded in the 1970s the intermediate steps were not reversed. This was a push towards dedifferentiation that survived to the end of the communist regime and beyond. Early legal reforms undertaken in post-communism often failed to appreciate the significance of this issue. The result often was a decline of confidence in the legal system, which occurred as soon as creditors realized that the hard-won judgements were worthless because they could not be enforced. In system-theory terms this decline is a destabilization of the legal system. Unfortunately, the judiciary could do rather little to restore confidence because legislation was needed to improve enforcement. Enforcement is about the use of powers which are under the control of the executive branch of government. Stability problems of the legal system are unlikely to bother an up-to-date (as opposed to a 19th century) parliament[71]. Legislative initiatives are likely to be successful only if they are supported by the ruling party and this party is likely to grant support only if the issue has turned into a political problem and, thus, has arrived in the political system. The crossing of system boundaries is likely to happen only if the problems of the legal system have resulted in serious economic problems like a crisis of the financial system and, consequently, of the real economy. Then matters can no longer be ignored. The theory presented in this paper suggests that strengthening enforcement is re-stabilization – remember the proposition, that functional differentiation transforms evolution into a three-step movement – and a sort of re-stabilization that is likely to occur only after a crisis extending far beyond the legal system.

---

[68] See, e.g. Hall (2010:7).

[69] See, e.g. Djankov et al. (2008).

[70] This was an explicit purpose of the reform of Czechoslovak civil procedure undertaken in 1964. See Rubeš et al. (1970:189). Similar developments occurred in the GDR and Yugoslavia, but the details differed a lot. See Thaetner (2003) and Schönfelder (2012:454-464 and 968f). The ideological background was that communists expected a merger of civil law and morals. After the merger, enforcement usually would be achieved by moral pressure. How about individuals who refuse to succumb to this pressure? Presumably, they would be removed. See the treatment of "outsiders" reported in Markovits (1995).

[71] For a striking example for the failure of the legislator to react to the requests of a judiciary that suffered from this crisis of confidence, see Crnić (2004).

This implication is empirically testable. Schönfelder (2012) conducts the test for four countries and finds Luhmann confirmed. Further research could extend this to other post-communist countries.

This paper focused on a small subset of formerly communist countries. It may be instructive to extend the analysis to countries such as Poland in which the dedifferentiation of the law was not pushed as far.[72] System-environment theory suggests that this should make post-communist transformation easier. If this can be shown, it would provide an empirical test for the theory.[73] If less dedifferentiation helped in Poland, why did it help so little in most if not all successor states of former Yugoslavia countries? Schönfelder (2012) demonstrates that in Croatia the legal system developed more slowly than in Slovakia and the Czech Republic. A Luhmannian attempt to explain this difference could point to the conflict between nation-building and functional differentiation. In Luhmann's words: "The idea of nation … belongs to the set of short-lived semantics that can exercise a fascination for a transitional period without betraying what societal system they refer to. It can therefore be assumed that this idea is now on the wane, a phase in which it does more harm than good and from a sociological point of view constitutes one of those *obstacles épistémologiques* that for reason of past plausibilities block urgently needed insights."[74] The idea of nation proposes a unity above function systems. However, no such unity can exist nowadays. Function systems operate on a world-wide scale. A far as the economic system and science are concerned, this is close to obvious, but it holds for the law and for politics as well. All attempts to build a national economy are futile and counterproductive. A comparison between Croatia and Slovakia may be illustrative at this point. In the 1992s, the Slovaks got their nation-state and this was an effortless achievement. As soon as the Slovaks had it, they found out that it did not solve any real problem. It took them little time to realize this and act accordingly. In contrast, the Croatian nation-state was hard-earned and came at the cost of great sacrifice. Small wonder that Croats have tended to overestimate this gain. To test the theory one might study Slovenia, which should come out as an intermediate case between Croatia and Slovakia.

In this paper, system-environment-theory has been contrasted with the approach taken by economists. A juxtaposition of legal theory-approaches may be instructive as well. Legal theorists coined the notions of ultra-formalism or hyperpositivism that are meant to capture certain characteristics of late socialist and post-communist judicial decision-making.[75] The theory presented in this paper should be seen as a congenial to this view. It makes a stab at deepening the analysis. The hyperpositivism-literature emphasizes that in the 1970s and 1980s judicial behav-

---

[72] For a comparison see Kühn (2011:25f).
[73] In the 1990s, Jacek Rostowski (oral communication), praised Polish citizens for being amazingly law abiding – amazingly in view of the problems of poverty that plagued the country. Rostowski was an astute observer. If he was right, this would count in favor of the Luhmannian approach taken in this paper.
[74] Luhmann (2013:289) uses the French expression *obstacles épistémologiques* himself.
[75] Kühn (2005, 2011) and Mańko (2013a,b).

iour changed. The bluntly political abjudications of Stalinist courts largely were a matter of the past. In the vein of system-environment theory this change can be classified as a first step towards a renewed "outdifferentiation" of the law. Adherence to judicial procedure and obedience to the letter of the law rendered judicial decision-making distinct from the working style of the executive branch of government and the Communist Party which continued to despise such attitudes.[76] In post-communism operational closure became feasible. Whatever judges had learnt in school, they couldn't help noticing that all of sudden they had to *decide* cases – and not just arbitrate them, as they often did in the GDR and Czechoslovakia, or double-check or implement decisions taken by the prosecutor or politics. And these weren't the petty and (for the most) easy cases which filled the civil dockets of communist courts.[77] Post-communist judges couldn't help noticing that "hard cases" cannot actually be *decided* within the narrow confines of hyperpositivism. System-environment theory proposes that operational closure creates indeterminacy inside the system. Indeterminacy is a prerequisite for decision-making.[78] Henceforth, decision-making needs to be guided by a "system memory", but this memory grows only as a by-product of system-operations. In this predicament, judges grasped at straws. They employed hyperpositivism to make excuses. In addition, hyperpositivism may have been another *obstacle épistémologique,* but the theory would suggest that in view of the paramount importance of operational closure this should be a matter of secondary importance. If this proposition is right, hyperpositivism should not be regarded as a "third legal tradition" on the same footing as Common Law and Roman Law.[79]

---

[76] In communist Czechoslovakia and the GDR public administration and the Party nearly always preferred to take an informal approach neglecting procedural niceties and emphasizing political correctness ("following the party line"). See Bernet (1995), Kabele (2004:46) and Schröder (2008:42). In Yugoslavia this held for party officials, but not necessarily for public administration.

[77] In Yugoslavia as well as in the GDR and Czechoslovakia a political decision-maker "guided" the court whenever a controversy was deemed politically important – except during the Croatian spring. See Pokrovac (1988) and Uzelac (2012). The handling of unimportant cases differed from the GDR and Czechoslovakia: they were often decided by Yugoslav judges rather than arbitrated. See Schönfelder (2012:472).

[78] Luhmann (2000:147) approvingly quotes Shackle's dictum: "Choice is an exploitation of unknowledge."

[79] If Uzelac (2012) insisted on equal footing, I would disagree.

REFERENCES

1. Alinčić, M. [et al.], 1994. *Hrvatsko građanskopravno uređenje i kontinental-noeuropski pravni krug.* Zagreb: Pravni fakultet.

2. Alinčić, M. [et al.], 2005. *Građansko pravo i pripadnost hrvatskog pravnog poretka kontinentalnoeuropskom pravnom krugu: teorijske osnove građan-skog prava.* Zagreb: Pravni fakultet.

3. Berend, T. I., 1990. *The Hungarian economic reforms 1953 – 1988.* Cambridge: Cambridge University Press.

4. Berman, H. J., 1983. *Law and Revolution. The Formation of the Western Legal Tradition.* Cambridge: Harvard University Press.

5. Bernet, W., 1995. "Verwaltungsrecht" In: Uwe-Jens Heuer, ed. *Die Rechtsord-nung der DDR: Anspruch und Wirklichkeit.* Baden-Baden: Nomos.

6. CBS, 1991. *Statistički Godišnjak Republike Hrvatske.* Zagreb: Croatian Bureau of Statistics, pp. 372-373.

7. Chloros, Alexander G., 1970. *Yugoslav Civil Law.* Oxford: Clarendon.

8. Christov, C., 2007. *Tajnite faliti na kommunizma.* Sofija: Ciela.

9. Crnić, I., 2004. "Uvodne napomene uz novelu Ovršnog zakona iz 2003." In: I. Crnić, ed. *Novo ovršno pravo.* Zagreb: Narodne novine.

10. Djankov, S. [et al.], 2008. Debt Enforcement around the World. *Journal of Political Economy*, 116(6), pp. 1105-1149. doi: 10.1086/595015

11. Drahokoupil, J. and Myant, M., 2011. *Transition Economies: Political Economy in Russia, Eastern Europe, and Central Asia.* New York: Wiley.

12. Durman, K., 1998. Útek od praporu. Praha: Karolinum.

13. Ehrlich, E., 1962. *Fundamental Principles of the Sociology of Law.* New York: Russell and Russell [1913].

14. Epstein, R., 1995. *Simple Rules for a Complex World.* Cambridge: Harvard University Press.

15. Fitzpatrick, S., 1999. *Everyday Stalinism. Ordinary life in extraordinary times: Soviet Russia in the 1930s.* New York: Oxford University Press.

16. Gajdar, E., 2007. *Gibel' imperii.* Moskva: Rosspen.

17. Gajdar, E., 2012. *Sobranie sočinenij, tom 2.* Moskva: Delo [1994].

18. Gambetta, D., 1988. Fragments of an economic theory of the mafia. *European Journal of Sociology,* 29(1), pp. 127-145. doi: 10.1017/S0003975600005610

19. Gams, A., 1998. *Svojina.* Beograd: Naučna knjiga.

20. Hall, R., 2010. Why Does the Economy Fall to Pieces after a Financial Crisis? *Journal of Economic Perspectives,* 24(4), pp. 3-20. doi: 10.1257/jep.24.4.3

21. Harašić, Ž., 2006. "Gerichtsargumentation" in Kroatien. In: D. Fischer, ed. *Transformation des Rechts in Ost und West.* Berlin: BWV, pp. 273-282.

22. Hardege, S. and Schmitz, E., 2008. *Die Kosten des Kündigungsschutzes in Deutschland.* Köln: Deutscher Institutsverlag.

23. Hardy, J., 2009. *Poland's New Capitalism.* London: Pluto Press.

24. Havel, V., 2012. *Moc bezmocných a jiné eseje.* Praha: Knihovna Václava Havla [1978].

25. Hayek, F. A., 1989. *The Fatal Conceit: the Errors of Socialism.* Chicago: University of Chicago Press.

26. Holmes, O. W. Jr., 1991. *The Common Law*. New York: Dover [1881].

27. Irwin, D., 2014. Adam Smith's "Tolerable Administration of Justice" and the Wealth of Nations. *NBER WP*, No. 20636.

28. Kabele, J., 2004. Výkladi vládnutí v reálném socialismus. Praha: Matfypress.

29. Kühn, Z., 2005. *Aplikace práva soudem v èře středoevropského komunismu a transformace*. Praha: Beck.

30. Kühn, Z., 2011. *The Judiciary in Central and Eastern Europe*. Leiden: Martinus Nijhoff. doi: 10.1163/9789047429005

31. Luhmann, N., 1990. *Die Wissenschaft der Gesellschaft*. Frankfurt: Suhrkamp.

32. Luhmann, N., 1995. *Social Systems*. Stanford: Stanford University Press.

33. Luhmann, N., 2000. *Die Politik der Gesellschaft*. Frankfurt: Suhrkamp.

34. Luhmann, N., 2004. *Law as a Social System*. Oxford: Oxford University Press.

35. Luhmann, N., 2012. *Theory of Society, vol. 1*. Stanford: Stanford University Press [1997].

36. Luhmann, N., 2013. *Theory of Society, vol. 2*. Stanford: Stanford University Press [1997].

37. Lydall, H., 1984. *Yugoslav Socialism*. Oxford: Clarendon.

38. Madžar, Lj., 1990. *Suton socialističkih privreda*. Beograd: Ekonomika.

39. Maine, H., 1906. *Ancient Law*. New York: Henry Holt [1861].

40. Mańko, R., 2005. The Culture of Private Law in Central Europa after Enlargement: A Polish Perspective. *European Law Journal,* 11(5), pp. 527-548. doi: 10.1111/j.1468-0386.2005.00275.x

41. Mańko, R., 2013a. Survival of the Socialist Legal Tradition? A Polish Perspective. *Comparative Law Review,* 4(2), pp. 1-28.

42. Mańko, R., 2013b. Wees in the Gardens of Justice? The Survival of Hyperpositivism in Polish Legal Culture as a Symptom/Sinthome. *Pólemos – Journal of Law, Literature and Culture,* 7(2), pp. 207-233.

43. Marković, R. and Vuković, I., 1978. *Socialistički samoupravni system SFRJ*. Beograd: Privredni pregled.

44. Markovits, I., 1995. *Imperfect Justice: An East-West German Diary*. Oxford: Oxford University Press.

45. Markovits, I., 2000. *Justice in Luritz: Experiencing Socialist Law in East Germany*. Princeton: Princeton University Press.

46. McKinnon, R., 1992. Spontaneous Order on the Road Back from Socialism: An Asian Perspective. *American Economic Review,* 82(2), pp. 31-36.

47. Moeller, H-G., 2006: *Luhmann Explained. From Souls to Systems*. Chicago: Open Court.

48. Možný, I., 2009. *Proč tak snadno…* Praha: Slon [1991].

49. Nelson, R. and Winter, S., 1982. *An Evolutionary Theory of Economic Change*. Cambridge: Harvard University Press.

50. Obstfeld, M. and Rogoff, K., 1996. *Foundations of International Macroeconomics*. Cambridge: MIT Press.

51. Pašić, N., 1981. O položaju, ulozi i perspektivama politikologije. *Politička misao,* 18(3), pp. 221-225.

52. Petráš, J. and Svoboda, L., 2014. *Osm let po válce*. Praha: Ustav pro studium totalalítnich režimu.

53. Pirjevec, J., 2013a. *Tito i drugovi*. I deo. Beograd: Laguna.

54. Pirjevec, J., 2013b. *Tito i drugovi*. II deo. Beograd: Laguna.

55. Pokrovac, Z., 1988. Realsocializam i pravo. *Savremenost*, 18(166-167), pp. 44-50.

56. Pokrovac, Z., 1999. "Zur Geschichte der Vergesellschaftung der Justiz in Jugoslawien" In: G. Bender and U. Falk, eds. *Recht im Sozialismus. Band 2. Justizpolitik*. Frankfurt: Klostermann.

57. Popović, S., 1981. *Samoupravno pravo*. Beograd: Privredni pregled.

58. Posner, R., 2011. *Economic Analysis of Law*. New York: Aspen.

59. Roland, G., 2000. *Transition and Economics*. Cambridge: The MIT Press.

60. Rouček, F. and Sedláček, J, 1998. *Komentář k Československému obecnému zákoníku občanskému a občanské právo platné na Slovensku a v Podkarpatske Rusi*. Praha: Aspi [1935-1937].

61. Rubeš, J. [et al.], 1970. *Občanský soudní řad*. Praha: Orbis.

62. Rychetský, P., 2004. "Gewährleistung der richterlichen Unabhängigkeit und des Rechtsschutzes unter veränderten Rahmenbedingungen" In: F. Schroeder, ed. *Justizrefrom in Osteuropa*. Frankfurt: Lang, pp. 251-256.

63. Sachs, J. and Woo, W. T., 1994. Structural factors in the economic reforms of China, Eastern Europe, and the Former Soviet Union. *Economic Policy,* 9(1), pp. 102-145. doi: 10.2307/1344459

64. Schönfelder, B., 2012. *Vom Spätsozialismus zur Privatrechtsordnung. Eine Untersuchung über die Interdependenz zwischen Recht und Wirtschaft am Beispiel von Gläubigerschutz und Kredit*. Berlin: Berliner Wissenschafts-Verlag.

65. Schröder, R., 2008. *Zivilrechtskultur der DDR, Vol 4*. Berlin: Duncker und Humblot.

66. Schürer, G., 1998. *Gewagt und verloren*. Frankfurt: Frankfurter Oder Editionen.

67. Stalin, I., 1947. *Fragen des Leninismus*. Moskva: Izd. lit. na ino. jazykach.

68. Stiglitz, J., 2003. *Globalization and its Discontents*. New York: Norton.

69. Thaetner, T., 2003. *Die Zwangsvollstreckung in der DDR*. Berlin: BWV.

70. Turek, O., 1995. *Podíl ekonomiky na pádu komunismu v Českolovensku*. Praha: Ústav pro soudobé dějine ČR.

71. Uzelac, A., 1992. Zavisnost i nezavisnost. *Zbornik Pravnog fakulteta u Zagrebu,* 42(Suppl. 4), pp. 575-594.

72. Uzelac, A., 2010. Survival of the Third Legal Tradition? *Supreme Court Law Review,* 49(2d), pp. 377-396.

73. Valentin, V., 2002. *Les Conceptions néo-libérales du droit*. Paris: Économica.

74. Wiles, P., 1992. "Capitalist triumphalism in Eastern Europe, or the economics of transition: an interim report" In: A. Clesse, ed. *Preventing a new East-West divide*. Baden-Baden: Nomos, pp. 389-407.

75. Wittreck, F., 2006. *Die Verwaltung der Dritten Gewalt*. Tübingen: Mohr.

Financial transactions taxation in the European Union and Croatia

MASSIMO MILEVOJ, MEcon*

_____
Massimo MILEVOJ
e-mail: massimo789@gmail.com

Abstract

The paper provides a systematic review of what is known about the financial transactions tax, its practical application in EU countries and elsewhere in the world and its impacts on financial markets. A special emphasis is placed on the outlook for the taxation of financial transactions in the EU and the possibilities and constraints in the Republic of Croatia. The results of the analysis show the propensity of countries to the taxation of financial transactions, particularly because speculations on the financial markets before the outbreak of the crisis enabled high economic rents to be generated, while during the crisis period, failed speculations were paid for with ample government support because of the fear that the stability of the financial sector would be distorted. Analysis of the possibility of taxing the Croatian capital market according to the proposed European model shows that the inclusion of the Republic of Croatia in the common taxation procedure would be justified, although the revenue from such a tax would be relatively small.

Keywords: financial transactions tax, taxation of the capital market in the EU and Croatia

1 INTRODUCTION

The world's financial market underwent a surge in development in the two decades before the outbreak of the financial crisis. The turnover of financial transactions at global level in 1995 exceeded the value of world GDP more than twenty times, and in 2007 more than seventy times the turnover of world GDP at that period (European Parliament, 2010). Financial markets, then, are an important tax source, and many states apply various forms of the taxation of financial transactions.

In 2009, leaders of G-20 countries suggested to the IMF that it should investigate the ways in which the financial sector could make a more equitable contribution to public finances. By way of answer, in 2010 the IMF proposed two approaches. The first was based on targeted levies that financial institutions would pay into a common fund, the resources of which would be used for the resolution of failed financial institutions, endeavouring to prevent future crises in the financial sector. The second proposal covered the possibilities of taxing the activities of the financial sector at a general level, that is, the introduction of a tax on financial transactions (FTT, Financial Transaction(s) Tax) and a tax on financial activities (FAT, Financial Activities Tax). Although no G-20 level agreement has been achieved, some countries have begun to tax the financial sector more stringently (IMF, 2010).

In 2011, forty states were implementing some form of tax on financial transactions, most frequently on a unilateral basis, with many specific features in the form and manner of the taxation. In spite of that fact, the relative revenue from the tax is very low, and hardly reaches 0.5% of the GDP of the given country. In the last twenty years before the economic crisis, the importance of this form of taxation had fallen because states had endeavoured to reduce the costs of capital in

order to enhance the competitiveness of their domestic financial markets (Matheson, 2011).

The following chapter sets out some theoretical considerations about the tax on financial transactions, individual effects on the market are investigated (such as volatility, liquidity, cost of capital, volume, efficacy), tests out the effect on tax revenue and ascertains the tax incidence. Part three refers to the European system of financial transaction taxation. Part four considers the outlook for the taxation of financial transactions in the Republic of Croatia, with a reference to budgetary capacities.

2 THE IMPACTS OF THE INTRODUCTION OF A TAX ON FINANCIAL TRANSACTIONS

The form in which a state becomes involved in market movements via taxation always has certain implications. Those in favour of introducing a financial transactions tax think that it can improve market movements, encourage the development of economic relations and reduce speculative trading, while opponents consider that this type of taxation should increase liquidity.

IMPACTS ON TRADING VOLUME

Empirical studies indicate that greater transaction costs will reduce the volume of trading, while the intensity will depend on market elasticity. Some studies have calculated elasticity exclusively on the basis of changes in the tax rate, but others have taken into their base the difference between purchase and sale, and the most accurate take into consideration the total costs of transactions. If trading volume elasticity is calculated on the basis of some of the components of total transaction costs (changes in the tax rate, bid-ask spreads) then elasticity with respect to total costs will be higher (Matheson, 2011:16).

Elasticity of trading volume with respect to transaction costs depends on the observation period. Thus in the UK, short-term elasticity is -0.5 and long term is -1.7 (Jackson and O'Donnell, 1985). For Sweden, Lindgren and Westlund (1990) determined the elasticity of trading volume with respect to total transaction costs in a range between -0.85 and -1.35.

In financial markets with fixed yields, research has suggested a much more pronounced market reaction to taxation. The introduction of a 0.2 to 3 basis points tax on bond trading in Sweden led to a considerable reduction in the volume of trade. Trading in long-term bonds, for which there is a large number of non-taxable alternatives, was reduced by 85% when the taxation was announced. Trading in short-term debt instruments was reduced by about 20% (Froot and Campbell, 1994).

Elasticity of trading volume on currency markets is estimated in general, with respect to transaction costs, at -0.4. This negative elasticity reflects a wide multilateral taxation base that reduces the possibility of tax evasion (Schmidt, 2007).

IMPACT ON VOLATILITY

Summers and Summers (1989) considered the impact on price volatility and established that there were positive impacts related to moderation of the instabilities caused by speculative trading because of the smaller amount to which funds were channelled to the financial sector.

Empirical research however confirms the much more complex link between the introduction of a tax and the appearance of short-term price volatility. Medhavan, Richardson and Roomans (1997) claim that price volatility can be explained with the use of four variables: public information, private information, transaction costs and market frictions. The impact of public information on volatility can range between 35 and 46%, of private information between 26 and 31%, of transaction costs between 22 and 35%, and of market frictions between 1 and 4% (Mehavan, Richardson and Roomans, 1997).

Matheson (2011:20) distinguishes short-term and long-term price volatility. By short-term, he means short-term market fluctuations in asset prices that do not deviate essentially from their fundamental value. Long-term and short-term volatility do not necessarily have to be correlated.

Studying the link between short-term price volatility and transaction costs in 23 countries, Roll (1989) found no casual link. Baltagi et al., (2006) also found no important impact of taxation on price volatility. Deregulation of the American stock market reduced transaction costs and price volatility (Jones and Seguin, 1997). Franch and Roll (1986) show that the actual process of trading causes short-term volatility and that the taxation of transactions, reducing the number of trades, can reduce price volatility. Habermeier and Kirilenko (2001) established that a tax increases price volatility and at the same time reduces the volume of trading on financial markets.

Since the tax has the same effect on well-informed investors, who reduce market volatility correcting asset values in the direction of their fundamental value and poorly informed or noise investors who increase price volatility, the net effect of taxation on volatility depends on the microstructure of a given market (De Long et al., 1989).

IMPACT ON LIQUIDITY

Preservation of the liquidity of markets is the primary task of all financial market regulators. Matheson (2011) particularly points out that a financial transactions tax reduces the volume of trading, which can also reduce liquidity and volatility. The influences of liquidity and volatility, then, are complementary to each other.

Although Oxera (2007) claims that the introduction of the UK stamp duty had negative impacts on liquidity, the UK is not considering abolishing this form of taxation, mainly from historical reasons. The stamp duty was the first tax on finan-

cial transactions and is also the oldest form of taxation in the country that is still being applied (Dieter, 2003). A study carried out with respect to Asian markets shows that a financial transactions tax can reduce liquidity as a result of the reduction of the number of financial transactions (Zhang, 2001).

Imposing a tax on stock exchange transactions can have various effects on liquidity in markets marked by information asymmetry. Subrahmanyan (1998) says that taxation of transactions will reduce liquidity on oligopolistic markets, but not on monopolist, where liquidity can be increased if a monopolist, a market maker, has information of which other traders are deprived, for taxation reduces information asymmetry on the market.

The tax rate has to be placed at a level low enough not to distort the market, while at the same time providing sufficient tax revenue.

IMPACT ON SHARE VALUE AND CAPITAL COSTS
Theoretical models confirm that higher transaction costs, including taxation costs, can lead to a reduction in the value of shares (Kupiecs, 1996). Investors have to bear larger costs when they buy shares, and will demand higher yields, which will tend to produce a reduction in share prices.

Higher transaction costs lead to an increase in the costs of capital of companies whose shares are being traded. A liquidity premium for shares can also be an important factor influencing evaluation. Block (2007) shows that retained shares of companies that are not traded are worth 20 to 25% less than comparable shares. Higher transaction costs, then, increase the cost of capital of companies that issue taxable securities.

It seems clear that because of taxation investors will be ready to keep shares longer and trade in them less, which can reduce their liquidity. Reduced liquidity thus reduces tax revenue, makes it more difficult to sell a financial instrument and ultimately reduces the value of the actual stock.

As for the effect of taxation on the cost of capital, Schwert and Seguin (1993) have estimated that the imposition of a 0.5% securities transaction tax in the USA would lead to an increase in the cost of capital of between 0.1 and 1.8 percentage points. Oxera (2007) estimates that repeal of the stamp tax in the UK would lead to a rise in share prices of 7.2% and a reduction of the cost of capital of between 0.66 and 0.8 percentage points.

IMPACT ON MARKET EFFICIENCY
In an efficient market, asset prices at once reflect all the available information. When new information arrives in an efficient market, investors adjust their assets according to the newly arriving information, which leads to changes in prices. If there are no transaction costs, this adjustment can go on incessantly, with differ-

ences in prices being pared down at the same time. Since for investors an FTT is a transaction cost, adjustment of prices will be slowed down or in other words a longer period will be required for the newly arriving information to be completely reflected in prices. In this way the taxation of financial transactions can reduce market efficiency (Habermeier and Kirilenko, 2001).

If taxation is considered from the viewpoint of its impact on market efficiency, it can be concluded that reduced transaction costs suit short-term trading in securities and derivatives. Many short-term trades are of a speculative nature based mainly on a technical analysis and hence lead to increased price volatility and the appearance of speculative bubbles. When transaction costs are increased, the number of short-term trades is reduced, which in turn reduces volatility and the possibility that securities will be incorrectly valued.

IMPACT ON TAX REVENUE

One of the main reasons for the introduction of an FTT is to increase tax revenues, which will depend on three parameters: the tax rate, the volume of trading weighted by the average price level and the number of transactions undertaken. When the rates are raised, then tax revenue will also rise; but at the same time the number of transactions effectuated and the volume of trading will fall, and the ultimate impact on tax revenue depends on the elasticity of the market to the increased transaction costs.

In order to reduce the tax burden, investors will endeavour to reduce their trading in short-term securities and will be oriented to long-term securities or to foreign capital markets with lower tax burdens.

Umlauf (1993) also mentions, in addition to the direct effects on tax revenues, that the introduction of an FTT would reduce income from capital gains tax.

TAX INCIDENCE

The greatest burden of the taxation would be shouldered by security holders, and since they are mainly individuals with high incomes, taxation can have a progressive effect (Matheson, 2011).

In the long term, the market will endeavour to equalise the rate of yield on capital in taxed and non-taxed markets. Because of the rise in the cost of capital, companies will try to finance their operations with non-taxable sources. Because of the smaller supply of taxed substitutes and the greater demand for non-taxed financing, yields on taxable sources of financing will fall and the yields on non-taxable will rise, until the yields even out. How much total investment will fall because of a rise in the cost of capital will depend on elasticity of supply and demand. In a small open economy return on capital depends on international market trends, and capital will drain out of a country until the yield after taxation is equal to the yield on the world market. In the long term, then, owners of capital will not bear the

burden of taxation entirely. Since supply of capital is not perfectly elastic in market conditions the yields on capital will fall. The final tax burden thus will be shared by owners of capital and by labour according to the relative elasticities of capital supply and demand (Matheson, 2011).

It can be concluded that taxation of securities transactions will reduce the volume of trading, the reduction depending on the tax rate, the specific features of a given market, the kind of securities, the existence of untaxed substitutes and taxation on competitive capital markets. The impacts of volatility and liquidity are complementary to each other, and the effect of taxation will depend on the microstructure of a given market. Taxation can reduce share value and increase the cost of capital. There is thus a dual effect of taxation on market efficiency. On the one hand taxation increases transaction costs and in this way slows down the adjustment of securities' prices, exerting pressures that reduce market efficiency; on the other hand, by reducing speculative trading, it can also increase market efficiency. Reduced volume of trading because of the reduction of investor profit will also tend to diminish revenue from the taxation of capital gains.

3 A COMMON TAX ON FINANCIAL TRANSACTIONS IN THE EUROPEAN UNION

According to the EC (2012) support of member states to the financial sector between October 2008 and December 31, 2011 came to 1.6 trillion euros, or about 13% of European GDP. Over 60% of this amount related to state guarantees to banks. The European Commission (2013c) estimated that direct benefits from the under-taxation of the financial system came at an annual level of about 18 billion euros (primarily in the context of VAT).

A common FTT should indirectly harmonise the taxation of the financial sector, essential for a proper functioning of the internal market, prevent damaging tax competition in the trading of financial instruments within the EU, and ensure that the financial sector makes a more equitable contribution to public finances.

An essential reason for the introduction of a common FTT is certainly the attempt to prevent suspect and speculative transactions and the provision of tax neutrality among countries, thus contributing to protection against tax evasion.

3.1 BASIC DETERMINANTS OF COMMON TAXATION OF FINANCIAL TRANSACTIONS IN THE EUROPEAN UNION

A common FTT would contribute significantly to the conditions of trading in financial instruments in the EU. Anticipated is the taxation of all financial institutions that trade in basic or derivative financial instruments, even outside organised markets. Transactions with original instruments would be taxed at a minimum rate of 0.1% on the amount of the transaction, and transactions in derivative instruments at a minimum rate of 0.01% on the notional value of the derivative contract. The notional amount of the transaction is considered the taxable amount. In order

to facilitate harmonisation of the application of the FTT, other forms of taxation of financial transactions would not be permitted. However, countries would be able to set their own rates, higher than the minimum, letting them provide additional sources of revenue. Accordingly, only transactions among financial institutions of the participating countries would be taxed. The proposal is based on the triple-A approach, meaning it refers to All markets, All instruments and All actors.

According to European Commission (2013b) the following would be considered financial institutions: investment companies, organised markets, credit institutions, insurance and reinsurance companies, investment funds and investment fund management companies, pensions funds and pensions funds management companies, holding companies, leasing companies, special purpose companies and other institutions as defined by statute. As well as these, other companies that undertake certain financial activities that exceed 50% of their annual income would also be considered financial institutions.

The FTT is primarily aimed at the taxation of financial transactions that are conducted by financial institutions, whether for their own account or the account of some other person. Transactions of the European Central Bank, the central banks of EU member states and competent bodies for the maintenance of financial system stability such as the European Financial Stability Facility and the European Stability Mechanism would be exempted from taxation.

Standard financial transactions of individuals and companies (making deposits, taking out loans, various forms of payment) would not be subject to taxation. The aim here is to protect the real sector against additional tax burdens. However, if individuals or business entities take up trading in securities or make derivative contracts via financial institutions, they will come within the scope of the FTT. Although natural persons and business entities of the real sector should not be considered liable to pay FTT, they could well end up bearing the economic burden of the taxation.

Although taxation would bypass the primary capital markets, the implications on the primary market itself are questionable. Will the economy be ready to finance its own operations with primary emission?

The implementation of an FTT would not be restricted to trading in organised markets, but would also take in over-the-counter markets, which would lead to an enlargement of the taxable amount. Transactions of larger amounts, because the regulatory requirements are not so exacting and because of the lower costs, are commonly conducted outside the organised markets. Also taxable would be repo and reverse repo operations, but only in the context of the initial exchange of instruments. The return transaction would not be subject to taxation (European Commission, 2103b).

According to the current proposal of the European Commission (2013b) structured products are explicitly subject to the FTT. The liability in derivatives trading

would be determined at the moment the derivative contract is made and at the time of the sale, transfer or replacement of derivative contracts, not taking cash flow into consideration, only the execution of contractual obligations. If it is generated, a transaction with an original financial instrument would be additionally taxed. It is necessary to point out that spot currency transactions are not subject to the FTT.

Derivative contracts that relate to trading in commodities (mostly gold, silver, precious metals or oil) without any real exchange of commodities would be subject to the payment of the FTT, for they are mostly speculative or aimed at the protection of assets without any real intention of physical assets being traded (European Commission, 2013b).

Investment and pension funds would be liable to pay the FTT in financial transactions in the management of the assets of the fund, but not when they are issuing their own securities (issue of certificates of the purchase or redemption of shares). Investment and pensions funds would be liable to pay the FTT in the scope of active trading.

It is planned to achieve the prevention of tax evasion, market disruptions and the redirection of transactions to a more favourable tax regime by a combination of the principle of residence as a basic principle of taxation with elements of the issuance principle. Such a regime will make it more difficult for taxpayers to relocate the place of their regular business activity, which will prevent tax evasion.

A financial institution that undertakes taxable financial transactions would be taxed according to the country of its residence, irrespective of the place of issuance of the actual transaction. If various financial institutions that have their place of residence in different member countries carry out financial transactions among themselves, each member state would have the right to withhold tax according to a set tax rate. The proposal of taxation according to European Commission based on double taxation in fact is not applicable nowadays in most European countries.

An FTT according to the proposal of the European Commission (2013b) could damage financial institutions, because taxation would diminish their competitiveness on the international market.

In the case of trading in securities carried out by non-residents, outside the jurisdiction of the FTT, issued in a country in which FTT is applied, the already mentioned residence principle is supplemented with the principle of country of issuance. Securities that are issued in a country where FTT is applied would be subject to the payment of the tax in the country of issuance, irrespective of where the transaction is conducted. If the parties in the transaction are not subject to the payment of the FTT according to the residence principle, the actual trading in securities issued in countries where FTT is applied would be taxable. Parties that trade in such securities would be considered liable to pay the tax in the country of issu-

ance, unless they can prove that there is no interrelation between the economic substance of the said financial transaction and the territory of the countries that do apply the FTT (European Commission, 2013b).

Such a provision does in fact reduce liquidity for all trading in such financial instruments will be taxable, which can have negative effects not only on investors but also on the issuers of securities. Proof of the lack of connection of a financial transaction with countries in which the FTT is applied is incumbent on the tax payers themselves.

3.2 THE MACROECONOMIC IMPACTS OF TAXATION
Most EU countries have important financial sectors, and their fiscal capacity is thus important.

TABLE 1
Revenue from the taxation of financial transaction in selected EU member countries, 2011

Member state	Taxation revenue	
	In million euros	As percentage of GDP
Belgium	132.0	0.04
Cyprus	1.4	0.01
Finland	249.0	0.13
France	1,100.0	0.06
Greece	92.0	0.04
Ireland	322.4	0.21
Luxembourg	605.0	1.41
Malta	2.1	0.03
United Kingdom	3,987.6	0.23

Source: European Commission (2013); Impact Assessment, pp. 62-68.

The figures in table 1 indicate how low the amount of tax collected is. Data from the European Commission (2013a) show that in 2011 various forms of the taxation of financial transactions including financial instruments existed in 11 member states. Although in 2011 there were some indications that financial transactions would be taxed in Italy, a new system of taxation was introduced in March 2013 for original securities and in July of the same year for derivatives (European Commission, 2016).

European Commission (2013a) predicts that the introduction of a common FTT would result in a fall of GDP of 0.28% over a period of 40 years. It predicts that if there is no FTT, that around 2050 European GDP would be 81.4% greater, and that it if is introduced, 81.1% greater than the present value. Some estimates say that the introduction of an FTT might lead to a long-term rise in GDP of at least 0.25% (Griffith-Jones and Persaud, 2012). However, the ultimate effect will depend primarily on the way in which the tax revenue is spent, that is, whether it will be channelled into boosting economic growth.

3.3 IMPACTS OF TAXATION ON INVESTOR BEHAVIOUR

The introduction of an FTT will increase transaction costs, which will lead to a change in the way in which investors act on the market, because trading in certain instruments might become more attractive.

An FTT might tend to bring about a reduction of short-term high-frequency trades that are on the whole based on algorithms. Since the tax will increase transaction costs and hence reduce the margin, certain short-term trading will not be represented any more to a sufficient extent, for example, high frequency trading and delta-hedging.

In addition, financial internalisation will, because of the increased transaction costs, probably be replaced with financial intermediation.[1] Increased transaction costs will thus lead to an increase in the effective tax burden and a reduction of the sales margin.

Certain changes can be looked for in the behaviour of financial institutions that can replace repo operations with overnight loans or short-term liquidity loans (European Commission, 2013a).

Trading in derivative contracts might, according to the estimates of the European Commission (2013a) experience a fall of as much as 75%, because investors will lose their interest in it. Taxation of derivative contracts will lead to an increase in the effective tax rate with an increase in the effect of scale. Since the proposed tax rate on derivative trading is several times small than the rate on basic instruments, it will be more acceptable to make contracts with a financial ratio of less than 1:10. With ratios of greater than 1:10, entering into a derivative contract will result in a rise in the effective tax rate, which can be clearly seen in the following case. If a derivative contract with a notional amount of 1 million euros is made with the use of a ratio of 1:100 the economic value of the financial transaction will come to 10,000 euros. Irrespective of the fact that 10,000 euros will be paid by the transaction, the taxable amount is still the notional amount of 1 million euros. With the application of a tax rate of 0.01% on the notional amount, a tax liability of 100 euros is incurred. But if the tax paid is expressed as a ratio of the transaction amount of 10,000 euros, the effective tax rate comes to 1%. Hence, more pronounced use of the effect of scale in derivative operations will lead to an increase in the effect tax rate.

Taxation would probably result in a change in management strategy in investment and pension funds, from active management to buy and hold strategies.[2] An active strategy would lead to greater taxation. The tax itself should not affect management strategies of money funds.

---

[1] Financial internalisation implies more successive purchases and sales of securities before ultimate sale to the end user, and financial intermediation means the creation of mediators between seller and buyer without trading for own account.

[2] An active management strategy implies vigorous changes in the structure of the fund, while buy and hold means that the portfolio is kept more or less the same.

Taking into account the very high quality protection mechanisms for preventing tax evasion with a combination of the principles of residence and issuance, it would not pay to shift the principal place of business of a financial institution outside the FTT jurisdiction as long as trading takes place with a counterparty located within the jurisdiction of the FTT or trading is done with instruments issued in a country in which FTT is applied. Payment of tax then can be avoided only by abandoning the market of the countries that apply the FTT, which is not very likely.

Because of the high fiscal and other aids to the financial sector, in the EU its fiscal contribution is being re-examined. By increased collaboration of member states the conditions might be created for easier harmonisation of taxation of the financial sector.

4 THE POSSIBILITIES FOR TAXING FINANCIAL TRANSACTIONS IN CROATIA

Although the FTT is not of recent date, and its application in various forms is widespread in the countries of the EU and elsewhere, the tax policy of the Republic of Croatia has avoided it, out of historical reasons and also in an attempt to increase the competitiveness of the domestic capital market.

Taking into account the stability of financial institutions, along with the regulatory measures of the Croatian National Bank (CNB), state interventions in the financial sector have not been necessary, on the contrary, in the crisis times, it made ample contributions to the public finances. For all these reasons, there has been little discussion of the introduction of an FTT.

Recent economic trends have reduced the attractiveness of investing in the financial markets. Regular turnover on the Zagreb Stock Exchange fell by 35% in 2012, trading in shares dropped by 44%, while because of investor caution, the turnover in bonds was more than twice that of 2011. In September 2012, trading in structured products started, including various forms of certificates (index, bonus, turbo certificates). Trading in these products was counted into regular trading (ZSE, 2013). The Zagreb Stock Exchange does not carry out trading in derivatives.

Since regulatory systems are an important component of contemporary stock exchange dealings for the sake promptly spotting and preventing manipulative trading patterns, early in 2015 the Zagreb Stock Exchange introduced a new system for the supervision of market segments – organised market, multilateral trading platforms and OTC markets (ZSE, 2016).

In 2013 there were no important changes in the total turnover, but it can be pointed out that there was a turnover 42 times greater in commercial bills, and a trade in structured products that was twice as high as in 2012 (ZSE, 2014).

In 2014 there was no significant rise in regular turnover in the total amount. Trade in shares, which accounted for almost 80% of overall turnover, remained at the

2013 level, while trade in bonds was increased by 65%, and structured product turnover fell by 44% (ZSE, 2015).

In 2015, regular trading was reduced by 9.3%, and OTC trading by 17.8%, which suggests that the capital market became less attractive to investors. Still there was a rise in the tourist industry segment (the CROBEXturist® index rose by 23.7% from 2014) (ZSE, 2016).

TABLE 2

Turnover on the Zagreb Stock Exchange 2011-2015 (in billion kuna)

	2011	2012	2013	2014	2015	Average	
						Value	Change (%)
Regular trading	5.9	3.9	3.8	3.9	3.5	4.2	-12.2
OTC transactions	18.2	19.3	18.2	26.6	21.9	20.8	4.7
Total	24.1	23.2	22.0	30.5	25.4	25.0	1.3

Source: Zagreb Stock Exchange (2013, 2014, 2015, 2016), author's elaboration.

Table 2 shows that in the 2011-2015 period the average value of a OTC transactions was almost five times the size of average regular trading on the Zagreb Stock Exchange.

Taking into consideration as taxable amount the average value of all transactions (regular trading and OTC transactions) the revenues from a possible introduction of an FTT might be simulated. On the assumption that both parties in a financial transaction are liable to the tax and that they are Republic of Croatia residents, the taxable amount (an average of 25 billion kuna) is doubled, and comes to 50 billion. If the minimum tax rate of 0.1% is applied, the tax revenue will come to 50 million kuna. For the sake of a comparison of the importance of these revenues from taxation in total revenues and in the tax revenues of the government budget, table 3 shows total tax revenues of the budget in 2015.

TABLE 3

Revenue of the government budget in 2015 (in billion kuna)

Kind of revenue	Amount of revenue
Tax revenues	68.0
Contributions	22.8
Grants	5.0
Revenue from assets	2.8
Revenue from administrative fees	3.6
Other revenue	6.9
Total revenue	109.1

Source: Ministry of Finance (2016).

If, then, the simulated revenue from FTT is compared with the figures in table 3, in 2015 it would have come to no more than 0.05% of total revenues of the government budget.

It can be concluded, then, that by taxing transactions on the Croatian capital market would be generated relatively small amounts of tax revenue and that the taxation of turnover in securities on the Zagreb Stock Exchange would be insufficiently effective. If one takes into consideration the essential administrative costs involved in the collection of the tax, although they should not be very great because of the well-arranged computer infrastructure, the actual amount of tax collected would not justify it being employed in practice. This form of taxation, then, would be insufficiently productive. If the possible impacts on the volume of trading, liquidity, cost of capital and market efficiency are considered, then there are no practical grounds for the taxation of the Croatian capital market.

However, when the fact that according to the EU-level proposal for the taxation of the transactions of financial institutions primarily according to the principle of residence is taken into consideration, the real tax revenues could be very different from those of the simulation. Transactions on the Zagreb Stock Exchange are conducted by residents and non-residents. In the Republic of Croatia, only the transactions of residents would be taxed, irrespective of the place in which the transactions are carried out. Real tax revenues would then include all the transactions of residents and the transactions of non-residents to the extent that they are trading in financial instruments issued in the Republic of Croatia and are not at the same time residents of some EU member state in which a common FTT would be in application. Hence all such transactions need simulating.

Tax revenues would certainly cover transactions in derivative contracts among financial institutions. According to projections of the European Commission (2013d) revenue from the taxation of derivative contracts would come to about two thirds of all revenue from an FTT.

A preliminary analysis of the possibilities of taxing financial transactions in the Republic of Croatia according to the proposal to introduce an FTT at the EU11 level suggests that the Croatian capital market is still not sufficiently developed to be able to generate abundant revenues from such a tax.

If domestic financial institutions carry out taxable financial transactions with financial institutions in the EU11 or trade in financial instruments issued in these countries, they would be obliged to pay FTT in the country of residence of the counterparty or the country of issuance of the security, which would lead to a drain of tax revenue outside the borders of the Republic of Croatia. Hence increased collaboration of the Republic of Croatia concerning the issue of taxing financial transactions would be justified in order to ensure that Republic of Croatia residents pay tax in their own country.

5 CONCLUSION

One of the possible solutions for a more equitable contribution of the financial sector to public finances is the reinforcement of the role of indirect taxation of financial transactions. The global trend toward the reduction of the taxation of financial transactions in the last two decades because of the liberalisation of capital flows has been replaced with a more powerful role for the state in the regulation of financial markets. Although taxation of financial transactions is not important in overall public revenues, developed world economies are not about to give up on this form of taxation, but want to increase its fiscal role.

If one considers the taxation from the standpoint of its impact on financial markets, an FTT will certainly increase transaction costs. Numerous empirical investigations about the impact of an FTT on the function of financial markets have led to various conflicting conclusions, largely related to the microstructure of a given market. There is a consensus that taxation will reduce the number of financial transactions generated as well as the value of financial instruments, which will bring about an increase in the cost of capital. There are opposed ideas about the effect on liquidity and volatility. Tax revenues, then, depend considerably on the reaction of the market to increased transaction costs consequent upon the introduction of the tax.

Expensive government rescue operations have forced the European Commission to re-examine the fiscal contribution made by the financial sector; but because of the different national systems of taxing financial transactions, there is a lack of harmonisation. With the establishment of the enhanced cooperation system in the EU11, conditions can be created for an easier establishment of a consensus among all the members of the EU and greater harmonisation. The actual execution of common taxation has a broad tax base since it provides for the taxation of all transactions in basic and derivative instruments both inside and outside organised markets. With a combination of the principles of residence and issuance, it is possible to have good mechanisms of protection against tax evasion. The future of common taxation at EU level will depend on the wiliness of all members to find a high quality, effective and generally accepted manner of taxing financial transactions.

A preliminary analysis of the possibilities of taxing the Croatian capital market has shown that there are no practical grounds for the implementation of such a tax, because the insufficiently developed market would be incapable of generating sufficiently abundant tax revenues. However, if financial transactions are not taxed in the Republic of Croatia, Croatian residents might be taxed in EU member states that do apply this tax, which would lead to an outflow of tax revenues, and so it would be justified for the Republic of Croatia to join in the procedure of enhanced cooperation to do with the FTT. It can, then, be concluded that the common taxation of financial transactions at the EU level could meet the interests of the Republic of Croatia better than unilateral taxation.

The costs and benefits of the taxation and also of the non-taxation of financial transactions in the Republic of Croatia will be able to be seen better when the FTT is in the phase of practical implementation in the EU. Then it will be certainly possible to carry out a better and more trustworthy analysis of the effects of the taxation.

REFERENCES

1. Baltagi, B., Li, D. and Li, Q., 2006. Transaction Tax and Stock Market Behavior: Evidence from an Emerging Market. *Empirical Economics,* 31, pp. 393-408. doi: 10.1007/s00181-005-0022-9

2. Block, S., 2007. The Liquidity Discount in Valuing Privately Owned Companies. *Journal of Applied Finance,* 17(2), pp. 33-40.

3. De Long, B. [et al.], 1989. Positive Feedback Investment Strategies and Destabilizing Rational Speculation. *NBER Working Paper,* No. 2880.

4. Dieter, H., 2003. Reshaping globalisation: a new order for international financial markets. *Centre for the Study of Globalisation and Regionalisation Working paper,* No. 103/02.

5. European Commission, 2012. *State aid: crisis-related aid aside, Scoreboard shows continued trend towards less and better targeted aid*. Available at: [http://europa.eu/rapid/press-release_IP-12-1444_en.htm].

6. European Commission, 2013a. *Commission staff working document – Impact assessment*. Available at: [http://ec.europa.eu/taxation_customs/resources/documents/taxation/swd_2013_28_en.pdf].

7. European Commission, 2013b. *Proposal for a Council Directive implementing enhanced cooperation in the area of financial transaction tax*. Available at: [http://ec.europa.eu/taxation_customs/resources/documents/taxation/com_2013_71_en.pdf].

8. European Commission, 2013c. *Financial Transaction Tax throught Enhanced Cooperation: Questions and Answers*. Available at: [http://europa.eu/rapid/press-release_MEMO-13-98_en.htm].

9. European Commission, 2013d. *Taxation of the financial sector*. Available at: [http://ec.europa.eu/taxation_customs/taxation/other_taxes/financial_sector/].

10. European Commission, 2016. *Financial Transaction Taxes in the European Union*. Available at: [http://ec.europa.eu/taxation_customs/resources/documents/taxation/gen_info/economic_analysis/tax_papers/taxation_paper_62.pdf].

11. European Parlament, 2010. *Financial Transaction Tax: Small is Beautiful*, IP/A/ECON/NT/2008-08, PE 429.989.

12. French, K. and Roll, R., 1986. Stock Return Variances: The Arrival of Information and the Reaction of Traders. *Journal of Financial Economics*, 17, pp. 5-26. doi: 10.1016/0304-405X(86)90004-8

13. Froot, A. K. and Campbell, J. Y., 1994. "International Experiences with Securities Transaction Taxes" In: Frankel, J. A., ed. *The Internationalization of Equity Markets*. University of Chicago Press, pp. 277-308.

14. Griffith-Jones, S. and Persaud, A., 2012. *Finantial Transaction Taxes*. Available at: [http://www.europarl.europa.eu/document/activities/cont/201202/20120208ATT37596/20120208ATT37596EN.pdf].

15. Habermeier, K. and Kirilenko, A., 2001. Securities Transaction Taxes and Financial Markets. *International Monetary Fund Working Paper* 01/51.

16. IMF, 2010. *Financial Sector Taxation - The IMF's Report to the G-20 and Background Material.* Washington: IMF.

17. Jackson, P. D. and O'Donnell A. T., 1985. The effects of stamp duty on equity transactions and prices in UK Stock Exchange. *Working paper*. London: Bank of England.

18. Jones, C. and Seguin, P., 1997. Transactions Costs and Price Volatility: Evidence from Commission Deregulation. *American Economic Review,* 87(4), pp. 728-737.

19. Kupiecs, P., 1996. Noise Traders, Excess Volatility, and a Securities Transaction Tax. *Journal of Financial Services Research,* 10, pp. 115-129. doi: 10.1007/BF00115671

20. Lindgren, R. and Westlund, A., 1990. How did the transaction costs on the Stockholm Stock Exchange influence trading volume and price volatility? *Skandinaviska Enskilda Banken Quarterly Review*, (2), pp. 30-35.

21. Madhavan, A., Richadson M. and Roomans, M., 1997. Why Do Securities Prices Change? A Transaction-Level Analysis of NYSE Stocks. *Review of Financial Studies*, 10, pp. 1035-1064. doi: 10.1093/rfs/10.4.1035

22. Matheson, T., 2011. Taxing Financial Transaction: Issues and evidence. *International Monetary Fund Working Paper* 11/54.

23. Ministry of Finance of the Republic of Croatia, 2016. *Račun prihoda i rashoda za 2015. godinu*. Available at: [http://www.mfin.hr/hr/drzavni-proracun-2015--godina].

24. Oxera, 2007. *Stamp Duty: Its Impact and the Benefits of its Abolition*. Report prepared for Association of British Insurers. London: City of London Corporation, Investment Management Association and London Stock Exchange.

25. Roll, R., 1989. Price Volatility, International Market Links and Their Implication for Regulatory Policies. *Journal of Financial Services Research,* 3(2-3), pp. 211-246. doi: 10.1007/BF00122803

26. Schmidt, R., 2007. *The Currency Transaction Tax: Rate and Revenue Estimates*. Ottawa: North-South Institute.

27. Schwert, G. W. and Seguin,P., 1993. Securities Transaction Taxes: An Overview of Costs, Benefits and Unresolved Questions. *Financial Analysts Journal*, (Sep–Oct), pp. 27-35. doi: 10.2469/faj.v49.n5.27

28. Subrahmanyam, A., 1998. Transaction Taxes and Financial Market Equilibrium. *Journal of Business,* 71, pp. 81-117. doi: 10.1086/209737

29. Summers, L. and Summers, V., 1989. When Financial Markets Work Too Well: A Cautious Case for a Securities Transaction Tax. *Journal of Financial Services Research,* 3, pp. 261-286. doi: 10.1007/BF00122806

30. Umlauf, S., 1993. Transaction Taxes and the Behavior of the Swedish Stock Market. *Journal of Financial Economics,* 33, pp. 227-240. doi: 10.1016/0304--405X(93)90005-V

31. Zagreb Stock Exchange, 2013. *Pregled trgovine u 2012. godini.* Available at: [http://www.zse.hr/UserDocsImages/reports/ZSE-2012.pdf].

32. Zagreb Stock Exchange, 2014. *Pregled trgovine u 2013. godini*. Available at: [http://zse.hr/UserDocsImages/reports/ZSE-2013.pdf].

33. Zagreb Stock Exchange, 2015. *Pregled trgovine u 2014. godini*. Available at: [http://zse.hr/UserDocsImages/reports/ZSE-2014.pdf].

34. Zagreb Stock Exchange, 2016. *Pregled trgovine u 2015. godini*. Available at: [http://zse.hr/UserDocsImages/reports/ZSE-2015.pdf].

35. Zhang, L., 2001. *The Impact of Transaction Tax on Stock Markets: Evidence from an emerging market*. East Carolina University, Department of Economics.

Using the age-based insurance eligibility criterion to estimate moral hazard in medical care consumption

YAN ZHENG, PhD*
TOMISLAV VUKINA, PhD*

Yan ZHENG
Credit Analyst, JP Morgan Chase and Co. 201 N. Walnut Street, Wilmington, DE 19801, USA
e-mail: zoeyyzheng@gmail.com

Tomislav VUKINA
Professor, Department of Agricultural and Resource Economics, North Carolina State University,
3332 Nelson Hall, Raleigh, NC 27695-8109, USA
e-mail: tom_vukina@ncsu.edu

FINANCIAL THEORY AND
PRACTICE
40 (3) 337-356 (2016)

YAN ZHENG AND TOMISLAV VUKINA:
USING THE AGE-BASED INSURANCE ELIGIBILITY CRITERION TO ESTIMATE MORAL HAZARD
IN MEDICAL CARE CONSUMPTION

Abstract

This paper uses fuzzy regression discontinuity design to estimate the moral hazard effect in health care consumption in the population of young adults. We use invoice data for outpatient hospital services from a regional hospital in Croatia. The estimation is complicated by the fact that the data set consists only of users of medical services, which would tend to underestimate the moral hazard effect. To address this issue we use a modified version of the instrumental variables approach. We find a 92% reduction in the number of hospital visits for individuals who lost insurance coverage when crossing the 18th birthday threshold.

Keywords: fuzzy regression discontinuity design, instrumental variables, health insurance

1 INTRODUCTION

A large body of health economics literature documents a strong association between health insurance status and patterns of health care utilization. People with more generous insurance coverage tend to consume more health care, a phenomenon known as moral hazard. The literature suggests that when individuals lose health insurance, they change their consumption of health care services, i.e. they seek medical attention less frequently. But would the uninsured consume more health care if they had health insurance? Such a causal inference is difficult because uninsured individuals are likely to have different health conditions, attitudes towards risk, disposable income, wealth, and so on than insured individuals.

To overcome this problem, we exploit quasi-experimental variation in insurance status resulting from the rules insurance companies use to establish the coverage eligibility of dependents. One of the largest segments of the population that lacks health insurance is young adults (age 19 to 29). For example, in the United States, 29% of the uninsured are young adults (Schwartz and Schwartz, 2008). Because many health insurance contracts cover dependents until the age of 18 and only cover older dependents if they are full time students, a significant number of teenagers become uninsured after they reach the threshold birthday. Because young adults are relatively healthier than the older population, it is reasonable to assume that they would consume relatively low levels of health care regardless of whether they are insured or uninsured and that the social cost of extending coverage to uninsured young adults would be relatively modest.

The use of terms moral hazard has its origins in the insurance literature and has subsequently spread into contract theory and information economics. Contract theory refers to moral hazard as an asymmetric information problem arising when an agent's (insured's) behavior is not observable by the principal (insurance company). The magnitude of moral hazard is then measured by the welfare difference between the first best (symmetric information) and the second-best (asymmetric information) outcome. In the context of health insurance, however, moral hazard is often used in reference to the price elasticity of demand for health care, condi-

tional on underlying health status (Paully, 1968; Cutler and Zeckhauser, 2000; Einav et al., 2013). The approach that we use in this paper is conceptually in line with this mainstream health insurance literature. In other words, our approach does not consider the potential impact of insurance on underlying health or its impact on risky behavior that may have detrimental effects on health, as for example in Yörük (2015). Instead, moral hazard is simply the consequence of the demand function for health care being less than perfectly inelastic (vertical) such that any reduction in the price of health care services resulting from owning an insurance policy increases the quantity demanded. The measure of moral hazard is then simply the difference in health care consumption between the full price of medical services and the lower price one pays as the owner of an insurance policy, conditional on health status.

The empirical evidence on the importance of moral hazard in health insurance markets varies by country, type of health service provided and socio-economic status of insured. For example, using Australian data, Cameron et al. (1988) established that more generous coverage leads to higher utilization of a broad range of services. They found a significant price effect in health care consumption, which implies that moral hazard is an important determinant of overall health care utilization. Also for Australia, Savage and Wright (2003) found that after correcting for endogeneity, the extent of moral hazard can increase the expected length of a hospital stay up to three times.

Liu, Nestic and Vukina (2012), using matching estimators, found the presence of a significant moral hazard effect in the Croatian health care system. In the case of Spain, Vera-Hernandez (1999) found no evidence of moral hazard for heads-of-households and strong evidence for other household members. Holly et al. (1998), using data for Switzerland, found evidence of moral hazard in hospital stays. Coulson et al. (1995) found that supplemental insurance increases the number of prescriptions filled by the elderly in the United States. Manning et al. (1987) used a randomized experiment and found that a catastrophic insurance plan reduces expenditures 31% relative to zero out-of-pocket price, indicating a large moral hazard effect. Cardon and Hendel (2001) integrated health insurance and health care demand using 1987 National Medical Expenditure Survey data and found that the gap in expenditure between insured and uninsured can be attributed to observable demographic differences and to price sensitivity. They interpreted the elasticity of demand with respect to price (coinsurance rate) as evidence of moral hazard.

In this paper, we rely on the fuzzy regression discontinuity (RD) design to estimate the effect of losing insurance on health care consumption in young adults. We use invoice data for outpatient hospital services from a regional hospital in Croatia. Croatia has a state-run health insurance system dominated by a single public health insurance fund, the Croatian Institute for Health Insurance (HZZO). The HZZO offers compulsory and supplemental insurance. The former covers some medical care services fully and some subject to co-payments. Full coverage

FINANCIAL THEORY AND PRACTICE
40 (3) 337-356 (2016)

YAN ZHENG AND TOMISLAV VUKINA:
USING THE AGE-BASED INSURANCE ELIGIBILITY CRITERION TO ESTIMATE MORAL HAZARD
IN MEDICAL CARE CONSUMPTION

of medical care services is provided to children younger than 18 and to all patients suffering from specific serious illnesses. All other health services are subject to co-payments. In order to avoid these co-payments, supplemental insurance is required. The compulsory insurance coverage is universal whereas supplemental insurance can be either bought or is extended automatically free of charge to some categories of citizens such as, for example, full time students.

In this context, the 18th birthday represents a threshold for supplemental insurance coverage; young adults crossing the threshold will lose full coverage and face three options: stay in school and continue being fully covered for free, buy supplemental insurance and continue enjoying full coverage or refuse to buy the supplemental coverage and pay co-payments as required.

Since crossing the 18th birthday threshold is not a unique determinant of assignment into the treatment (losing the coverage), the problem fits into the fuzzy RD design. RD design can be used to determine the treatment effects in quasi-experimental settings where treatment is determined by a forcing variable exceeding the threshold. The probability of being treated at the threshold jumps from zero to one (see: Lee and Lemieux, 2010). In the fuzzy RD design, the forcing variable does not exclusively determine the treatment assignment, hence the discontinuity in the probability of being treated at the threshold is less than one.

Regression discontinuity design has been used extensively to determine the causal effect from an intervention. For example, Card, Dobkin, and Maestas (2009) compared health care consumption among people just before and just after the age of 65, the threshold for Medicare eligibility in the U.S., in a sharp RD design. They found that the Medicare eligibility causes a discontinuous increase in health care utilization. Van der Klaauw (2002) identified the causal effect of financial aid on college enrollment decisions using fuzzy RD. He found that enrollment rate increased by about 0.2 at the financial aid eligibility threshold. Chen and van der Klaauw (2008) evaluated the effect of disability insurance program on labor force participation among disability insurance beneficiaries using fuzzy RD design. They found that the labor force participation rate among those beneficiaries would have been 20% higher if none had received the benefits. Anderson, Dobkin, Gross (2012) used the age 19 as an instrument to identify the causal effect of loss of health insurance at age 19 on health care consumption. They found that not having insurance leads to a decreasing level of health care consumption, manifested in a 40 percent reduction in emergency department visits and a 61 percent reduction in inpatient hospital admissions.

A distinct feature of our data set which considerably complicates the estimation is that it consists of users only. People that did not use the medical services of the hospital during the time period covered by the data do not show up in the data. The problem of estimating the moral hazard effect with users-only data is caused by the fact that some sick people do not seek medical attention at all or seek it less

often precisely because they do not have the insurance. Using users-only (patients) data clearly underestimates the moral hazard effect. To deal with this attenuation bias we use the instrumental variables (IV) approach from Anderson, Dobkin and Gross (2012), which relies on the assumption that the net change in the observed hospital visits after the age threshold of 18 is driven only by individuals who have lost their insurance coverage, an assumption implied by the standard IV exclusion restriction. We found a statistically significant reduction in the number of hospital visits by 92% for young adults who had lost their supplemental insurance after turning 18, confirming the moral hazard hypothesis.

2 INSTITUTIONAL FRAMEWORK AND DATA DESCRIPTION

The healthcare system in Croatia is still largely dominated by the institutional setup inherited from socialism. Despite the fact that the generous benefits and exemptions inherited from the old system have been politically difficult to roll back, Croatia has embarked on a number of reform initiatives that resulted in a relative decline in total spending on health care. Reforms have included enlarging the participation scheme (co-payments), reducing the number of individuals exempt from participation, introduction of administrative fees, and some cost savings in prescription drugs expenditures (Liu, Nestic and Vukina, 2012). Croatia still spends 7.9% of its GDP on health, among the highest for new EU members. In a fiscally constrained environment, the Croatian health system faces a mismatch between declining available public resources, growing expenditures and the increasing needs of an ageing population (Word Bank, 2015).

The main characteristics of the system provided by the HZZO can be summarized as follows. The compulsory insurance, which is funded by a 15% payroll tax, covers two kinds of medical care services: one with full coverage and the other with a system of co-payments. Full coverage medical care services are provided to children up to 18 years of age, pregnant women and for everybody else for life-threatening types of conditions such as infectious diseases, psychiatric care, surgeries, cancers and mandatory vaccinations, all other health services (including but not limited to primary care, hospitals stays and prescription drugs) are subject to a system of co-payments. The patients are required to pay 20% of the full price of medical care, with the largest out-of-pocket cost share amount set at 3,000.00 HRK per invoice.[1] Supplemental insurance is voluntary and can be acquired by a person 18 years or older who has compulsory insurance, by signing a contract with the HZZO. Certain categories of citizens are entitled to the supplemental insurance free of charge, i.e. their premiums are covered from the state budget. The list of people entitled to free supplemental insurance includes full time secondary school and college students. For those not entitled to free supplemental coverage, the premiums range from 50 to 130 HRK per month depending on income and whether the insured is active or retired. A person having the supplemental insur-

---

[1] The figures are reflective of the year 2009 which is the year covered by the dataset. The exchange rate for the local currency, Croatian kuna (HRK), as of June 20, 2009 was 1USD=5.19 HRK. A more detailed description of the health insurance system in Croatia can be found in Liu, Nestic and Vukina (2012).

FINANCIAL THEORY AND PRACTICE
40 (3) 337-356 (2016)

YAN ZHENG AND TOMISLAV VUKINA:
USING THE AGE-BASED INSURANCE ELIGIBILITY CRITERION TO ESTIMATE MORAL HAZARD
IN MEDICAL CARE CONSUMPTION

ance is entitled to full waiver of all the medical expense co-payments mentioned before.

The original data set consists of all invoices for all outpatient services from a small hospital in Croatia during the period from March 1 to June 30, 2009.[2] The data set consists of 105,646 observations. Each observation reflects one hospital visit (invoice). The data contains the following set of variables: a numeric code for the type of hospital service provided, compulsory health insurance number, supplemental insurance number (if the patient has one), period covered by the supplemental insurance, numeric code for categories entitled to supplemental insurance free of charge, eligibility category for compulsory insurance, cost of hospital service, part of the cost covered by compulsory insurance, part of cost covered by supplemental insurance, part of cost covered by participation (co-payment), date of birth and sex of the patient. The invoices do not record the exact date when the patient visited the hospital, but are chronologically ordered. To determine the date of the visit, we first divided all invoices into 122 days (March 1 to June 30) and designated the first 866 invoices as March 1st invoices, the next 866 invoices as March 2nd invoices, etc. This is a fairly innocuous simplification because the actual number of daily appointments is determined by the hospital's outpatient capacity and it is, therefore, reasonable to assume that the average number of patients treated each day is approximately the same. Since we know the date of birth of each patient and knowing the day of the visit, we can calculate the number of days each patient was away from the 18th birthday when they visited the hospital. For patients younger than 18, the days away from the 18th birthday are recorded as negative numbers and for patients older than 18 as positive numbers.

3 REGRESSION DISCONTINUITY DESIGN

In the basic setting for the regression discontinuity design researchers are interested in the causal effect of a binary intervention or treatment. Individuals are either exposed or not exposed to a treatment. Let $W_i Î \{0, 1\}$ denote a treatment with $W_i = 1$ if unit i was exposed to the treatment and $W_i = 0$ otherwise. Let $Y_i(W_i)$ denote the outcome where $Y_i(0)$ is the outcome without the exposure to treatment and $Y_i(1)$ is the outcome with exposure to treatment. We are interested in the difference $Y_i(1) - Y_i(0)$. Since one can never observe $Y_i(0)$ and $Y_i(1)$ at the same time, one needs to use the average effects of the treatment over the population or a sample to estimate the average treatment effect.

It is important to distinguish two RD settings: sharp and fuzzy regression discontinuity designs. Based on Imbens and Lemieux (2008), in the sharp regression discontinuity (SRD) design, the assignment W_i is a deterministic function of one

---

[2] The hospital in question is one of the general medicine regional hospitals in Croatia which provides all typical services such as internal medicine, surgery, anesthesiology, pediatrics, psychiatry, neurology, obstetrics and gynecology, radiology, pathology and emergency care at a fairly standard (basic) level. All complicated cases are typically referred to one of the larger medical centers in the capital Zagreb. The name and the location of the hospital are suppressed for confidentiality reasons.

of the covariates, i.e. the forcing (or treatment-determining) variable X such that $W_i = 1$ if $X_i \geq$ c and $W_i = 0$ if $X_i <$ c, where c is some constant. Therefore, all units with the value of the forcing variable $X_i = x$ at least c are assigned to the treatment group and participation is mandatory whereas all units with the value of forcing variable $X_i = x$ less than c are assigned to the control group. In the SRD design, the interest is to use the discontinuity in the conditional expectation of the outcome given the forcing variable in order to estimate the average treatment effect:

$$\lim_{x \downarrow c} E[Y_i | X_i = x] - \lim_{x \uparrow c} E[Y_i | X_i = x], \tag{1}$$

which is interpreted as the average causal effect of the treatment:

$$\tau_{SRD} = E[Y_i(1) - Y_i(0) | X_i = c] \tag{2}$$

at the discontinuity point.

In the fuzzy regression discontinuity (FRD) design, however, the probability of being treated does not need to jump from zero to one at the threshold. The design allows other factors to influence the assignment to treatment, besides the forcing variable. Therefore, only a jump in the probability of assignment to the treatment at the threshold is required:

$$\lim_{x \downarrow c} \Pr(W_i = 1 | X_i = x) \neq \lim_{x \uparrow c} \Pr(W_i = 1 | X_i = x). \tag{3}$$

The average causal effect of the treatment:

$$\tau_{FRD} = \frac{\lim_{x \downarrow c} E[Y | X = x] - \lim_{x \uparrow c} E[Y | X = x]}{\lim_{x \downarrow c} E[W | X = x] - \lim_{x \uparrow c} E[W | X = x]} \tag{4}$$

is identified as the ratio of the jump in the outcome at the threshold over the jump in the treatment indicator at the threshold. In our particular case expression (4) measures the average causal effect on health care utilization (as measured by the number of hospital visits) of losing insurance as the result of turning 18.

3.1 INSTRUMENTAL VARIABLE APPROACH
An important reason to choose FRD instead of some other causal inference methods is that it naturally solves the endogeneity problem by using the discontinuity as an instrument (Lee and Lemieux, 2010). To estimate the ratio in equation (4) we follow the approach of Anderson, Dobkin and Gross (2012). To explain the procedure, consider the following reduced form model of the effect of health insurance coverage on health care utilization:

$$Y_i = \gamma_0 + \gamma_1 D_i + \varepsilon_i, \tag{5}$$

where Y_i indicates health care utilization of individual i, D_i is an indicator variable equal to 1 if individual i has health insurance and 0 otherwise and γ_0 and γ_1 are coef-

FINANCIAL THEORY AND PRACTICE 40 (3) 337-356 (2016)

YAN ZHENG AND TOMISLAV VUKINA: USING THE AGE-BASED INSURANCE ELIGIBILITY CRITERION TO ESTIMATE MORAL HAZARD IN MEDICAL CARE CONSUMPTION

ficients to be estimated. All other determinants of health care utilization are summarized in the error term ε_i. Therefore, the coefficient γ_1 denotes the causal effect of health insurance on health care utilization, also known as the moral hazard effect.

However, this moral hazard effect is contaminated by other factors because the insurance coverage variable D_i is correlated with other determinants of health care utilization in ε_i, for example, the unobserved health status of that individual. An individual chooses to acquire health insurance based on idiosyncracies that simultaneously affect the decision to be insured and the consumption of health care. Therefore the estimate for γ_1 is inconsistent due to endogeneity. To solve the problem one typically relies on the IV approach. The objective is to estimate the causal effect of losing health insurance coverage on the number of hospital visits. Using crossing the 18th birthday threshold as an instrument, we estimate the first stage equation – the share of young adults who lose the insurance coverage while crossing the age 18 threshold and the reduced form equation – the change in the number of hospital visits associated with attaining the age of eighteen. We identify the effect of the moral hazard (coefficient γ_1) by dividing the reduced form estimate (the effect of turning 18 on the number of visits) by the first stage estimate (the effect of turning 18 on health insurance coverage).[3]

Notice however, that having hospital invoices data introduces sample selection bias in the first stage estimation because we only observe the insurance status for individuals who do show up in the hospital and do consume some services for which they or their insurer is charged. Regression estimate of the change in the proportion of uninsured after crossing the age 18 threshold understates the true size of this change. Because loss of insurance reduces the likelihood of a hospital visit and therefore affects the probability of appearing in the sample, this selection mechanism leads to an attenuation bias when estimating the change in the insurance coverage because newly uninsured individuals are more likely to leave the sample.

To correct for the bias in the first-stage estimates we assume that the net change in the observed hospital visits after crossing the age 18 threshold is driven only by individuals who have lost insurance coverage. This assumption is implied by the standard IV exclusion restriction and is quite reasonable. Because all our patients in the *Y-18* group are fully insured, those who did not go to the hospital before 18 (when they had insurance) and hence did not show up in the data as users, are unlikely to visit hospital immediately after turning 18 even if they have insurance and even less likely if they don't have insurance.

Let D_i denote the insurance coverage indicator and $A_i = a$ denote the age of an individual. Then the effect of crossing the age 18 threshold on insurance coverage at the population level can be expressed as:

---

[3] This strategy is analogous to using the age 18 discontinuity as an instrument to identity the causal effect of health insurance; see Hahn, Todd and van der Klaauw (2001).

$$\lim_{a\downarrow 18} E[D_i | A_i = a] - \lim_{a\uparrow 18} E[D_i | A_i = a], \tag{6}$$

Next, let $D_i(1) = 1$ indicate an individual older than 18 with insurance coverage and $D_i(1) = 0$ an individual older than 18 without insurance. $D_i(0) = 1$ and $D_i(0) = 0$ are defined similarly for individuals younger than 18. Also, let $Y_i(1) = 1$ denote an individual older than 18 who did visit the hospital and $Y_i(1) = 0$ an individual older than 18 who did not visit the hospital. $Y_i(0) = 1$ and $Y_i(0) = 0$ are defined similarly for individuals younger than 18. Since we could only observe individuals who visited the hospital (users), the effect of turning 18 on insurance coverage among users is estimated as:

$$E[D_i(1) | Y_i(1) = 1] - E[D_i(0) | Y_i(0) = 1]. \tag{7}$$

However, we would like to estimate the effect of turning 18 on insurance coverage at the population level, given that they visited the hospital when they were younger than 18:

$$E[D_i(1) | Y_i(0) = 1] - E[D_i(0) | Y_i(0) = 1]. \tag{8}$$

The desired effect is estimated as follows. First we denote the number of visits made before 18 as y_o, the number of insured visits made before 18 as d_o, the number of visits made after 18 as y_1 and the number of insured visits made after 18 as d_1. The ratios $\frac{d_0}{y_0}$ and $\frac{d_1}{y_1}$ represent the fractions of insured visits before and after the age of 18. The corresponding fractions of uninsured visits before and after the age of 18 are denoted as $(1 - \frac{d_0}{y_0})$ and $(1 - \frac{d_1}{y_1})$. It can be shown that the bias-corrected estimate for the effect of crossing 18 on the insurance coverage is obtained as:

$$\frac{d_1 - d_0}{y_0} \overset{p}{\to} E[D_i(1) - D_i(0) | Y_i(0) = 1], \tag{9}$$

where $\overset{p}{\to}$ denotes the convergence in probability. Expression (9) then provides the population level estimate defined by equation (8).[4]

Under a simplifying assumption that number of visits per patient is constant[5], equation (9) can be estimated as:

$$\frac{\frac{d_1}{y_1} * y_1}{y_0} - \frac{d_0}{y_0} = \frac{\lim_{a\downarrow 18} E[D_i | A_i = a] * \lim_{a\downarrow 18} E[Y_i | A_i = a]}{\lim_{a\uparrow 18} E[Y_i | A_i = a]} - \lim_{a\uparrow 18} E[D_i | A_i = a]. \tag{10}$$

Since we are interested in estimating how an increase in the proportion of the uninsured affects the change in hospital visits as people turn 18, we need to estimate the following equation:

---

[4] For proof see web support to Anderson, Dobkin and Gross (2012).
[5] As explained later in the estimation procedure section, this approach does not exploit the panel structure of the data set, i.e. we are only using the number of visits (insured and uninsured) in each week regardless of the number of patients that generated those visits.

FINANCIAL THEORY AND
PRACTICE
40 (3) 337-356 (2016)

YAN ZHENG AND TOMISLAV VUKINA:
USING THE AGE-BASED INSURANCE ELIGIBILITY CRITERION TO ESTIMATE MORAL HAZARD
IN MEDICAL CARE CONSUMPTION

$$\frac{d_0 - d_1}{y_0} = \frac{\lim_{a\downarrow 18} E[U_i \,|\, A_i = a] * \lim_{a\downarrow 18} E[Y_i \,|\, A_i = a]}{\lim_{a\uparrow 18} E[Y_i \,|\, A_i = a]} + (1 - \frac{\lim_{a\downarrow 18} E[Y_i \,|\, A_i = a]}{\lim_{a\uparrow 18} E[Y_i \,|\, A_i = a]})$$

$$- \lim_{a\uparrow 18} E[U_i \,|\, A_i = a] \tag{11}$$

where U_i is an indicator equal to one if individual i is uninsured, i.e. $U_i = 1 - D_i$.

In the second step, we estimate the reduced form equation, i.e. the percentage decline in visits due to crossing the age 18 threshold, which can be written as:

$$\frac{y_1 - y_0}{y_0}, \tag{12}$$

and obtained by the following estimator:

$$\frac{\lim_{a\downarrow 18} E[Y_i \,|\, A_i = a] - \lim_{a\uparrow 18} E[Y_i \,|\, A_i = a]}{\lim_{a\uparrow 18} E[Y_i \,|\, A_i = a]}. \tag{13}$$

It can be shown that (12) converges in probability to:

$$\frac{y_1 - y_0}{y_0} \xrightarrow{p} E[Y_i(1) - Y_i(0) \,|\, D_i(1) - D_i(0)$$

$$= -1, Y_i(0) = 1] * E[D_i(1) - D_i(0) \,|\, Y_i(0) = 1] \tag{14}$$

where the first factor on the right hand side denotes the measure of moral hazard for individuals who visited the hospital before turning 18 and lost the insurance coverage after turning 18 and the second factor denotes the first stage estimate from equation (9), i.e. the change in insurance coverage due to reaching one's 18th birthday.[6] Therefore, the measure of moral hazard is computed as the ratio of the reduced form estimate and the first stage estimate.

3.2 ESTIMATION PROCEDURE

The key decision in implementing RD design relates to the choice of bandwidth. In most current empirical studies researchers often choose a bandwidth by either cross-validation or by *ad hoc* methods. We rely instead on the optimal, data-dependent, bandwidth choice rule introduced by Imbens and Kalyanaraman (2012). Estimation of the optimal bandwidth was performed with the invoice level data using uniform kernel on [-1,+1] interval. Both a one-year and a two-year data window surrounding the 18th birthday yield very similar estimates of the optimal bandwidth of 1.24 and 1.23 years, respectively (64 weeks surrounding the threshold).[7]

Once optimal bandwidth has been determined, all invoices for the patients outside the 64 weeks surrounding the 18th birthday threshold were dropped, leaving the

---

[6] The proof of this result is also contained in the web support to Anderson, Dobkin and Gross (2012).
[7] A detailed technical description of the estimation algorithm is available from authors upon request.

FINANCIAL THEORY AND
PRACTICE
40 (3) 337-356 (2016)

YAN ZHENG AND TOMISLAV VUKINA:
USING THE AGE-BASED INSURANCE ELIGIBILITY CRITERION TO ESTIMATE MORAL HAZARD
IN MEDICAL CARE CONSUMPTION

sample of 1,883 invoices in total. The summary statistics for this 64 weeks window are presented in table 1. In the left panel we look at the number of visits (invoices). If the invoice shows that the patient visited the hospital with both compulsory and supplemental insurance, we call it an insured visit, otherwise it is an uninsured visit. In other words, an uninsured visit in this paper means a visit to the hospital when the patient did not have the supplemental insurance. Therefore, the younger than 18 group (*Y-18*) has only insured visits whereas the older than 18 group (*O-18*) has both insured and uninsured visits. The number of uninsured visits is only 3% of the total visits.

TABLE 1

Summary statistics: 64-week window around the age 18

Sample	Visits					Patients			
	Invoices	Insured	Uninsured	Male	Female	Number	Male	Female	Visits/patient
Y-18	1,063	1,063	0	480	583	393	181	212	2.71 (3.16)
O-18	820	763	57	372	448	360	152	208	1.98 (1.85)
Total	1,883	1,826	57	852	1,031	753	333	420	2.36 (2.64)

Note: The numbers in parentheses are standard deviations.

We also compared the number of visits by gender. In both *Y-18* and *O-18* groups, women visited the hospital more frequently than men. In the right panel we look at the summary statistics at the patient level. We see that, overall, the sample contains more women than men, which probably explains why we observe more female than male visits. Also, the average number of visits per patient has dropped when the age 18 threshold is crossed from 2.71 to 1.98 visits per patient. The *O-18* group has fewer people visiting the hospital and a lower number of visits per patient but the number of female patients dropped less than the number of male patients as they passed their 18th birthday.
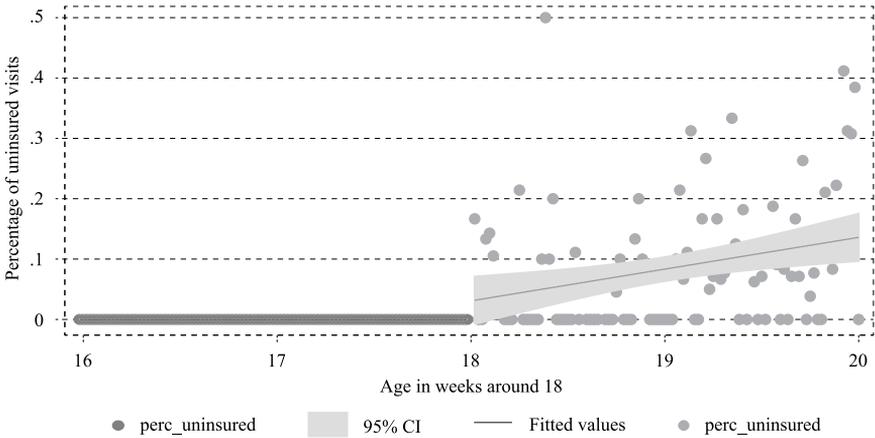
We start the estimation by collapsing the individual invoices data into weekly data and count the number of total and uninsured visits in each week around the patients' 18th birthday. The numbers of days away from the 18th birthday are converted into the numbers of weeks away from the 18th birthday.

The scatter plot and fitted values of the percentages of uninsured visits in each of the 64 weeks away from age 18 are depicted in figure 1. The fitted lines are estimated separately on each side of the cutoff point. There is a clear jump in the percentage of uninsured visits at the cutoff point at age 18. The proportion of uninsured visits is zero for people younger than 18 because all are fully covered by both compulsory and supplemental insurance.

The scatter plot and fitted values of the log of the total visits in each of the 64 weeks surrounding the 18th birthday are displayed in figure 2. There is a sharp drop in hospital visits at the cutoff point at age 18. The graphs provide powerful visual evidence supporting our choice of the regression discontinuity design.
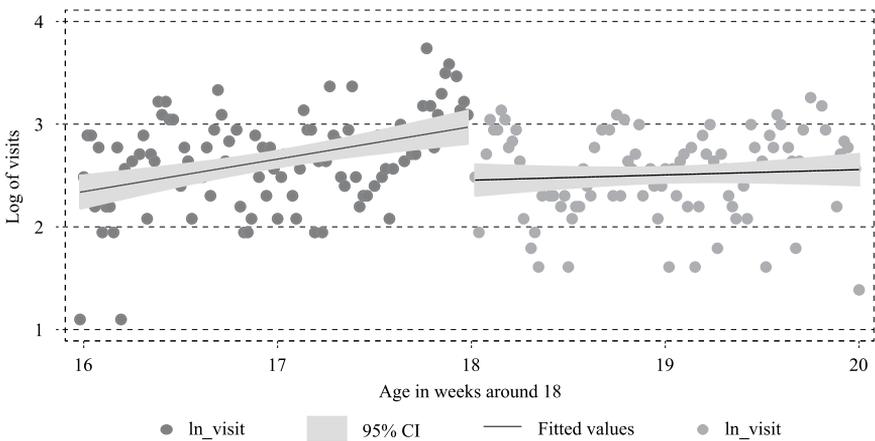
FINANCIAL THEORY AND
PRACTICE
40 (3) 337-356 (2016)

YAN ZHENG AND TOMISLAV VUKINA:
USING THE AGE-BASED INSURANCE ELIGIBILITY CRITERION TO ESTIMATE MORAL HAZARD
IN MEDICAL CARE CONSUMPTION

FIGURE 1

Uninsured visits surrounding patients' 18th birthday (64-week window)

FIGURE 2

Hospital visits around patients' 18th birthdays (64-week window)

The primary concern in the RD design is that factors other than insurance coverage such as high school graduation, starting college, obtaining driving license, entering legal drinking age and starting employment could change discontinuously at the age of 18 and dramatically alter the need to access health care. Because we measure age at the weekly level, only factors that change sharply within a few weeks around the age 18 threshold could bias our estimates. As it turns out, most of those obvious confounders should not bias our estimates. First, high school graduations occur in June and universities start classes in September or October but 18th birthdays are distributed throughout the year. Second, Croatia has a law on the books prohibiting people younger than 18 to purchase or consume alcohol. Yet, there is ample anecdotal evidence that the enforcement is fairly lax, especially in establishments outside the larger cities and along the Adriatic coast

during the tourist season. Finally, a potentially interesting factor is the legal driving age. At age 18 young adults in Croatia become eligible to drive. However, since the drivers' education process is quite lengthy and expensive, most young adults will not obtain their licenses exactly on their 18th birthday but rather during that year or even later.

FINANCIAL THEORY AND PRACTICE 40 (3) 337-356 (2016)

YAN ZHENG AND TOMISLAV VUKINA:
USING THE AGE-BASED INSURANCE ELIGIBILITY CRITERION TO ESTIMATE MORAL HAZARD
IN MEDICAL CARE CONSUMPTION

Based on the framework developed before, in the first step we estimate the limit expectations in equation (11) by seemingly unrelated regression (SUR) using the following two equations:

$$1 - \frac{d_j}{y_j} = \alpha_0 + \alpha_1 I_{over18j} + \alpha_2 aweek_j + \alpha_3 aweek_j * I_{over18j} + \varepsilon_j$$

$$\ln y_j = \beta_0 + \beta_1 I_{over18j} + \beta_2 aweek_j + \beta_3 aweek_j * I_{over18j} + \mu_j \quad (15)$$

where $j=1, 2, ..., 128$; d_j and y_j are weekly counts of insured visits and total visits respectively, $aweek_j$ is age in weeks equals $\frac{1}{64}$ if the visits are made in the first week after turning 18, equals $\frac{2}{64}$ if the visits are made in the second week after turning 18, etc., $aweek_j$ equals 0 if the visits are made on the 18th birthday, $aweek_j$ equals $-\frac{1}{64}$ if the visits are made in the first week before turning 18, etc., $I_{over18j}$ is an indicator variable equal to one if observation j is older than 18 and ε_j, μ_j are random errors with $E(\varepsilon_j) = 0$, $E(\mu_j) = 0$ and $Cov(\varepsilon_j, \mu_j) \neq 0$. The dependent variables are the weekly percent of uninsured visits $1 - \frac{d_j}{y_j}$ and the natural log of weekly number of visits $\ln y_j$. Since the errors affecting percentage of uninsured visits and total visits in each week are likely to be correlated, the use of SUR is justified.

Using estimated coefficients from equation (15), the limit expectations in equation (11) can be recovered as follows:

$$\lim_{a \downarrow 18} E[U_i | A_i = a] = \hat{\alpha}_0 + \hat{\alpha}_1$$

$$\lim_{a \downarrow 18} E[Y_i | A_i = a] = \exp(\hat{\beta}_0 + \hat{\beta}_1)$$

$$\lim_{a \uparrow 18} E[Y_i | A_i = a] = \exp(\hat{\beta}_0)$$

$$\lim_{a \uparrow 18} E[U_i | A_i = a] = \hat{\alpha}_0$$

where $\hat{\alpha}_2$, $\hat{\alpha}_3$, $\hat{\beta}_2$ and $\hat{\beta}_3$ drop out because $aweek = 0$ at the limit of age 18. Therefore, the first stage estimator for the increment in the proportion of uninsured at age 18 becomes:

$$\frac{(\hat{\alpha}_0 + \hat{\alpha}_1) \exp(\hat{\beta}_0 + \hat{\beta}_1)}{\exp(\hat{\beta}_0)} + 1 - \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1)}{\exp(\hat{\beta}_0)} - \hat{\alpha}_0 \quad (16)$$

In the second step we estimate the limit expectations in equation (13) using the following equation:

$$\ln y_j = \beta_0 + \beta_1 I_{over18j} + \beta_2 aweek_j + \beta_3 aweek_j * I_{over18j} + \mu_j \qquad (17)$$

which turns out to be identical to the second equation from (15). Therefore, the reduced form estimator for the percentage change in hospital visits for individuals turning 18 is:

$$\frac{\exp(\hat{\beta}_0 + \hat{\beta}_1) - \exp(\hat{\beta}_0)}{\exp(\hat{\beta}_0)} = \exp(\hat{\beta}_1) - 1 \qquad (18)$$

where $\hat{\beta}_2, \hat{\beta}_3$ drop out in the limit.

Based on equation (14), it is straightforward to see that the moral hazard effect y_1 is identified by dividing the effect of turning 18 on hospital visits, i.e. the reduced form estimator from equation (18), by the effect of turning 18 on insurance coverage, i.e. the first stage estimator from equation (16). Thus, the IV estimator can be written as:

$$\tau_{IV} = \frac{\exp(\hat{\beta}_1) - 1}{\dfrac{(\hat{\alpha}_0 + \hat{\alpha}_1)\exp(\hat{\beta}_0 + \hat{\beta}_1)}{\exp(\hat{\beta}_0)} + 1 - \dfrac{\exp(\hat{\beta}_0 + \hat{\beta}_1)}{\exp(\hat{\beta}_0)} - \hat{\alpha}_0} \qquad (19)$$

4 RESULTS

The first-step SUR results of equation (15) are presented in table 2. The top panel indicates that the percentage of uninsured visits increases after people turn 18 by about 5%. This represents the short-run effect of loss of insurance. Splitting the sample by gender reveals exactly the same pattern of behavior for young men and women.

TABLE 2

Seemingly unrelated regression results: 64-week window

Variable		Total	Male	Female
Dependent	**Independent**			
$1 - \dfrac{d_j}{y_j}$	$I_{over18j}$	0.051 (0.023)**	0.051 (0.023)**	0.051 (0.023)**
	$aweek_j$	0.000 (0.023)	0.000 (0.023)	0.000 (0.023)
	$aweek_j * I_{over18j}$	0.018 (0.033)	0.018 (0.033)	0.018 (0.033)
	Constant	0.000 (0.016)	0.000 (0.016)	0.000 (0.016)
$\ln y_j$	$I_{over18j}$	-0.485 (0.149)***	-0.488 (0.149)***	-0.485 (0.149)***
	$aweek_j$	0.504 (0.146)***	0.491 (0.145)***	0.504 (0.146)***
	$aweek_j * I_{over18j}$	-0.548 (0.209)***	-0.536 (0.208)**	-0.548 (0.209)***
	Constant	2.994 (0.104)***	2.997 (0.103)***	2.994 (0.104)***

Note: The dependent variables in the regressions are proportion of uninsured visits and log of visits at each age in weeks. Standard errors are in parentheses. *** – 1% significance level; ** – 5% significance level; * – 10% significance level.

We expect the coefficient reflecting the long-term effect of loss of insurance, $aweek_j * I_{over18j}$, to be negative. This would indicate that with the passage of time, the percentage of uninsured people should decrease, as those who lost the supplemental insurance after the 18th birthday would gradually become more inclined to purchase the coverage. However, our estimate of this coefficient is positive, though not significantly different from zero, indicating that, at least during the first year after reaching the age of 18, the percentage of young adults with supplemental insurance will not change significantly. This seems to indicate that all those that have supplemental insurance beyond the 18th birthday have it by default, i.e. by maintaining their full-time student status rather than by actually purchasing the policy.

The bottom panel of table 2 results indicate that the number of hospital visits decreases by about 32% (exp(0.485)-1) as young adults cross the 18th birthday threshold. Again, there is hardly any difference in the behavior between genders. In this equation, the coefficient on $aweek_j * I_{over18j}$ is negative and statistically significant indicating that the long-term effect of loss of insurance is negative, at least within the first 64 weeks, given that the number of visits further decreases with time (age).

The second-step OLS regression results of the reduced form equation (17) are displayed in table 3. The coefficients are the same as those obtained using SUR (bottom panel of table 2) but the SUR results are more efficient.

TABLE 3

OLS regression results: 64-week window

Variable		Total	Male	Female
Dependent	**Independent**			
$\ln y_j$	$I_{over18j}$	-0.485 (0.152)***	-0.488 (0.151)***	-0.485 (0.152)***
	$aweek_j$	0.504 (0.148)***	0.491 (0.147)***	0.504 (0.148)***
	$aweek_j * I_{over18j}$	-0.548 (0.212)**	-0.536 (0.211)**	-0.548 (0.212)**
	Constant	2.994 (0.106)***	2.997 (0.105)***	2.994 (0.106)***

Note: The dependent variable in the regression is log of visits at each age in weeks. Standard errors are in parentheses.

The results presented in tables 2 and 3 were obtained with hospital users-only data and as such are not generally valid. We are interested in results that reflect the population as a whole. Table 4 summarizes these results. The asymptotically correct (bias corrected) estimates of the increment in the percentage of uninsured at age 18 (based on equation 16) are displayed in the left-hand column. We estimated a 42% increase in the percentage of uninsured in the whole population, with negligible differences between the female and the male population.

FINANCIAL THEORY AND PRACTICE
40 (3) 337-356 (2016)

YAN ZHENG AND TOMISLAV VUKINA:
USING THE AGE-BASED INSURANCE ELIGIBILITY CRITERION TO ESTIMATE MORAL HAZARD
IN MEDICAL CARE CONSUMPTION

FINANCIAL THEORY AND
PRACTICE
40 (3) 337-356 (2016)

YAN ZHENG AND TOMISLAV VUKINA:
USING THE AGE-BASED INSURANCE ELIGIBILITY CRITERION TO ESTIMATE MORAL HAZARD
IN MEDICAL CARE CONSUMPTION

TABLE 4

Percent change in uninsured (first stage) and total visits (reduced form) and the moral hazard (IV): 64-week window

Sample	First stage	Reduced form	IV estimator
Total	0.416 (0.086)***	-0.385 (0.093)***	-0.924 (0.054)***
Male	0.418 (0.086)***	-0.386 (0.093)***	-0.924 (0.054)***
Female	0.416 (0.086)***	-0.385 (0.093)***	-0.924 (0.054)***

Note: Standard errors are in parentheses.

The percentage change in hospital visits at the age 18 (based on equation 18) is displayed in the middle column of table 4. The three segments of the population all experience a close to 38.5% decrease in the number of hospital visits as they cross their 18th birthday threshold.

The right-hand-side column of table 4 shows the estimates of the moral hazard effect, i.e. the reduction in hospital visits for individuals who visited the hospital before the age of 18 but lost their health insurance after reaching the age of 18. There is a 92% decrease in visits in this group of people. This figure reflects both increased probability of becoming uninsured and the drop in the numbers of hospital visits.

How can one interpret these results? Relying on the summary statistics from table 1, we see that the number of uninsured visits in the *O-18* group is 57, which indicates a 92% reduction in the number of visits due to loss of insurance. Without the moral hazard effect, the number of visits would have been 57/(1-0.924)=750. In other words, 750-57=693 visits never happened because people who lost insurance either never came to the hospital or came less frequently.

We can further compare the actual work load of the hospital staff with the hypothetical workload that would have occurred if nobody lost the insurance after passing their 18th birthday. Assuming the hospital is opened 6 days a week, the actual workload of 17-19 year olds was 18.1 visits per day (1,883 visits/104 days), whereas the hypothetical workload would have been 24.8 visits per day ((1,883+693)/104), a substantial 37% increase.

Another way to look at the result is to take the total number of visits in the *Y-18* group (1,063, table 1) and apply a 38.5% decrease in hospital visits based on the reduced form estimate (middle column, table 4). This would give 409 visits. Taking the number of insured visits in the *Y-18* group (which is also 1,063 because all visits are insured) and applying a 41.6% increase in uninsured visits from the first stage estimates (table 4) yields 442 as the number of uninsured visits. Finally dividing 409/442=0.92 gives us the population-wide estimate of the reduction in the number of visits due to the loss of insurance. This figure coincides with the instrumental variables estimate of moral hazard (table 4, right-hand column).

Since the data used in the estimation covers only a 4-month period, we need to multiply the number of visits that never happened by 3 to obtain an annual estimate of moral hazard at the hospital level. This calculation gives an estimate of 2,079 visits per year.

Finally, the reduction in the number of visits can be converted into a monetary measure of moral hazard by using the average cost per visit (HRK 157.76). This yields around HRK 328,000 (USD 63,200) worth of savings at the hospital level per year.[8] Assuming that the convergence region for the regional hospital in question coincides with the population of the county where the hospital is located, we can obtain a monetary estimate of the moral hazard per capita. Because our estimates of moral hazard are based on a 1.24-year window (64 weeks), we approximate the relevant number of people who passed the age of 18 by the entire 18 year-old cohort plus 24% of the 19 year-old cohort. Based on the 2011 Census, the population of 18-year-olds in the county was 1,401 and the population of 19-year-olds was 1,497, so the moral hazard per person is calculated as 328,000/ (1401+0.24*1497) = HRK 186.

How economically significant is this effect at the national level? Based on the 2011 Census figures, the population of 18-year-olds in Croatia was 47,960 and the population of 19-year-olds was 50,790. This gives an estimate of 60,150 young adults who crossed the 18th birthday threshold during the 1.24-year window and the total moral hazard cost of approximately HRK 11.2 million. Therefore, the total cost of extending the supplemental insurance for all young adults in the country from the current age 18 to age 19.4 would consist of two components: (a) the direct loss in collected premiums for supplemental insurance, and (b) the indirect cost of moral hazard. Assuming the highest premium for supplemental insurance of HRK 130 per month gives us a total cost of HRK 7.8 million per year in lost premiums plus the indirect cost of moral hazard of HRK 11.2 million per year, for the total of HRK 19 million.

5 CONCLUSION

In this paper we implemented fuzzy regression discontinuity design to estimate the moral hazard effect in the health care consumption of young adults drawing on invoice data from one small regional hospital in Croatia. The challenge in estimating moral hazard with such a dataset is that we only observe people who actually used medical care services within the time period covered by the data. To deal with this sample selection bias, we estimate the causal effect of insurance on medical care consumption using the 18th birthday as an instrumental variable. The 18th birthday represents a threshold at which young adults will, by default, lose the full

---

[8] A more precise monetary estimate of the moral hazard effect would be possible by looking at the change in the cost structure of visits before and after crossing the age 18 threshold. The average cost per visit for the *Y-18* group is HRK 132.90, for *O-18* group HRK 189.99, and for both groups combined HRK 157.76. However, since our estimation procedure relies on the visits data and not the cost data, this calculation would be less reliable.

FINANCIAL THEORY AND PRACTICE 40 (3) 337-356 (2016)

YAN ZHENG AND TOMISLAV VUKINA: USING THE AGE-BASED INSURANCE ELIGIBILITY CRITERION TO ESTIMATE MORAL HAZARD IN MEDICAL CARE CONSUMPTION

coverage unless they remain full time students or decide to buy the coverage on their own. The estimation uses weekly counts of hospital visits. Results based on the 64 weeks window estimation show that there is an 92% decrease in hospital visits due to loss of insurance among individuals who did visit the hospital before age 18 and lost their insurance after passing the age of 18.

As an illustration, the above numbers can be related to an alternative program recently promulgated by the Croatian government. Given a severe youth unemployment problem[9], starting January 1, 2015, the government has launched a program whereby the employers are excused from paying the payroll taxes of 17.2% (health insurance earmarked tax rate of 15% plus some other taxes of 2.2%) for a period of up to 5 years for every new employee below the age of 30. For an employee with an average net salary of HRK 5,800 per month the total savings for the employer amount to HRK 11,971 per year. Of course, the amount of money saved by an employer represents a loss to HZZO or the state budget. Relying on the data published by the Croatian Employment Service (Hrvatski zavod za zapošljavanje, HZZ), the total unemployment in the 15-29 age cohort amounts to 102,483 persons. Assuming an optimistic scenario whereby the newly introduced measure will reduce youth unemployment by about 10% or 10,250 people per year, the total cost to the HZZO or the state budget can be estimated at HRK 122.7 million per year.

A comparison of benefits of the two mentioned programs would require a more elaborate modeling and simulation, which is outside the scope of this paper. Nevertheless, the above example highlights several interesting points. First, the cost of moving the threshold for free supplemental insurance from 18 to 19.4 years would cost less than one sixth of the employment stimulus package currently in place. Second, a significant part of that cost is the cost of moral hazard and not the budgetary costs due to lost insurance premiums, hence, the actual impact on hospitals' workloads in terms of increased number of visits could be significant. However, the national health effect of such a program could be substantial. Recall that losing the supplemental insurance at the age of 18 caused a reduction in the number of hospital visits. Some of those lost visits could have profound impacts on early disease detection and prevention and could actually save a lot of money for the health insurance system in the future.

---

[9] According to Eurostat, youth unemployment in the fourth quarter of 2013 was a staggering 48.6%.

REFERENCES

1. Anderson, M., Dobkin, C. and Gross, T., 2012. The Effect of Health Insurance Coverage on the Use of Medical Services. *American Economic Journal: Economic Policy,* 4(1), pp. 1-27. doi: 10.1257/pol.4.1.1

2. Cameron, A. C. [et al.], 1998. A Microeconometric Model of the Demand for Health Care and Health Insurance in Australia. *Rev. of Economic Studies,* (1), 85-106.

3. Card, D., Dobkin, C. and Maestas, N., 2009. Does Medicare Save Lives. *The Quarterly Journal of Economics,* 124(2), pp. 597-636. doi: 10.1162/qjec.2009.124.2.597

4. Chen, S. and van der Klaauw, W., 2008. The work disincentive effects of the disability insurance program in the 1990s. *Journal of Econometrics,* 142, pp. 757-784. doi: 10.1016/j.jeconom.2007.05.016

5. Coulson, N. E. [et al.], 1995. Estimating the Moral-Hazard Effect of Supplemental Medical Insurance in the Demand for prescription Drugs by the Elderly. *American Economic Review,* 85(2), pp. 122-126.

6. Cutler, D. and Zeckhauser, R., 2000. "The Anatomy of Health Insurance" In: A. J. Culyer and J. P. Newhouse, eds. *Handbook of Health Economics Vol. 1A*. Amsterdam: North Holland. doi: 10.1016/S1574-0064(00)80170-5

7. Einav, L. [et al.], 2013. Selection on Moral Hazard in Health Insurance. *The American Economic Review,* 103(1), pp. 178-219. doi: 10.1257/aer.103.1.178

8. Hahn, J., Todd, P. and van der Klaauw, W., 2001. Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica,* 69(1), pp. 201-209. doi: 10.1111/1468-0262.00183

9. Holly, A. [et al.], 1998. An Econometric Model of Health Care Utilization and Health Insurance in Switzerland. *European Economic Review,* 42, pp. 513-522. doi: 10.1016/S0014-2921(98)00003-8

10. Imbens, G. W. and Kalyanaraman, K., 2012. Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *Review of Economic Studies,* 79(3), pp. 933-959. doi: 10.1093/restud/rdr043

11. Imbens, G. W. and Lemieux, T., 2008. Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics,* 142(2), pp. 615-635. doi: 10.1016/j.jeconom.2007.05.001

12. Lee, D. S. and Lemieux, T., 2010. Regression Discontinuity Designs in Economics. *Journal of Economic Literature,* 48, pp. 281-355. doi: 10.1257/jel.48.2.281

13. Liu, X., Nestic, D. and Vukina, T., 2012. Estimating Adverse Selection and Moral Hazard Effects with Hospital Invoices Data in a Government--Controlled Healthcare System. *Health Economics,* 21(8), pp. 883-901. doi: 10.1002/hec.1756

14. Paully, M., 1968. The Economics of Moral Hazard: Comment. *The American Economic Review,* 58(3), pp. 531-537.

15. Savage, E. and Wright, D. J., 2003. Moral Hazard and Adverse Selection in Australian Private Hospitals: 1989-1990. *Journal of Health Economics,* 22, pp. 331-359. doi: 10.1016/S0167-6296(02)00104-2

FINANCIAL THEORY AND PRACTICE
40 (3) 337-356 (2016)

YAN ZHENG AND TOMISLAV VUKINA:
USING THE AGE-BASED INSURANCE ELIGIBILITY CRITERION TO ESTIMATE MORAL HAZARD
IN MEDICAL CARE CONSUMPTION

16. Schwartz, K. and Schwartz, T., 2008. Uninsured Young Adults: A Profile and Overview of Coverage Options. *The Kaiser Commission on Medicaid and the Uninsured* Issue Paper.

17. van der Klaauw, W., 2002. Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach. *International Economic Review,* 43(4), pp. 1249-1287. doi: 10.1111/1468-2354.t01-1-00055

18. Vera-Hernandez, A. M., 1999. Duplicate Coverage and Demand for Health Care. The Case of Catalonia. *Health Economics,* 8, pp. 579-598. doi: 10.1002/(SICI)1099-1050(199911)8:7<579::AID-HEC478>3.0.CO;2-P

19. World Bank, 2015. Overview. Available at: [http://www.worldbank.org/en/country/croatia/overview].

20. Yörük, B. K., 2015. Health Insurance Coverage and Risky Health Behaviors among Young Adults. *CESIFO Working paper*, No. 5495. Available at: [https://ag-econ.ncsu.edu/wp-content/uploads/2015/11/FWB_2015_Final.pdf].

Financijske institucije i tržišta Europske unije - regulacija i supervizija

Financial institutions and markets in the European Union - regulation and supervision

MARTA BOŽINA BEROŠ
Juraj Dobrila University, Pula, 2015, pp. 246

Ana GRDOVIĆ GNIP, PhD
Dr. Mijo Mirković Faculty of Economics and Tourism, Juraj Dobrila University of Pula, Preradovićeva 1/1,
52100 Pula, Croatia
e-mail: agrdovic@unipu.hr

FINANCIAL THEORY AND
PRACTICE
40 (3) 357-360 (2016)

MARTA BOŽINA BEROŠ:
FINANCIAL INSTITUTIONS AND MARKETS IN THE EUROPEAN UNION – REGULATION
AND SUPERVISION

"And once the storm is over, you won't remember how you made it through, how you managed to survive. You won't even be sure, whether the storm is really over. But one thing is certain. When you come out of the storm, you won't be the same person who walked in. That's what this storm's all about."

Haruki Murakami

The economic and financial crisis that reached its peak in 2008 left no part of the financial system untouched. At the European level, it highlighted the need for dauntless initiatives to improve the insufficient regulatory rules, since the European system of regulation was in great need of *"repairs"* (de Larosière, 2009). One consequence of the crisis, and not in Europe alone, was that more regulation was widely advocated and indeed implemented. To rephrase part of Marukami's quote "When we come out of the crisis, regulation won't be the same as when we walked in. That's what the crisis is all about."

The financial system is by its nature dynamic and unstable, speculation and risk being inherent in its origins and development (Galbraith, 1993). On the other hand, regulation and rules come after facts (*ex facto oritur jus*), meaning that financial rules are not capable of being designed so as to anticipate financial facts, i.e. financial creativity and innovation. Accordingly, financial regulation finds it difficult to manage and control all the risks involved within the financial system. The textbook entitled *Financial institutions and markets in the EU – regulation and supervision* written by Marta Božina Beroš examines the regulation of financial institutions and markets at the European level, contributing to a critical understanding of the dynamics of financial regulation in the EU through combined studies of law and economics.

The textbook is organized in six chapters, the first serving as an introduction. The author briefly explains the genesis of financial integration, emphasizing its developmental stages in the EU context and pinpointing the challenges it poses to in European regulatory policies.

With the second chapter, the author begins to critically elaborate on the topic and introduces all the main theoretical foundations needed to puzzle out the rest of the book. Since the regulation of financial institutions and markets combines the study of regulatory theories from law and from economics, the author first of all relates these two fields by explaining the three main economic reasons for the need to regulate economic relationships juridically. Firstly, if there were no regulatory policies, trading with financial instruments, financial assets and financial transactions would be exposed to the perils of *asymmetric information* since the trading price is based on contractual terms and market developments. The author points out that one of the goals of financial regulation is to balance the rights and obligations between the contracting parties and thus allowing an informed and easier financial decision making process. *Systemic risk externalities* represent the

second underlying reason for financial regulation. The aim is to avoid or alleviate the contagion effect initialized by the collapse of one financial institution that spills over to other institutions included in the system. On one hand, financial institutions often embark on excessively risky financial transactions for the sake of profit or the status of market leader. On the other hand, such negligent behaviour can endanger the fulfilment of the financial contract. Financial regulation has the goal of curtailing these two risks, i.e. the *adverse selection and moral hazard risks*, both listed as the third reason by the author.

Besides explaining the interplay of law and economics in the field of financial regulation, this (second) chapter examines the thin dividing line between regulation and supervision. Although in practice regulation and supervision frequently overlap, the author clearly distinguishes supervision from regulation and explains the models of supervision used to monitor the outcomes of implemented regulatory measures.

In the following three chapters the author carefully delineates the three areas that are subject to financial policing: (1) financial assets and transactions, (2) financial institutions, and (3) financial markets. Although each chapter outlines the main theoretical definitions and assumptions, the author focuses on the specificities within the EU's regulatory framework. So, she fleshes out assets and transactions typical for most member states (securities, bank deposits, bank loans and financial derivatives), analyses financial institutions through the lenses of European (banking) regulation as well as Basel Accords and also surveys financial markets under the MiFID (Markets in Financial Instruments Directive) framework. Moreover, in all of the three mentioned chapters, the author relates the topic to the latest global financial crisis and gives a critical discussion. For instance, special emphasis within the third chapter (Financial assets and transactions) is given to securities lending and securitization, because these have often been mentioned by public experts as the main causes of the financial collapse.

The last chapter constitutes a concluding critical review of the topic of regulatory governance in the EU, clearly distinguishing the pre- from the post-crisis period. In fact, the crisis has left no part of the financial system untouched and the European Commission has undertaken a long list of reforms narrowing down the discretionary powers of EU members. The author pinpoints centralisation and institutional consolidation as the main characteristics of post-crisis EU financial regulation. However, it is clear within the discussion that the author does not completely agree with such principles and gives arguments why the proposition of "maximum harmonization" might not be the best solution for the EU. Having in mind that the EU tightened the discretionary powers of its members not only in the field of financial regulation but also in economic governance as a whole, in the very last paragraph (p. 223) of the book she concludes the following:

> *"Unfortunately, excessive centralization and regulatory uniformity will simply ignore the interests of small(er) member states in favour of the*

greater good, i.e. the stability of the Euro Area and European finance. However, the role of members can not be so easily reduced. EU policy makers and the public can continue with their plans assuming that the EU is on track to become a perfectly integrated super-state. However, it is time to become aware of the limitations of integration – political and economic. Member states are still the EU reality, even when they are part of a sui generis *political and economic integration such as the EU, and their diversity must be respected, as must their right to protect their own economic interests."*

Financial regulation is not an easy task. Neither is writing a textbook about it. Marta Božina Beroš's manual reinforces the understanding of financial systems and regulatory governance, a topic often neglected in the Croatian literature. Although the author is familiar with regulatory theories from law and economics, the textbook discusses financial regulation from an economic perspective in the European context. The writing is not particularly challenging to read, but requires a knowledge of the basics of economics and finance. It is appropriate for use in advanced undergraduate and graduate finance courses and can serve as an excellent reference for educators and practitioners. Giving that each of the chapters begins with an overview of the chapter, a reader will easily spot and locate specific concepts of interest.

Recognizing that textbooks are not static publications, the author might pursue the matter further in a subsequent edition, detailing the particularities of the complex European multi-jurisdictional system (legislation, institutions and decision-making) and showing the legal perspective of financial regulation as well. Moreover, more emphasis might be devoted to financial supervision and its country-case examples. This would enhance the book's usefulness, develop its contribution to the topic of financial regulation and supervision, and at the same time make it accessible to a wider audience.