Essays in Economics

DISSERTATION
of the University of St. Gallen,
School of Management,
Economics, Law, Social Sciences
and International Affairs
to obtain the title of
Doctor of Philosophy in Economics and Finance

submitted by
Adam Aleksander Pigoń
from
Poland

Approved on the application of
Prof. Dr. Winfried Koeniger,
Prof. Dr. Almuth Scholl,
and
Prof. Dr. Reto Föllmi

Dissertation no. 4864

Krystel, Warsaw, 2019

Contents

1 Acknowledgements and Summary

Writing a dissertation is always a great intellectual endeavour and it can be considered a great pleasure, at least from the perspective of having a finished thesis on one's desk. Having worked on my PhD project for more than four years I can only state that it was a period in which I could develop my knowledge, of both economics as a science and of the world itself, and simultaneously experience many great moments. It is clear to me that the whole community of the University of St. Gallen (HSG) contributed to all these happy memories I have. Nevertheless, there are some people to which I am particularly grateful.

First and foremost, I would like to thank my supervisor Winfried Koeniger, who offered me the opportunity to pursue the PhD studies at HSG. Moreover, many long discussions and comments from his side made it possible to me to finish my research projects. His effort was absolutely indispensable for this doctoral thesis to materialize. Without any doubts, Winfried Koeniger is the person from whom I learnt most during my studies in St. Gallen.

Secondly, I would like to express my gratitude to my co-authors Michał Ramsza and Gyula Seres, who are not only my scientific collaborators but also closest friends.

Furthermore, I have benefited from interactions with other HSG professors: Monika Bütler, Guido Cozzi, Reto Föllmi, Roland Hodler and Michael Lechner. My teachers of economics and econometrics from the Warsaw School of Economics, Tilburg University, St. Gallen University and the University of Zurich also must be acknowledged: Agata Boratyńska, Bogusław Czarny, Marek Góra, Marcin Kolasa, Felix Kübler, Jochen Mankart, Wieland Müller, Martin Salm, Agnieszka Stążka, Uwe Sunde, Frederic Vermeulen and Łukasz Woźny. I would like to thank Almuth Scholl from University in Konstanz for being the external referee of the dissertation.

The time I spent in St. Gallen would not have been that pleasant without my colleagues and friends: Melanie Annaheim, Hugo Bodory, Margaret Davenport, Irene Fensore, Hans Fricke, Daniel Goller, Charles Gottlieb, Katharina Jaronicka-Hofer, Lukas Inderbitzin, Adrian Jäggi, Michael Knaus, Alex Krumer, Benedikt Lennartz, Georgi Lautliev, Stefan Legge, Piotr Łukaszuk, Gabriel Okasa, Marc-Antoine Ramelet, Thomas Spycher, Carina Steckenleiter, Arnau Valladares, Janosch Weiss and Carlo Zanella.

I am also very grateful to our secretaries at the SEW: Conny David, Susanne Moser, Sabine Peters and Caroline Ubieto. I would like to thank them for helping me in fulfilling all the administrative duties.

The financial support of the 'Mitteleuropastiftung' is acknowledged and praised for its generosity.

Lastly, special words of gratitude are given to my family for the ongoing support in writing my thesis. Here, I would like to thank my beloved wife Renata Pigoń, my mom Małgorzata Sobieszek-Warzecha, my grandparents Anna and Kazimierz Pigoń, my parents-in-law Anna and Tadeusz Śmietanka and my sister and siblings-in-law: Monika Warzecha, Katarzyna Dziopa and Wojciech Śmietanka.

St.Gallen & Warsaw, December 2018

Adam Pigoń

Summary

The dissertation consists of three independent chapters. Although they use different methodologies and contribute to different fields of economics, they all try to answer the same question: 'how can government improve the 'state of affairs' by changing its policy?'. In my essays, I analyze topics where government policy is concerned with mechanism design for public goods auctions, fiscal rules governing sustainable fiscal policy, or with targeting specific household groups with stimulus packages.

The first chapter investigates the incidence of the so-called *hand-to-mouth*, or simply *HtM*, consumers, i.e. households that consume their whole disposable income. Based on existing models, which introduced a two-dimensional analysis of *HtM* households with respect to their liquid and illiquid stocks of assets, I consider an additional important feature in the analysis: human capital. The chapter presents a statistical description of various types of *HtM* and *non HtM* households, listing their most important socio-economic features. Moreover, the analysis is augmented with a theoretical model explaining the households' behavior regarding human capital accumulation. Recognition of the characteristics of *HtM* consumers, including their education, is important as it allows governments to target specifically these households, which have a high marginal propensity to consume, with expansionary fiscal policy measures. This would increase the efficiency of fiscal policy in stabilizing the business cycle.

The second chapter, written together with Michał Ramsza, explores fiscal rules that govern the fiscal policy in Switzerland, Germany and Poland. Using a Monte Carlo simulation, we generate artificial time series of GDP and government revenues that, together with government expenditures implied by the rules, create paths of public debt and deficits. It gives us a framework in which we compare the performance of various fiscal rules for stabilizing debt at low levels and for inducing anticyclical fiscal policy.

The third chapter, co-authored with Gyula Seres, delves into the efficiency of mechanism design in the field of public procurement auctions. Using a novel data set with public contracts for highway construction in Poland, we analyze effects of screening, i.e. rejections of offers that do not fulfill legal requirements of a contract, on entry and the bidding policy of firms. We conclude that large firms are better prepared to defend their offers legally and small firms are discouraged from entry as they anticipate potential rejections. Screening policy results thus in a decrease of competition, in terms of the number of auction participants, and in an increase of highway contract prices.

Zusammenfassung

Die Dissertation besteht aus drei unabhängigen Kapiteln. Obwohl sie unterschiedliche Methoden anwenden und zu verschiedenen Bereichen der Volkswirtschaftslehre beitragen, versuchen sie alle, die gleiche Frage zu beantworten: wie kann der Staat die 'Lage' verbessern, indem er eine Politikmassnahme einführt oder ändert? In meinen Essays analysiere ich Themen, bei denen es um die Gestaltung von Mechanismen für öffentliche Güterauktionen, um Fiskalregeln für eine nachhaltige Fiskalpolitik oder um das gezielte Erreichen bestimmter Haushaltsgruppen mit Konjunkturpaketen geht.

Das erste Kapitel untersucht die Inzidenz der so genannten *Hand-to-Mouth*, oder einfach *HtM*, Verbraucher, d.h. Haushalte, die ihr gesamtes verfügbares Einkommen verbrauchen. Basierend auf bestehenden Modellen, die eine zweidimensionale Analyse von *HtM* Haushalten hinsichtlich ihrer liquiden und illiquiden Vermögensbestände eingeführt haben, berücksichtige ich ein weiteres wichtiges Merkmal für die Analyse: das Humankapital. Das Kapitel enthält eine statistische Beschreibung der verschiedenen Arten von *HtM* und *non-HtM* Haushalten und listet ihre wichtigsten sozioökonomischen Merkmale auf. Darüber hinaus wird die Analyse durch ein theoretisches Modell ergänzt, welches das Verhalten vom Haushalt hinsichtlich der Humankapitalakkumulation erklärt. Die Berücksichtigung von den Merkmalen der *HtM* Konsumenten, einschließlich ihrer Ausbildung, ist wichtig, da sie es den Regierungen ermöglicht, mit expansiven fiskalpolitischen Maßnahmen gezielt diese Haushalte mit einer hohen marginalen Konsumneigung erreichen. Dies würde die Effizienz der Fiskalpolitik bei der Stabilisierung von Konjunkturschwankungen erhöhen.

Das zweite Kapitel, zusammen mit Michał Ramsza verfasst, untersucht die Fiskalregeln in der Schweiz, Deutschland und Polen. Mit Hilfe einer Monte-Carlo-Simulation generieren wir künstliche Zeitreihen des BIP und der Staatseinnahmen, die zusammen mit den durch die Regeln implizierten Staatsausgaben Pfade der Staatsverschuldung und -defizite ergeben. Dies gibt uns einen Rahmen, in dem wir die Leistung verschiedener Fiskalregeln zur Stabilisierung der Verschuldung auf niedrigem Niveau und zur Herbeiführung einer antizyklischen Finanzpolitik vergleichen.

Das dritte Kapitel, das gemeinsam mit Gyula Seres verfasst wurde, befasst sich mit der Effizienz des Mechanism Design im Bereich der öffentlichen Auftragsvergabe. Anhand eines neuartigen Datensatzes mit öffentlichen Aufträgen für den Autobahnbau in Polen, analysieren wir die Auswirkungen des Screenings, d.h. der Ablehnung von Angeboten, die nicht den gesetzlichen Anforderungen eines Vertrages entsprechen, auf den Zugang und die Ausschreibungspolitik von Unternehmen. Wir kommen zu dem Schluss, dass große Unternehmen besser darauf vorbereitet sind, ihre Angebote rechtlich zu verteidigen, und kleine Unternehmen vom Beitritt abgehalten werden, da sie potenzielle Ablehnungen erwarten. Die Screening-Politik führt somit zu einem Rückgang des Wettbewerbs in Bezug auf die Anzahl der Auktionsteilnehmer und zu einem Anstieg der Preise für Autobahnaufträge.

2 A Simple Model of ***Educated Hand-to-Mouth*** Consumers

Abstract:

The main goal of this paper is to analyze the role of human capital for the incidence of the so called *hand-to-mouth* (*HtM*) consumers: those having no liquid resources. The proposed model is based on the recently published two-asset model by Kaplan, Violante, and Weidner (2014) with an extension to allow for endogenous accumulation of human capital. I show how the *HtM* status of consumers depends on their innate abilities, time preference and initial resources. *Wealthy HtM* households, i.e. households with illiquid resources but with little or no liquidity, are more able, more patient and initially richer than *poor HtM*. As a consequence, they accumulate more human capital than *poor HtM* households. For both types of households their status depends on having a steep income path, which is endogenous because of the endogenous human capital accumulation. The correlation of observable characteristics with *HtM* behavior may be of interest for economic policy: these consumers have a high marginal propensity to consume so that targeting them could increase the effectiveness of fiscal policy.

JEL-Classification: D31, E21, H31

Keywords: fiscal policy, hand-to-mouth, household consumption, liquidity

2.1 Introduction

The models of consumption in the spirit of Friedman, Modigliani or Hall have shaped the thinking about households' consumption decisions at all possible time horizons (for an excellent survey about different aspects of life-cycle models see Browning and Crossley (2001)). Nevertheless, the permanent income/life-cycle hypothesis is prone to some puzzles. One of them is the 'excess sensitivity' - a too large response of consumption to expected changes of income while the theory predicts that the consumption response should be smoothed. Together with the fact that consumption/saving decisions do not react much to interest rate changes (see e.g. Campbell and Mankiw (1989)) this suggests that a significant fraction of households does not satisfy their Euler equations because of liquidity constraints.

Standard incomplete market models (e.g. Aiyagari (1994) and Huggett (1996)), with ex-ante identical consumers that face uninsurable income shocks, can generate a fraction of the population that is liquidity constrained. Furthermore, there are saver-spender models (e.g. Campbell and Mankiw (1989) or Gali, Valles, and Lopez-Salido (2007)) that assume two groups of consumers: one group that optimizes in a forward-looking manner and another group that consumes all income they earn immediately. In both types of models a group of *hand-to-mouth* (*HtM*) households consumes everything they earn, yet both types of models miss a significant fraction of *wealthy hand-to-mouth* households. These consumers are wealthy as they possess a positive stock of illiquid assets like housing, often considerable in size, and, simultaneously, are constrained in terms of liquidity.

The existence of the *wealthy hand-to-mouth* households, or just *wHtM*, is the main focus of Kaplan and Violante (2014) and Kaplan et al. (2014). The authors suggest a model in which there are two types of assets: liquid ones, and illiquid ones with a higher rate of return. Based on a wide battery of robustness checks the authors document that about 1/3 of consumers in the US are hand-to-mouth, out of which 2/3 are *wealthy HtM* (i.e. those consuming all their current income but having simultaneously a strictly positive stock of illiquid assets like housing). Ignoring such a group of consumers would lead to spurious conclusions on the effects of fiscal policy as *wealthy hand-to-mouth* households exhibit, in comparison with other households, a particularly high marginal propensity to consume (MPC). This group of households is crucial for fiscal policy, whose effectiveness for stimulating aggregate demand hinges on correctly targeting transfers or tax rebates to groups with high MPC.

A fundamental question, from both a policy and theoretical perspective, is the reason for the existence of the *wealthy-hand-to-mouth* consumers. Possible explanations, mentioned already in Kaplan et al. (2014), are specific preferences or exogenous characteristics of buying some types of goods like houses; idiosyncratic bad shocks like health problems or job loss; or, finally, poor ability to plan expenses and foresee income. Another potential reason for the incidence of *wealthy hand-to-mouth* households is related to the accumulation of human capital in the life-cycle. The idea of this paper is to analyze potential explanations, with focus on the human capital channel, so that the characteristics of the *wealthy-hand-to-mouth* consumers can be precisely pinpointed.

Although the *wealthy HtM* share some important features, such as similar fertility, with *poor HtM* (other than the one that defines both groups: no liquidity), they are also similar in some respects to the *non-hand-to-mouth* (*nHtM*) households, e.g. in the probability of having an unemployed member of the household or the probability of receiving social benefits. Most distinctive differences between *poor HtM*, *wealthy HtM* and *nHtM* households are their income and education that are monotonically increasing between these groups (see Figure 2.1). Especially the latter difference is interesting and its magnitude means that it is worth extending the seminal model in Kaplan and Violante (2014) by endogenizing educational attainments and labor income, which is the main purpose of this paper.

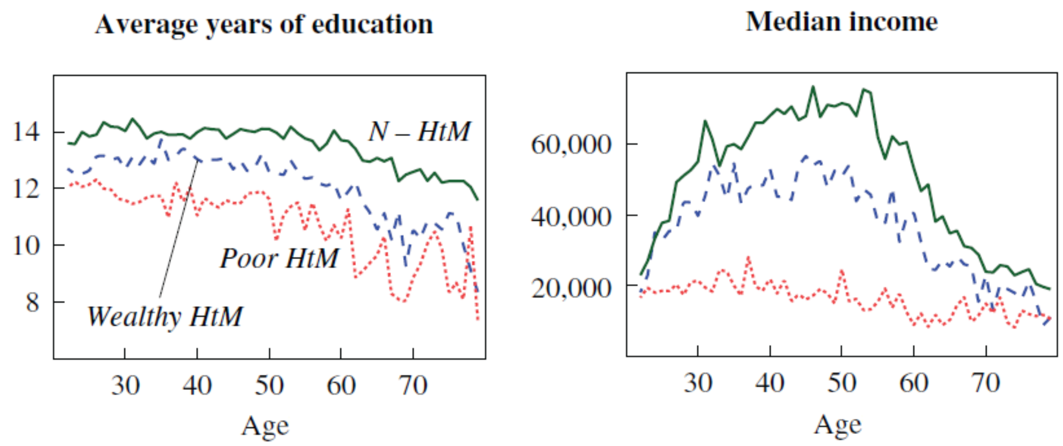


Figure 2.1: *Education and income in various types of households (values for heads of households). Source: Kaplan et al. (2014)*

The modelling of educational decisions in the style of BenPorath (1967), Huggett, Ventura, and Yaron (2006) or Huggett, Ventura, and Yaron (2011) in the two-asset 'Kaplan-Violante' framework is expected to give us answers to the question about the differences in human capital across the different consumer types depicted in the left panel of Figure 2.1 and, together with the labor supply, about the differences in income depicted in the right panel of the figure. Altogether, this analysis of the endogenous current and expected income of a consumer should shed further light on the incidence of the *hand-to-mouth* behavior.

There are many potential channels of causality that are presumed to be involved in interdependencies between education and *HtM* status. Accumulation of large stocks of human capital increases the steepness of earnings over the life cycle and thus magnifies the incidence of liquidity constraints. Moreover, expenses on human capital may be affected themselves by the liquidity constraints. Lastly, educational expenditures can also imply different timing of other expenditures, particularly, those related to mortgages.

The analysis treats human capital as another type of asset with a given rate of return. An agent chooses a portfolio composition that consists of liquid and illiquid assets and also of human capital. Different characteristics of these three types of assets are relevant for the portfolio choice. Obviously, the relative rates of return on all these assets are important but other features as well: human capital is inalienable and it gives a return only through exerting labor supply. Furthermore, human capital cannot be 'made liquid', while illiquid assets such as housing can be sold at some transaction cost. Finally, the use of human capital as a collateral for credit is more limited than for illiquid assets such as housing.

Nevertheless, human capital can be analysed as an asset, mainly because of the fact that it earns a return that can be compared with returns of other assets. Furthermore, a decision to accumulate more education is also an economic decision that contains a 'cost-benefit' analysis, which is similar to decisions regarding financial investments. Most recently, Huggett and Kaplan (2015), have treated human capital as a financial asset.

It is interesting for several reasons to endogenize educational decisions and income, while maintaining the general two-asset framework proposed by Kaplan and Violante (2014). The main reason is that this allows to capture a relevant dimension of consumer characteristics as documented in Figure 2.1. In view of observed heterogeneity in the population (analysis of such heterogeneity in the context of macroeconomic modelling is discussed in Heathcote, Storesletten, and Violante (2009), also with comments about educational attainments), this additional dimension will help answering important theoretical or policy questions, especially because decisions about human capital formation affect a wide spectrum of socio-economic outcomes (see e.g. Card (1999) or Bhuller, Mogstad, and Salvanes (2017)). Applications of the two-asset incomplete market model with endogenous human capital and income can be interesting from both the theoretical point of view and an economic policy perspective:

- Fiscal policy was the main motivation for Kaplan and Violante (2014) and Kaplan et al. (2014) as, according to their analysis, the *wealthy HtM* households exhibit the highest MPC. These households are though difficult to discern as they share socioeconomic characteristics with households having lower MPC values. The most unambiguous feature of the wealthy HtM is the high probability of having a mortgage loan yet it is politically difficult to target expansive fiscal policy on mortgage-indebted households. Investigation on the role of human capital could potentially reveal other discernible characteristics of the group with a high MPC. Therefore, it could lead to policy implications that are more politically feasible.

- Life-cycle models offer a coherent framework for analysing life-time choices (see e.g. Browning and Crossley (2001)). The theory should not abstract from decisions on accumulating human capital and various types of physical assets as these choices are linked and shape the consumption patterns over the lifecycle.

The rest of this paper is organized as follows. Section 2.2 contains a descriptive statistical analysis of the US households with respect to their *hand-to-mouth* status and their assets portfolio composition. Section 2.3 presents a stylized three-period model of the *hand-to-mouth* behavior with human capital accumulation. Section 2.4 shows a partial analytical solution to the model, while Section 2.5 describes numerical results of the consumer's optimization problem. Finally, Section 2.6 concludes and the Appendix contains formal derivations and details of the theoretical and empirical models presented in the paper.

2.2 Analysis of the survey data on US households

This section presents an analysis of the Survey of Consumer Finances (SCF) data on US households and their assets portfolio composition. A special attention is given to demographic and economic characteristics of different household types. The analysis builds on Kaplan et al. (2014), henceforth called 'KVW', but extends their paper in two dimensions. Firstly, it uses the newest data from waves 2010, 2013 and 2016 in order to establish the robustness of the *hand-to-mouth* phenomenon across time in the post-crisis reality, while KVW use waves of SCF from 1989 to 2010. Secondly, the analysis attaches more importance to the human capital accumulation decisions by exploring student loans. In order to maintain comparability of results, all definitions are identical as in KVW.

2.2.1 Defining *wHtM*, *pHtM* and *nHtM* households in the data

Methodology In order to define household types we need data on households' income y, liquid assets m and illiquid assets a. Plain definitions are given as follows: *wHtM* households do not have liquid assets m but have a positive stock of illiquid assets a, *pHtM* households have neither liquid assets m nor illiquid assets a and *nHtM* consumers have positive stocks of liquid assets m. More formally, the *hand-to-mouth* status can show up at two kinks of the intertemporal budget constraint: at zero liquid assets or at the credit limit. The stock of liquid assets should be measured at the end of a pay period so that it is clear if a household can move liquid wealth between periods or it is at its kink. Unfortunately, SCF reports balances at some random dates (dates of interview), which means that measuring household types needs further assumptions that may introduce measurement error.

As KVW, we assume that households receive their income discretely and they consume continuously over the whole pay period at a constant rate. We abstract from discrete consumption commitments like rents, mortgage and loan payments, alimony etc. or, equivalently, we assume they are incurred in the middle of the pay period or are incurred smoothly. Liquid resources of a *hand-to-mouth* household peak then in the payday at the value of y and are depleted to zero (or to the credit limit) at the end of the pay period. Average liquid holdings within the pay period equal $y/2$ due to the constant rate of consumption. Then, *hand-to-mouth* households are those whose liquid wealth is positive and lower or equal to half of their income (households at the zero kink) or whose liquid wealth is negative and lower or equal to half of income plus the credit limit (households at the credit constraint kink). Formally, *wHtM* and *pHtM* households are defined as follows:

- *wHtM* at the zero kink: $a > 0$ and $0 \leq m \leq y/2$
- *wHtM* at the credit constraint kink: $a > 0$ and $m \leq 0$ and $m \leq y/2 - \underline{m}$
- *pHtM* at the zero kink: $a \leq 0$ and $0 \leq m \leq y/2$
- *pHtM* at the credit constraint kink: $a \leq 0$ and $m \leq 0$ and $m \leq y/2 - \underline{m}$

Effectively, variable m depicts average liquid wealth. Variable $\underline{m}$ is the credit limit for a given household. The pay period is set to two weeks and the credit limit is equal to monthly income. As KVW point out, the criteria presented above are conservative and give a lower limit on *hand-to-mouth* incidence. The reason is that households that start a given period with any liquid wealth above the kink at zero or above the credit limit kink but end the period exhausting all possible liquid resources carry average liquid holdings above the limits listed above. All other households are called *non-hand-to-mouth*.

Contrary to KVW, who treat all *nHtM* consumers identically, we differentiate in the statistical analysis between *poor nHtM* (*pnHtM*), who have negative or zero illiquid assets, and *wealthy nHtM* (*wnHtM*) households, who have strictly positive illiquid assets. Both consumer types are not expected to differ in their MPC out of transitory income shocks as both have positive stocks of liquid assets implying that they are not at their budget/liquidity constraints. Nevertheless, they may differ systematically in their socioeconomic characteristics or in phases of the life-cycle, in which given household types typically occur. The latter feature is fundamental in view of the analysis of the human capital accumulation in the life-cycle and it warrants their differentiation.

Negative positions in illiquid assets a, both in case of *HtM* and *non-HtM* households, are very rare. These positions occur as a result of decreasing house prices, which push house values below residual values of mortgage loans. Although these consumers own illiquid assets (houses and flats), they cannot use them to smooth consumption and, therefore, they resemble *poor HtM* or *poor nHtM* households and they are counted as such.

Income Income y is defined as the sum of all annual labor income and other income including self-employment income, pension income, government benefits (SSI, food stamps, unemployment benefits etc.), private transfers like alimony and other sources of regular income other than investment income. The exclusion of interest, dividend and other capital income is dictated by its irregular character that may not be captured by SCF correctly.

Liquid wealth Liquid assets are a sum of checking, saving, money market and call accounts, mutual funds, stocks, corporate and government bonds. SCF data does not contain information on cash holdings. Therefore, we have to use an imputation method, which is used by KVW, and which is based on the data from the Survey of Consumer Payment Choice (SCPC). We use editions 2010, 2013 and 2015 to match them with corresponding SCF waves (SCPC 2016 edition is not available). The imputation implies computation of the ratio of average cash holdings taken from SCPC to the median holdings of checking, saving, money and call accounts from SCF. Then, the aforementioned holdings of assets are inflated by the computed ratio. Cash holdings in 2016 dollars were 150, 236, 204 in years 2010, 2013 and 2016 resulting in imputed cash ratios of about 1.05, 1.07

and 1.05, respectively.[1] Liquid wealth is equal to liquid assets minus liquid debt, which contains credit card and installment debt.

Illiquid wealth Illiquid assets are defined as a sum of housing residential and nonresidential real estates, private retirement accounts, life insurances, certificates of deposits and saving bonds. Illiquid wealth is a sum of illiquid assets net of mortgage debt. Similarly as in the baseline calculation method in KVW, illiquid wealth does not contain wealth in vehicles nor student loans.

Sample selection All graphs and statistics in the paper use a subsample of the SCF households, in which the household head is between 22 and 79 years of age. In order to maintain comparability with KVW, we drop households with negative income and those whose income entirely comes from self-employment. Standard SCF weights, calculated to account for oversampling of certain households groups, nonresponsiveness and hiding the identity of certain households, are applied to construct all statistics.

Names of SCF variables used for the construction of above explained measures are provided in Appendix 2.2.

2.2.2 Results

Introduction This section provides descriptive statistics regarding the *hand-to-mouth* phenomenon. By comparing the results with KVW we analyze potential volatility of the *HtM* households before and after the 2008 crisis. In all graphs we depict weighted mean or median values that are calculated on a pooled data combining the SCF 2010, 2013 and 2016 waves. All financial variables are in 2016 US dollars and age pertains to the head of the household. Graphs showing fractions of financial assets, net primary housing and retirement accounts to liquid and illiquid wealth are trimmed, eliminating the 0.1 percent of largest and smallest values of a given variable to reduce sensitivity to outliers. All fractions are (weighted) mean fractions. In order to smooth graphs visually a rolling mean is applied to all raw statistics. The rolling mean is defined in the following way: $\beta_{age} = (b_{(age-1)} + b_{age} + b_{(age+1)})/3$, where b is the original raw value.[2] Finally, we compute statistics only if the sample is large enough to draw meaningful conclusions, which is relevant for older households in the analysis of mortgage and student debt.

[1] The average cash holdings in SCPC are computed without taking into account the largest 2% of holdings. This is motivated by the existence of relatively limited number of households saving predominantly in cash. Their behavior distorts though the average values of cash holdings.

[2] For the first and last observation in a series we use directly their raw values: $\beta_{22} = b_{22}$ and $\beta_{79} = b_{79}$.

	SCF 2010			SCF 2013			SCF 2016		
	median	mean	fraction positive	median	mean	fraction positive	median	mean	fraction positive
income	53,350	86,175	100%	51,154	86,037	100%	54,682	96,353	100%
net worth	68,565	387,935	87.8%	68,563	394,022	88.6%	82,066	469,024	89,7%
liquid wealth	1,860	101,151	73.0%	2,281	110,495	76.1%	2,628	147,719	74.7%
illiquid wealth	63,556	286,781	77.5%	61,609	283,527	76%	73,000	321,305	78.1%
net housing	26,528	109,410	61.0%	24,747	104,245	59.7%	33,000	119,021	61.6%
mortgage debt	0	83,963	49.4%	0	73,840	45.4%	0	69,519	43.6%
cash, checking, saving, call & MM accounts	3,034	32,381	91.4%	3,639	34,589	91.9%	4,204	39,358	91.9%
credit card debt	0	3,267	41.3%	0	2,399	40.1%	0	2,653	46%
directly held stocks	0	30,255	14.7%	0	36,089	13.9%	0	39,030	13.7%
directly held bonds	0	9,031	1.5%	0	7,719	1.3%	0	7,745	1.1%
mutual funds	0	32,752	8.2%	0	34,497	8.0%	0	64,239	9.6%
retirement accounts	1,990	97,752	52.9%	1,031	107,457	52%	2,000	120,277	53.7%
net car wealth	10,722	16,793	86.5%	10,105	10,105	85.2%	9,596	15,678	83.1%
life insurance	0	5,824	19.0%	0	7,010	18.7%	0	6,328	18.8%
student loan	0	5,783	20%	0	6,470	21.1%	0	8,133	23.5%

Table 2.1: *Household income, liquid and illiquid wealth holdings and portfolio composition in SCF 2010, 2013 and 2016 data.*

Aggregate portfolio composition Table 2.1 lists basic statistics regarding income and the portfolio composition of US households across the post-crisis years. The SCF data clearly shows the repercussions of the crisis: a stagnation, or even decrease, of mean and median income and net worth between 2010 and 2013. The latest wave in 2016 confirms recovery, although together with slightly rising inequality as mean to median ratios of income and net worth increased, respectively, from 1.62 and 5.66 in 2010 to 1.76 and 5.72. Another noticeable fact is a clear deleveraging of mortgage debt. The household portfolio composition has not changed significantly in comparison to SCF 1989-2010 data presented by KVW. The largest component of net wealth of the median household is its housing, with all financial assets, but the retirement accounts, being zero. Finally, an increasing mean amount and incidence of student loans, already noted as a long-run trend in Brown, Haughwout, Lee, Scally, and van der Klaauw (2014), is confirmed by the newest SCF data.

***HtM* and *nHtM* incidence** Using the definitions listed in the preceding section we can compute the incidence of different household types, which are shown in Table 2.2. The *hand-to-mouth* phenomenon is rather stable and robust across time. Despite being hit by a large economic crisis, the fraction of *HtM* households has not changed dramatically and it fluctuates around 30% of all households, which is similar to the pre-crisis levels reported by KVW. A small decline in the number of *HtM* consumers may be caused by deleveraging of mortgage debt as debt is highly correlated with the *HtM* status (see KVW for a discussion). The ratio of *wealthy* to *poor HtM* consumers is 1.5 to 1 and is stable, both in the pre- and post-crisis period.

The table contains also additional groups of households: the *wealthy nHtM* and the *poor nHtM*. The *pnHtM* are those who have simultaneously liquid wealth with zero or negative illiquid wealth and the *wnHtM* are those who have both positive liquid and illiquid wealth stocks. The *poor nHtM* consumers form a small fraction of the *nHtM* households, equal to about 15% of them.

	SCF 2010	SCF 2013	SCF 2016
HtM	32.3%	30.2%	28.6%
wHtM	19.7%	17.3%	17.2%
pHtM	13%	12.9%	11.4%
nHtM	67.3%	69.8%	71.4%
wnHtM	57.7%	58.8%	60.9%
pnHtM	9.4%	11%	10.5%

Table 2.2: *Incidence (fraction) of different household types in SCF 2010, 2013 and 2016 data.*

Figure 2.2 depicts the incidence of all household types conditional on age. Qualitative results regarding the *wHtM*, *pHtM* and *nHtM* incidence are nearly identical as in KVW. The *poor HtM* status is monotonically decreasing in age, while the probability of having the *wealthy HtM* status increases in the beginning of the life cycle and peaks up at around 40 years of age to decrease later. A difference between the post-crisis and pre-crises data is the fact that both graphs are less steep, i.e. the *pHtM* starts at the fraction of around 20% instead of 30% and decreases at a slower pace. The same applies to the incidence of *wealthy HtM* households that peaks at about 20%, which is about 5pp lower than in KVW, but then decreases more slowly. The fraction of the *wealthy nHtM* consumers increases monotonically through all age groups whereas the incidence of *poor nHtM* decreases monotonically. The dynamics of the *poor nHtM* status differs though through age as it decreases much faster between 22 and 30-35 years of age and after that its dynamics resembles the one of the poor HtM consumers.

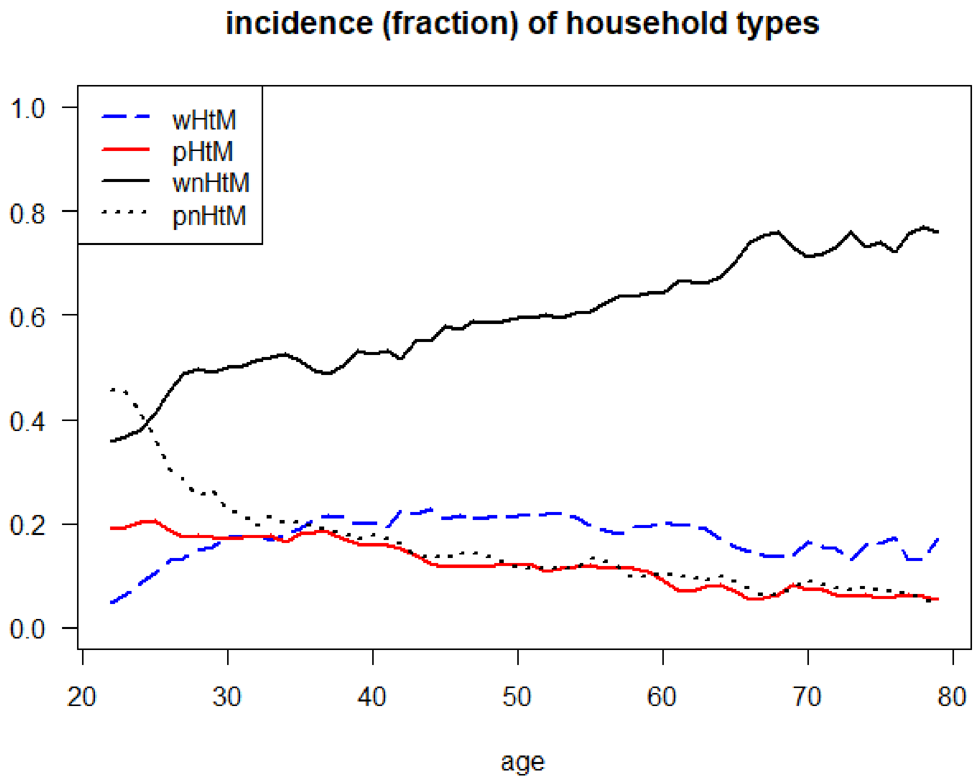


Figure 2.2: *Incidence (fraction) of different household types conditional on age. Pooled SCF 2010, 2013 and 2016 data.*

Demographic and social characteristics The most important sociodemographic indicators of all household types are depicted in Figure 2.3.[3] It is particularly noteworthy that in all graphs, with the exception of the average number of children in a household, *poor nHtM* and *poor HtM* are virtually identical across all age groups. The *wealthy nHtM* households differ considerably in all categories from the *poor HtM/nHtM* consumers: they have less children, are more often married, more often is the head of a household white and male, and they work more often, claiming less benefits. The *wealthy HtM* represent intermediate positions in all categories but they resemble the *poor HtM* in terms of the average number of children. Moreover, they work with the same probability as the *wealthy nHtM*.

[3]'Working in some way' means working full-time, part-time or being self-employed. Government benefits do not contain pension benefits.

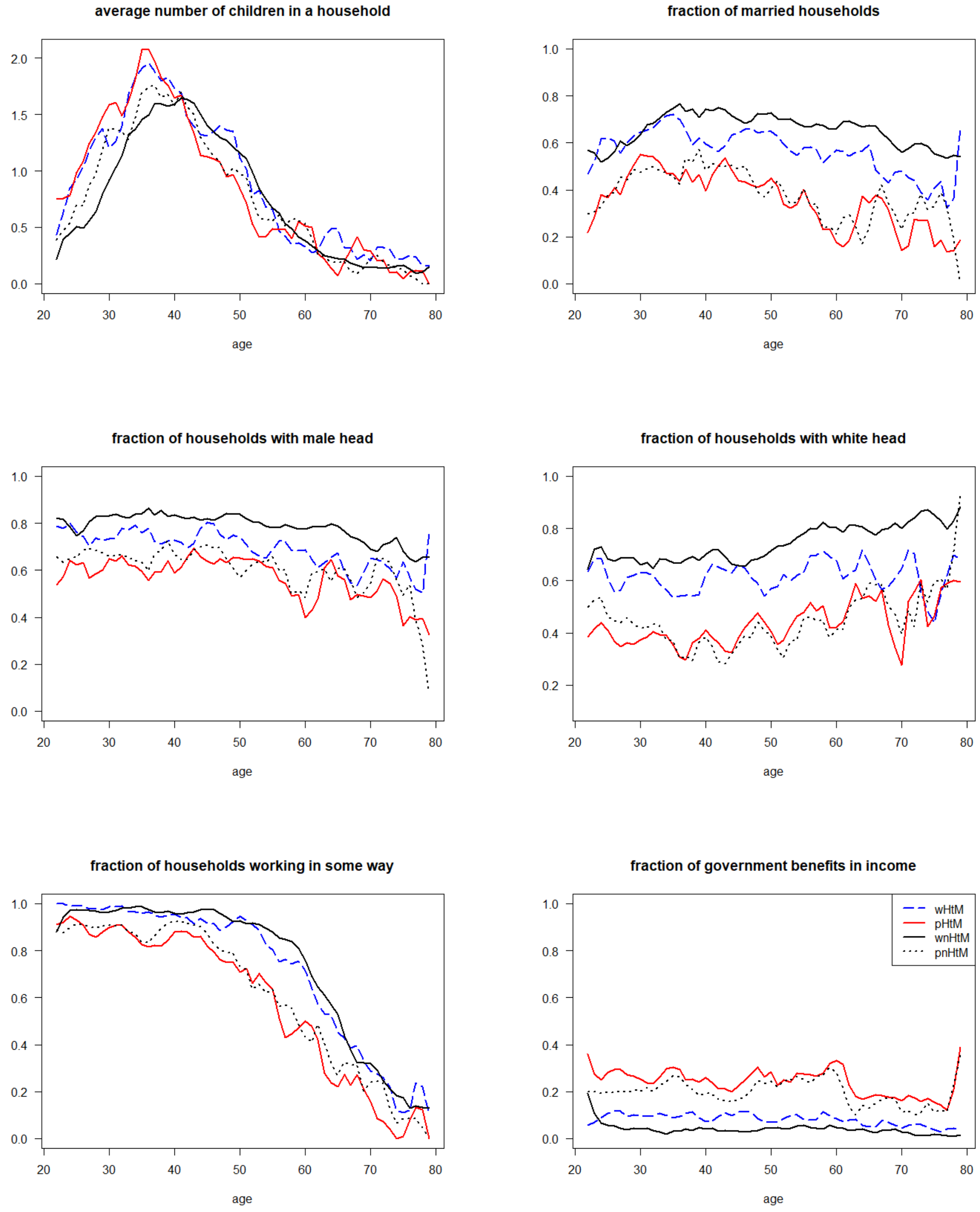


Figure 2.3: *Demographic and social characteristics of different household types conditional on age. Pooled SCF 2010, 2013 and 2016 data.*

Figure 2.4: *Wealth portfolio composition of different household types conditional on age. Pooled SCF 2010, 2013 and 2016 data.*

Portfolio characteristics Graphs of Figure 2.4 show various parts of the portfolio composition. Similarly as in the other figures, qualitative differences between the post- and pre-crises data are immaterial. It is particularly important to note that *poor nHtM* consumers, although fulfilling the condition to be listed as *non-hand-to-mouth*, own very little liquid wealth. Obviously, as an immediate consequence of the definition, they do not have any positive illiquid wealth, resembling thus the *poor HtM*. A striking fact in the comparison with the pre-crisis data is the larger stock of mean liquid wealth, which is mainly dictated by the increase in the value of the stock of financial assets among the richest households. Simultaneously, median liquid asset wealth peak up at 20-25k USD and median holdings of illiquid wealth reach maximally 300k USD. Average values of net assets in vehicles point out the similarity between the *poor HtM* and *poor nHtM* and show that the portfolio position in net vehicles of the *wealthy HtM* is between these two groups and the *wealthy nHtM*.

Income, education and credits Figure 2.5 reveals that *poor nHtM* and *poor HtM* are very similar in their income and education grade[4], conditional on age. The only exception is the educational grade before 30-35 years of age, which is systematically larger for *poor nHtM*. The shape of income paths indicates that *pHtM* and *pnHtM* are more likely to expect stagnating income in the future, conditional on staying in their current household type, whereas *wHtM* and *wnHtM* have a much steeper income path. Qualitatively, Figure 2.5 does not exhibit any differencess with Figure 2.1, which presents KVW results.

The right side of Figure 2.6 delves into the size and incidence of student loans. Contrary to phenomena depicted in other figures, student loans are most relevant for young age groups. Although the *poor nHtM* resemble *poor HtM* with respect to the probability of having a student loan, the average size of their loans is larger than for *wHtM*. As with most other characteristics, the *wnHtM* are distant from other households by exceeding them with both the loan size and probability. The *wHtM* are nearly identical to the *wnHtM* with their probability of having educational debt but its size is smaller on average.

The left side of Figure 2.6 presents statistics on mortgage credits. The revealed patterns suggest that there are two distinct groups of households when it comes to the probability of having a mortgage credit: *poor HtM* and *poor nHtM* are much less likely to have such a liability than the *wealthy HtM* and *wealthy nHtM*. The difference between the *wealthy* types of households emerges though in the average size of a credit, which is larger for the *wealthy nHtM*. Due to smaller sample size, it is difficult to draw conclusions from volatile data on average liabilities among indebted *poor* households.

[4]Education grade in the graph is a qualitative indicator for attaining a given educational level. Values below 8 mean different years of completed primary or high school education, 8 means high school diploma, 9 is some college without degree, 10 and 11 are vocational or academic college programs, 12 is a bachelor degree, 13 is a master degree and 14 is a PhD degree.

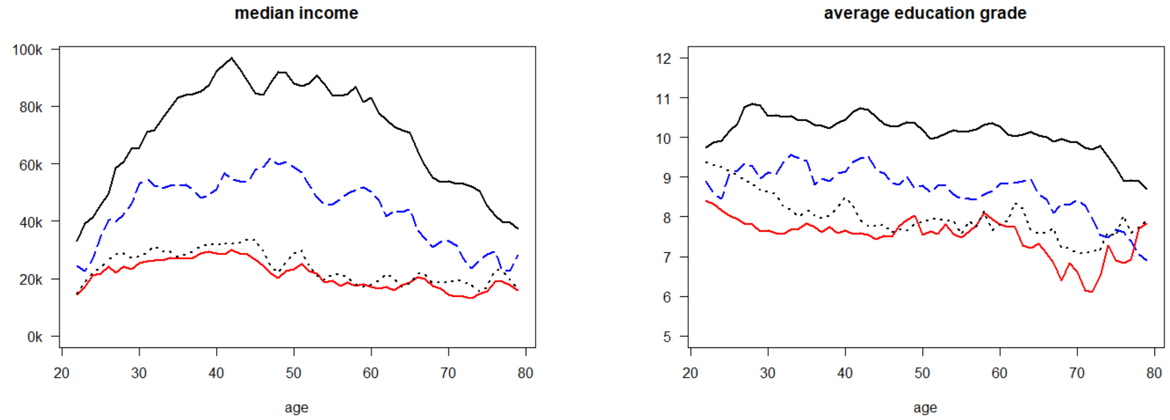


Figure 2.5: *Income and education of different household types conditional on age. Pooled SCF 2010, 2013 and 2016 data.*

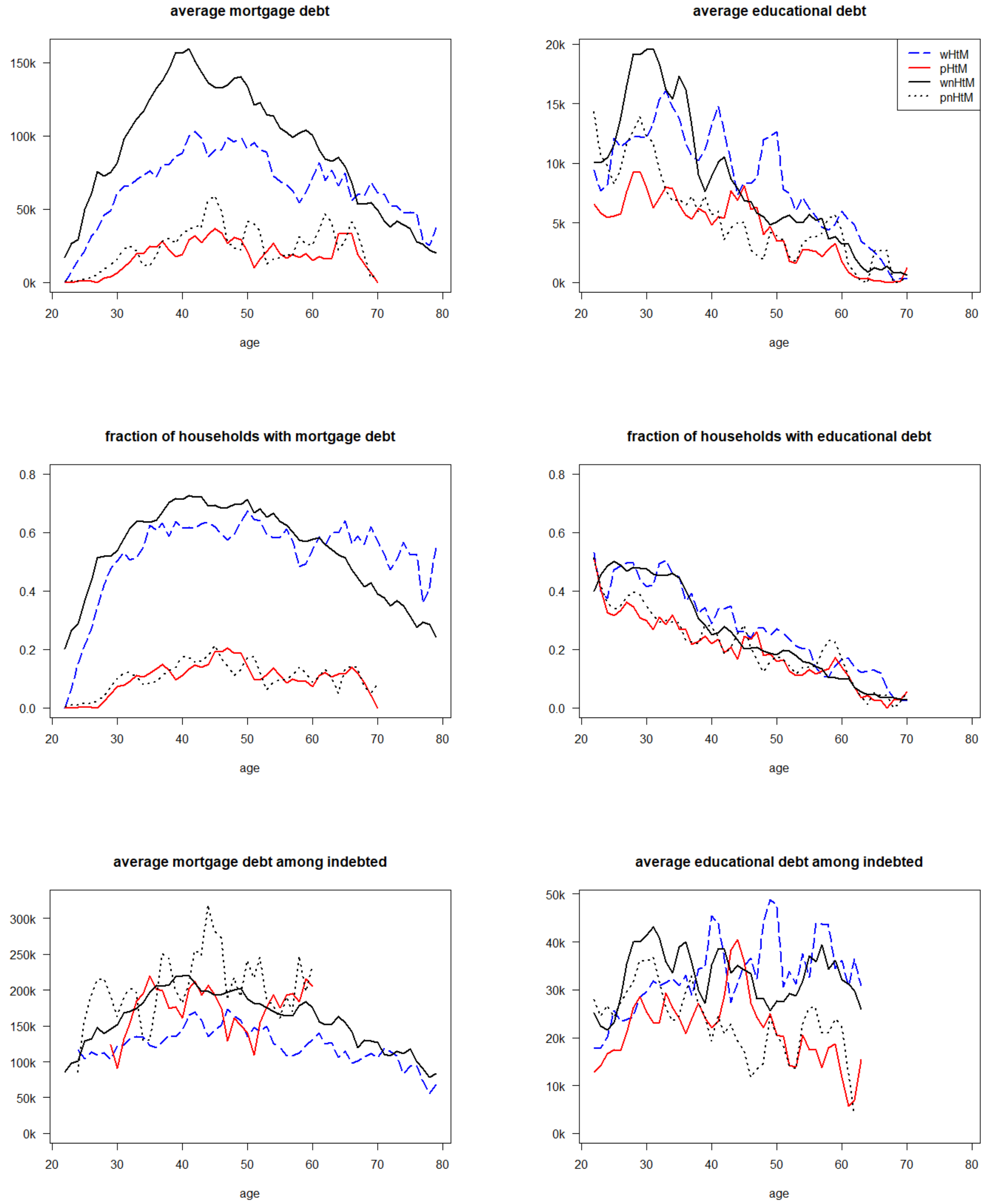


Figure 2.6: *Mortgage and student debt characteristics of different household types conditional on age. Pooled SCF 2010, 2013 and 2016 data.*

Transitions between household groups The SCF data set is not panel but a repeated cross sectional survey. This makes it impossible to trace given households over time and analyze directly the dynamic evolution of their *hand-to-mouth* status. KVW use the available SCF panel data between years 2007 and 2009 to measure the probability of transition between the *pHtM*, *wHtM* and *nHtM*. Table 2.3 displays their results and shows that *wHtM* status is most transitory with about 45% of these household becoming *nHtM* after two years. The *pHtM* were less likely to change their type to *nHtM* (only 1/3 underwent this transition) and the *non-hand-to-mouth* status was the most stable one.

07 → 09	*wHtM*	*pHtM*	*nHtM*
wHtM	45.5%	10.1%	44.4%
pHtM	12.7%	54.8%	32.6%
nHtM	12.9%	5.5%	81.6%

Table 2.3: *Transition matrix between household types for years 2007-2009 based on the Panel of the SCF. Each number is a fraction of a given row type in 2007 being a given column type in 2009. Source: Kaplan et al. (2014).*

Conclusions The statistical analysis of the empirical data leads to the following conclusions. First, the *hand-to-mouth* behavior is a stable and robust phenomenon. The latest, post-crisis data does not change qualitatively any results of KVW regarding the incidence, demographic characteristics nor portfolio composition. Existing differences are of quantitative nature, comprising effects of housing deleveraging and increases in the value of stocks of financial assets.

Second, the statistical analysis contributed to better understanding of interdependencies between student loans, and therefore human capital accumulation, and *hand-to-mouth* status. The analysis demanded a further division of households: the *non-hand-to-mouth* consumers were divided into the *poor* and the *wealthy* ones. In the groups above 35 years old the *poor nHtM* resemble almost perfectly the *poor HtM* (in some sociodemographic statistics they resemble them even earlier), which is intuitive in view of the fact that the *hand-to-mouth* definition we use is rather liberal and some truly *poor HtM* households, having no liquid nor illiquid assets, may be considered as *poor nHtM*. Moreover, a small fraction of the *poor nHtM* are those, whose mortgage surpasses the value of housing implying negative stock of illiquid assets. Among youngest households, who are below 30 years of age, the *pnHtM* are those who have large student loans, comparable to the ones of *wealthy HtM* or even *wealthy nHtM*. Over time, it seems that *poor nHtM* turn into *wealthy HtM* (possibly by taking a mortgage loan) or directly into the *wealthy nHtM*. The group of the *wealthy HtM* consumers is intermediate as it has a similar probability of having a student loan as the *wealthy nHtM* and the average size of a loan among the indebted is between the *wnHtM* and *pHtM*.

2.3 Explaining the household problem: exposition of the model

In order to show that wealthy and poor educated consumers may exhibit *hand-to-mouth* behavior and to track what characteristics of the consumers drive this kind of behavior, it is useful to take a look at a simple three period model with endogenous investment in human capital. The model is an extension of the model presented by Kaplan et al. (2014), which is, as in Section 2.2, called 'KVW' throughout the text.

Let us consider a three-period model, $t \in \{0, 1, 2\}$, in which a consumer maximizes his utility from consumption in the intermediate and final period by making decisions about the portfolio allocation (in the initial period) and consumption and labor supply (in the intermediate and final period). As in KVW, we assume that in the initial period the consumer does not consume and makes the decision only about the division of the initial resources ω (that are assumed to be strictly positive, $\omega > 0$). The consumer decides whether to invest the resources in liquid assets m_1, which can be consumed in the intermediate period 1, and illiquid assets a, which can be consumed only in the final period 2. Illiquid assets, such as housing and pension savings, earn interest factor $R > 1$, while liquid assets have an interest factor that is normalized to 1. Moreover, illiquid assets, a major part of which is housing (see the empirical analysis in KVW), generate direct utility in both periods, as, for example, in Kaplan and Violante (2014). A novelty in comparison to the KVW model is the possibility of investment in human capital h that generates income in both periods through the production function $A \cdot f(h, l_t)$. The human capital has properties of an asset but it exhibits certain distinctive features: it is more difficult to use as collateral than an illiquid asset and it only generates resources if labor effort l_t is exerted, which generates disutility because of a loss of leisure. The parameter A is interpreted as the innate ability affecting the productivity of an individual.

Formally, the agent's maximization problem is:

$$\max_{c_1, c_2, l_1, l_2, a} \quad u(c_1, l_1, a) + \beta \cdot u(c_2, l_2, a), \quad \beta > 0$$

$$s.t.$$

$$c_1 = A \cdot f(h, l_1) + m_1 - m_2$$

$$c_2 = g \cdot A \cdot f(h, l_2) + m_2 + R \cdot a$$

$$\omega = m_1 + a + h$$

with $c_1 \geq 0,\ c_2 \geq 0,\ l_1 \geq 0,\ l_2 \geq 0,\ m_1 \geq 0,\ m_2 \geq 0,\ a \geq 0,\ h \geq 0, \omega > 0, A > 0$. Households cannot take any debt.[5] The parameter $g \geq 0$ models the possibility of having an earnings path that is not constant over the life cycle - if $g > 1$, then the labor income path is increasing, ceteris paribus. β is the standard patience parameter. For simplicity, we assume that human capital and illiquid assets do not depreciate at all.

Schematically, the model can be depicted as follows:

	initial period	*intermediate period*	*final period*
	$t = 0$ $\longrightarrow$	$t = 1$ $\longrightarrow$	$t = 2$
decisions :	$a,\ h,\ m_1$	$m_2,\ l_1$	l_2
production :		$A \cdot f(h, l_1)$	$g \cdot A \cdot f(h, l_2)$
utility :	$u(c_1, l_1, a) + \beta u(c_2, l_2, a)$	$u(c_1, l_1, a) + \beta u(c_2, l_2, a)$	$u(c_2, l_2, a)$
consumption :		$\underbrace{A \cdot f(h, l_1) + m_1 - m_2}_{c_1}$	$\underbrace{g \cdot A \cdot f(h, l_2) + m_2 + Ra}_{c_2}$

The utility and production functions are assumed to fulfill the following conditions[6] for $t \in \{1, 2\}$ and for all combinations of $\{h, l_t\}$ with $h > 0$ and $l_t > 0$:

$$
\begin{aligned}
u_{c_t}(c_t, l_t, a) > 0, \quad & u_{c_t c_t}(c_t, l_t, a) < 0, & f_h(h, l_t) > 0, \quad & f_{hh}(h, l_t) < 0, \\
u_{l_t}(c_t, l_t, a) < 0, \quad & u_{l_t l_t}(c_t, l_t, a) < 0, & f_{l_t}(h, l_t) > 0, \quad & f_{l_t l_t}(h, l_t) < 0, \\
u_a(c_t, l_t, a) > 0, \quad & u_{aa}(c_t, l_t, a) < 0 & f_{l_t h}(h, l_t) \ = \ & f_{h l_t}(h, l_t) > 0
\end{aligned}
$$

Both production factors are necessary for production: $f(0, l_t) = f(h, 0) = f(0, 0) = 0$ and, therefore, $f_h(h, 0) = f_{l_t}(0, l_t) = 0$.

[5]This simplifying assumption can be obviously relaxed by introducing a borrowing limit, without changing results qualitatively. Potentially, households could borrow against some of their future labor earnings or their illiquid assets. The empirical data shows negative positions not only of liquid but also illiquid assets. The latter arise because of negative shocks to house prices, which results in mortgage credits surpassing house values. The model abstracts from such phenomena and does not allow for negative stocks of illiquid assets.

[6]Throughout the text partial derivatives are written like this: $\frac{\partial^2 f(x,y)}{\partial x \partial x} = f_{xx}(x, y)$. Sometimes, arguments of a function are dropped to conserve space and make the text more readable, using the notation $f_{xx}(x, y) = f_{xx}$. Note also that, for example, u_{c_1} pertains to $u_{c_1}(c_1, l_1, a)$ and not to $u_{c_1}(c_2, l_2, a)$.

Furthermore, the utility function is assumed to be separable:[7] $u(c_t, l_t) = U_U(c_t) - U_V(l_t) + U_W(a)$. Thus, the cross-derivatives are zero, in particular $u_{c_t l_t}(c_t, l_t, a) = 0$. Moreover, both $u(c_t, l_t, a)$ and $f(h, l_t)$ fulfill the regularity conditions:

$$\lim_{h\to 0} f_h(h, l_t) = \infty, \quad \lim_{l_t\to 0} f_{l_t}(h, l_t) = \infty, \quad \lim_{l_t\to 0} u_{l_t}(c_t, l_t, a) = 0$$
$$\lim_{c_t\to 0} u_{c_t}(c_t, l_t, a) = \infty, \quad \lim_{a\to 0} u_a(c_t, l_t, a) < \infty$$

The regularity conditions are crucial in ruling out corner solutions of zero labor supply and zero human capital as marginal productivity becomes infinite as $h \to 0$ or $l_t \to 0$ and marginal disutility of labor supply is zero as $l_t \to 0$. The utility function cannot though exhibit infinite marginal utility of illiquid assets for $a \to 0$ as it would rule out the corner solution of zero illiquid asset holdings. The economic interpretation of this property indicates that illiquid assets (housing) are treated as luxury goods.

An example for such utility and production functions are the commonly used CRRA/CIES utility function with labor disutility entering the utility as in King, Plosser, and Rebelo (1988), with an additional term for the utility flow derived from the illiquid asset, and a standard Cobb-Douglas production function:

$$u(c_t, l_t, a) = \frac{c_t^{1-\sigma} - 1}{1-\sigma} - \frac{\mu}{\alpha} l_t^{\alpha} + \frac{\psi}{\chi}(a+\varphi)^{\chi}, \quad \sigma \geq 0,\ \alpha > 1,\ \mu > 0,\ \psi > 0,\ \varphi > 0,\ \chi \in (0,1)$$
$$A \cdot f(h, l_t) = A \cdot h^{\gamma} l_t^{1-\gamma}, \quad A > 0,\ \gamma \in (0,1)$$

2.4 Analytical solution of the model

The model can be solved using the method of backward induction. We begin with the final period 2, in which the labor supply l_2 is chosen, then we proceed to the intermediate period 1, in which agents choose labor supply l_1 and liquid savings m_2. Finally, we get to the initial period 0, in which a decision about the composition of the portfolio consisting of a, m_1 and h is made (with $a + m_1 + h = \omega$). The model is fully deterministic - there is no uncertainty about any of the variables. There are no restrictions on the upper limit of the labor supply to avoid further complications, yet the presence of increasing disutility of labor ensures its finiteness. The *hand-to-mouth* (*HtM*) behavior is defined in the same way as in the KVW model: if a consumer decides in the intermediate period to hold no liquid assets $m_2 = 0$ then he is called *HtM*, while if $m_2 > 0$ then he is called

[7] This assumption is useful in proving the concavity of the problem.

non-hand-to-mouth (*nHtM*). The decision in the initial period determines whether a consumer is *wealthy* (*n*)*HtM* or *poor* (*n*)*HtM*, which depends on whether the illiquid asset stock a is positive or zero. The backward induction solution is given as follows:[8]

Final period 2:

$$\max_{l_2} \quad u(\underbrace{g{\cdot}A{\cdot}f(h,l_2)+m_2+R\cdot a}_{c_2},\ l_2,\ a)$$

$$s.t.$$

$$l_2 \geq 0$$

The assumptions that the utility and production function satisfy the regularity conditions rule out zero labor supply[9] and because of that the analysis below focuses only on the choice of positive labor supply. Then, the optimality conditions imply $u_{c_2}\cdot f_{l_2}\cdot g{\cdot}A = -u_{l_2}$. This is a standard intratemporal optimality condition, which equates the marginal rate of substitution $-\frac{u_l}{u_c}$ to the relative price of labor w, which is equal in the competitive labor market to the marginal product of labor $g{\cdot}A{\cdot}f_{l_2}$. This condition determines labor supply l_2 for any amount of assets m_2 and a, which were chosen in the intermediate and initial period and, together with the labor supply l_2, determine the consumption c_2.

Intermediate period 1:

$$\max_{m_2,l_1} \quad u(\underbrace{A{\cdot}f(h,l_1)+m_1-m_2}_{c_1},\ l_1,\ a)+\beta\cdot u(\underbrace{g{\cdot}A{\cdot}f(h,l_2^{opt})+m_2+R\cdot a}_{c_2},\ l_2^{opt},\ a)$$

$$s.t.$$

$$l_1 \geq 0, \quad m_2 \geq 0$$

The same assumptions as in the final period 2 rule out zero labor supply and lead to an analogous intratemporal optimality condition $u_{c_1}{\cdot}A{\cdot}\, f_{l_1} = -u_{l_1}$. The optimality condition regarding m_2 leads, precisely like in the KVW model, to the so called 'short-run' Euler equation $u_{c_1} \geq \beta u_{c_2}$. If $m_2 > 0$, then $u_{c_1} = \beta u_{c_2}$: the consumer is unconstrained and the Euler equation is satisfied. If $\beta = 1$, this implies perfect consumption smoothing, i.e. $c_1 = c_2$. If $\beta < 1$, then $c_1 > c_2$ and if $\beta > 1$, then $c_1 < c_2$.

[8]The formal analysis is presented in the Appendix, see Section 2.7.1. A discussion on sufficiency of first-order conditions is also given in the Appendix, see Section 2.7.3

[9]For a detailed analysis of this issue see Section 2.7.2 of the Appendix.

If $m_2 = 0$, then a consumer is constrained and cannot fully adjust consumption to the preferred level. The decision about m_2 is affected by the path of income in $t = 1$ and $t = 2$ and by preferences represented by parameter β. The more increasing (or 'less decreasing') the path of income the more likely it is that $m_2 = 0$ for a given β. If $m_2 = 0$ then the household is called *hand-to-mouth*, while households with positive m_2 are named *non-hand-to-mouth*.

Initial period 0:

$$\max_{a,h,m_1} \quad u(\underbrace{A \cdot f(h, l_1^{opt}) + m_1 - m_2^{opt}}_{c_1},\ l_1^{opt},\ a) + \beta \cdot u(\underbrace{g \cdot A \cdot f(h, l_2^{opt}) + m_2^{opt} + R \cdot a}_{c_2},\ l_2^{opt},\ a)$$

$$s.t.$$

$$\omega = a + h + m_1$$

$$a \geq 0, \quad h \geq 0, \quad m_1 \geq 0$$

The analysis of the household problem in the initial period has to distinguish two cases: $m_2 > 0$ and $m_2 = 0$ or, in other words, *nHtM* and *HtM* households, respectively. Regularity conditions for the production function rule out zero human capital accumulation in presence of positive labor supply for both *HtM* and *nHtM* households, while a choice of zero human capital and zero labor supply is also ruled out because of the regularity conditions - therefore the analysis below pertains only to the situation of positive labor supply and positive human capital. Details of the analysis are presented in Section 2.7.2 of the Appendix.

Ruling out the cases with no human capital for *HtM* households leaves us with 4 potential subcases - a combination of zero or positive liquid assets m_1 and illiquid assets a in the presence of positive human capital h. The model thus may generate both *wealthy* and *poor HtM* consumers. The *wealthy HtM* are those having positive illiquid assets in the final period and their stock of liquid assets in the intermediate period might be both positive or zero. Similarly, the *poor HtM* may have no illiquid assets m_1 or they might have a positive stock of them but they have always no illiquid assets a.

The model rejects another two potential outcomes for *nHtM* households as it is not optimal to save in the asset that is return dominated - we can rule out $m_1 > 0$ for these households. The *nHtM* households want to save some income in the intermediate period 1 in order to have a higher consumption in the final 2nd period. For these households $m_1 > 0$ and $m_2 > 0$ cannot be the optimal choice because the illiquid assets have the higher return factor $R > 1$. Hence, *nHtM* consumers choose their portfolio in the initial period so that it consists entirely of human capital or divide it between illiquid assets and human capital depending on the relation between the marginal return on education and on the illiquid asset a. In the latter case the optimal choice between the illiquid asset a, earning the gross return after two periods and generating direct utility in both periods, and h, earning return of $f_h(h, l_i)$ in periods 1 and 2, is determined by $A \cdot u_{c_1} \cdot [f_h(h, l_1) + g \cdot f_h(h, l_2)] = R \cdot u_{c_1} + u_a \cdot (1 + \beta)$. The corner solution with $a = 0$ and $h > 0$ is possible if allocation of all the initial wealth ω to human capital gives higher return than for the illiquid asset (the return includes R and the additional utility flow u_a).

Figure 2.7 depicts the six potential outcomes that can be chosen by agents as solutions to their optimization problem. Note that the graph depicts only solutions with $h > 0, l_1 > 0, l_2 > 0$ (in order to make the graphic more readable, decisions about values of h, l_1 and l_2 are not given explicitly; solutions with combinations of $h = 0$ and $l_1 = l_2 = 0$ cannot occur given the assumptions made).

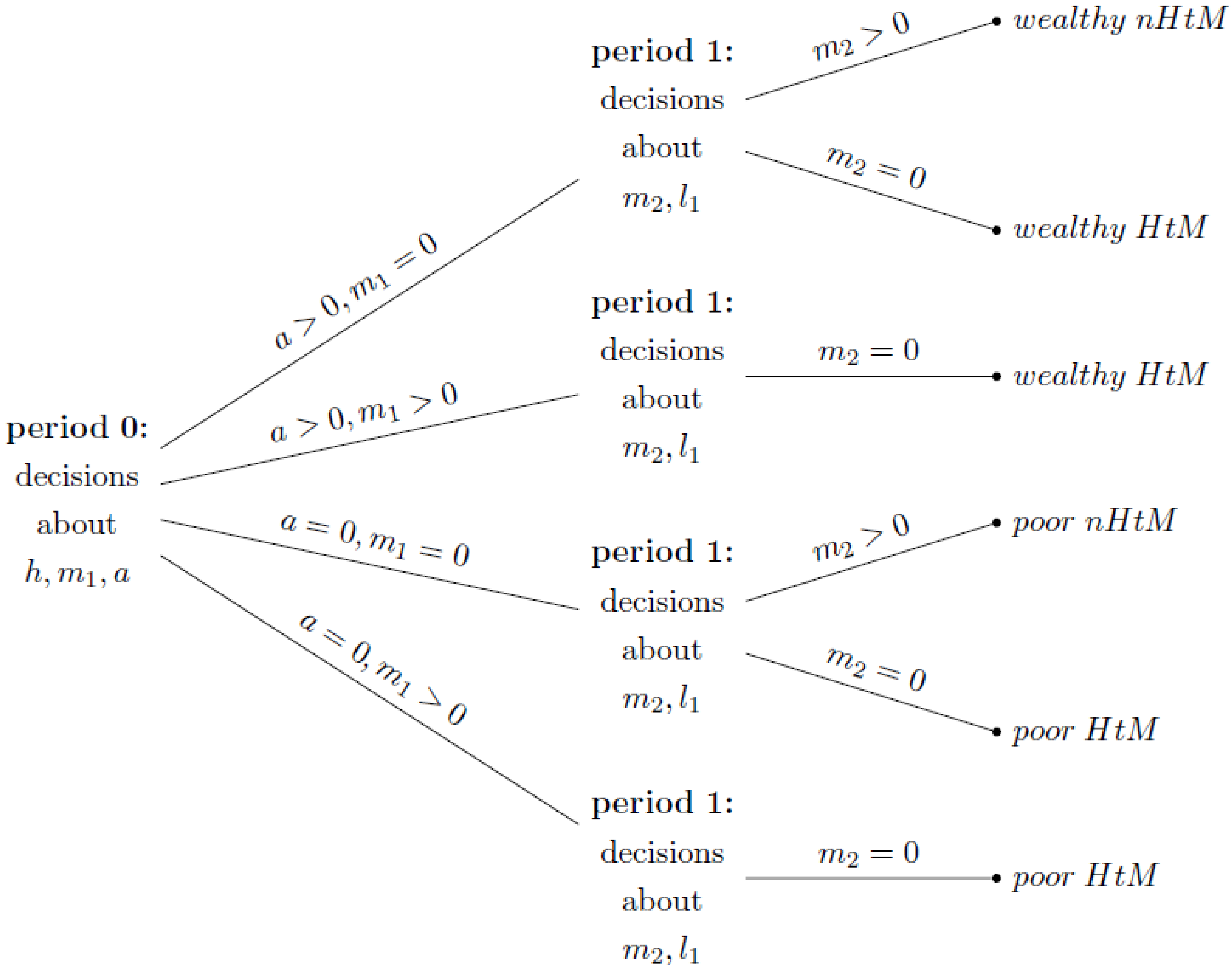


Figure 2.7: *Potential solutions to the agent's decision problem. Note that all six solutions involve* $h > 0, l_1 > 0, l_2 > 0$.

Unfortunately, the model, although conceptually simple, is rich in nonlinearities and constraints and therefore does not give a tractable closed-form solution, even for simple utility and production functions and for a model with exogenous labor supply. Nevertheless, the possibility of existence of *nHtM* (both *poor* and *wealthy*), *pHtM* and *wHtM* consumers is shown analytically. Their emergence depends on the slope of the endogenous income path, which means that the characteristics of consumers like time preference β, innate ability A or initial endowment ω are crucial in making households' decisions about their (n)*HtM* status or the decisions about accumulating illiquid assets a and human capital h.

Some of the results of the model are similar to KVW. Firstly, both *HtM* and *nHtM* households emerge, where one of the determinants of their appearance is the slope of the income path. If the household has higher total income in the final period than in the intermediate one, $y_1 < y_2$, then it is more likely to be *HtM* as it is more likely to decide to save no liquid assets. If $y_1 > y_2$ instead, then the household is more likely to be *nHtM* because of holding a positive stock of liquid assets.

Secondly, the possibility of saving both in liquid assets m_1 and m_2 and, simultaneously, in illiquid assets is ruled out because of the fact that the return factor of the liquid asset is lower than of the illiquid one, $R > 1$, which is again in line with KVW. Furthermore, households may be willing to attach less value to consumption smoothing if they receive sufficiently high return on the illiquid assets, which gives rise to *HtM* households. A discussion of empirical relevance of such behavior can be found in Browning and Crossley (2001).

A novelty in comparison with KVW is the introduction of endogenous labor income, which provides a motive to invest in human capital as a part of the portfolio. Because of the assumptions on the production function, we can rule out zero human capital and also zero labor supply, which is in accordance with empirics.

This makes it possible to have larger sets of *wealthy* and *poor hand-to-mouth* consumer types as, contrary to the KVW model, they can also be households with zero liquid wealth in the first period, i.e. $m_1 = 0$. This is because our assumptions preclude zero income so that consumers can get positive consumption in the intermediate period without a necessity of having positive liquid assets $m_1 > 0$. The assumptions on the production function imply that the set of *nHtM* consumer types is also larger as it consists of those dividing their initial endowment between the human capital and illiquid assets and those investing only in the human capital, deciding to have no illiquid assets. In Section 2.2 these two groups of households were called the *wealthy nHtM* and the *poor nHtM*, respectively.

2.5 Numerical solution of the model

The model is solved using numerical methods. The utility and production functions are parametrized as follows:

$$u(c_t, l_t, a) = \frac{c_t^{1-\sigma} - 1}{1 - \sigma} - \frac{\mu}{\alpha} l_t^{\alpha} + \frac{\psi}{\chi}(a + \varphi)^{\chi}$$

$$A \cdot f(h, l_t) = A \cdot h^{\gamma} l_t^{1-\gamma}$$

The following parameter values are chosen for illustration purposes: $\sigma = 2$, $\alpha = 3$, $\mu = 1$, $\psi = 0.015$, $\varphi = 0.5$, $\chi = 0.3$, $\gamma = 0.5, R = 1.5$. The parameters β (time preference), A (innate ability), ω (initial endowment) and g (slope of the labor income path) are considered as those driving the *HtM* behavior and the model is solved for a variety of their values. The considered values of A and ω are 25 equidistant points from the intervals $[3, 5]$ and $[1, 3]$, respectively. A and ω axes in the figures are depicted with ordinal values that place the parameters in the aforementioned vectors of these parameter values. Parameter g takes 4 values in the set $\{0.5, 0.75, 1, 1.25\}$.

For illustrative purposes a functional dependence is introduced between parameters β and A. The form of the dependence is $\beta = (\Gamma \cdot A)^{\Phi}$, with $\Gamma = 0.25$ and $\Phi = 2$, which generates values of β in the interval $[\frac{9}{16}, \frac{25}{16}]$. First of all, the dependence reduces the degrees of freedom and makes it possible to depict the model graphically in a 3-dimensional parameter space. Secondly, the direct positive relationship between β and A makes it possible to generate highly able (and therefore highly patient) *non-hand-to-mouth* consumers who invest both in human capital and in illiquid assets. The economic reasoning behind such a relationship is the fact that the time preference can be perceived as closely connected with working/learning productivity (see e.g. Cronqvist and Siegel (2015)). Without delving into a discussion about what causes consumers to be more or less patient, it seems credible to assume that those who are more patient are more reliable and productive workers. Therefore, both these characteristics are driven by the exogenously given innate ability, no matter if 'exogenously given' means genetics or upbringing or if the dependence between A and β is a causal one or it is merely a correlation.

The numerical solution of the consumer's problem is presented in Figure 2.8. The graphs show values of the optimal allocation of resources between illiquid assets a (red surface), human capital h (blue surface) and liquid assets m_2 (yellow surface; to improve readability of the graph, the m_2 surface is moved up by 5 units) for different values of the initial endowment ω and innate ability A. Each graph is computed for a different value of the relative productivity between periods g.

Figure 2.9 shows qualitatively the household's status for a given combination of initial endowment ω, innate ability A and the relative productivity parameter g. Households that exhibit *wealthy non-hand-to-mouth* behavior are depicted with black dots, the *poor non-hand-to-mouth* are denoted by black rings, while *wealthy hand-to-mouth* and *poor* *hand-to-mouth* are denoted by blue rings and red dots, respectively.

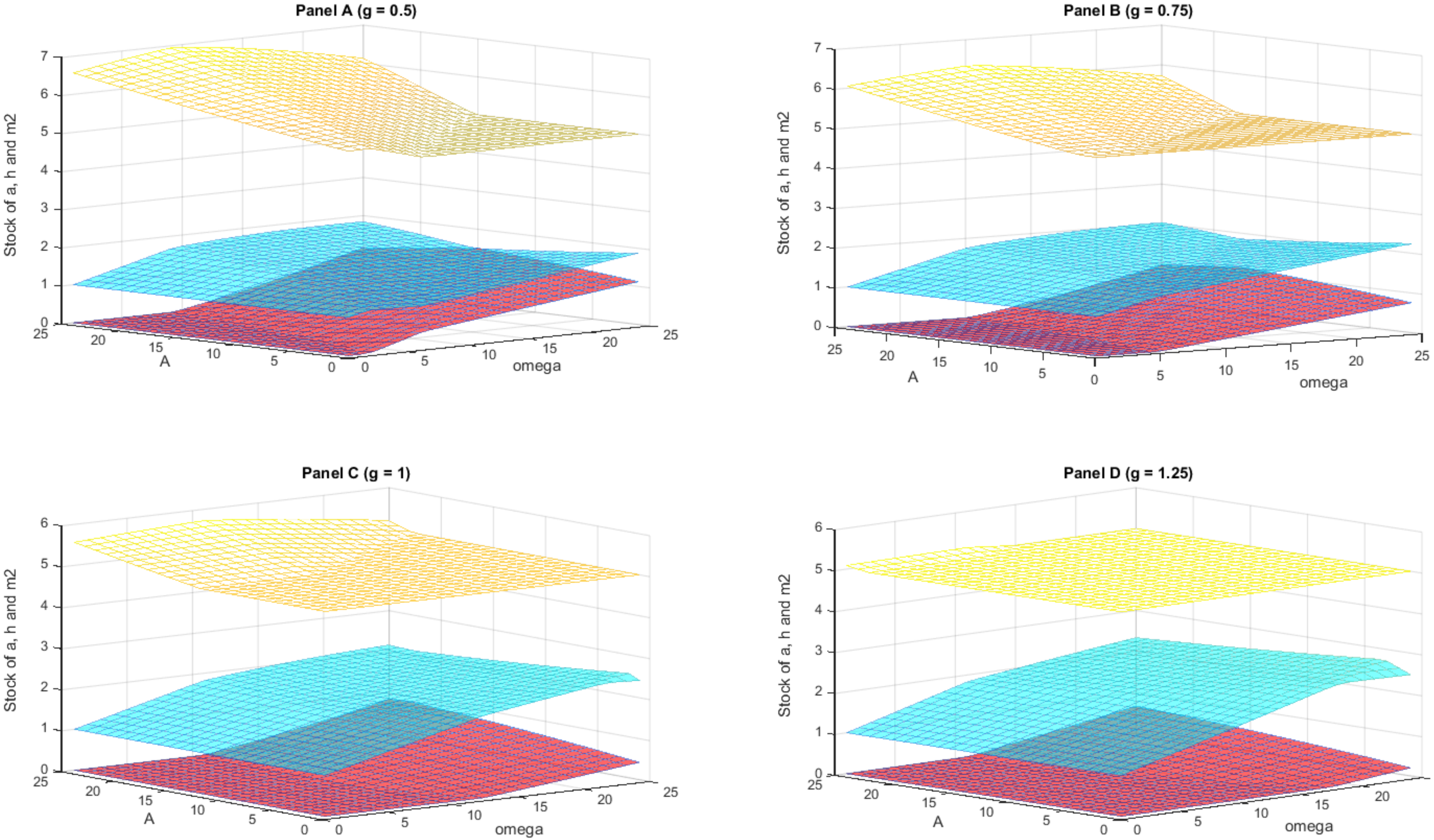


Figure 2.8: *Stocks of illiquid assets a (**red surface**), human capital h (**blue surface**) and liquid assets m_2 (**yellow surface**) for different values of g, ω and A. Note that the m_2 surface is moved up by 5 units to ensure graph readability (i.e. $m_2 = 5$ in the graph means that $m_2 = 0$ in the model).*

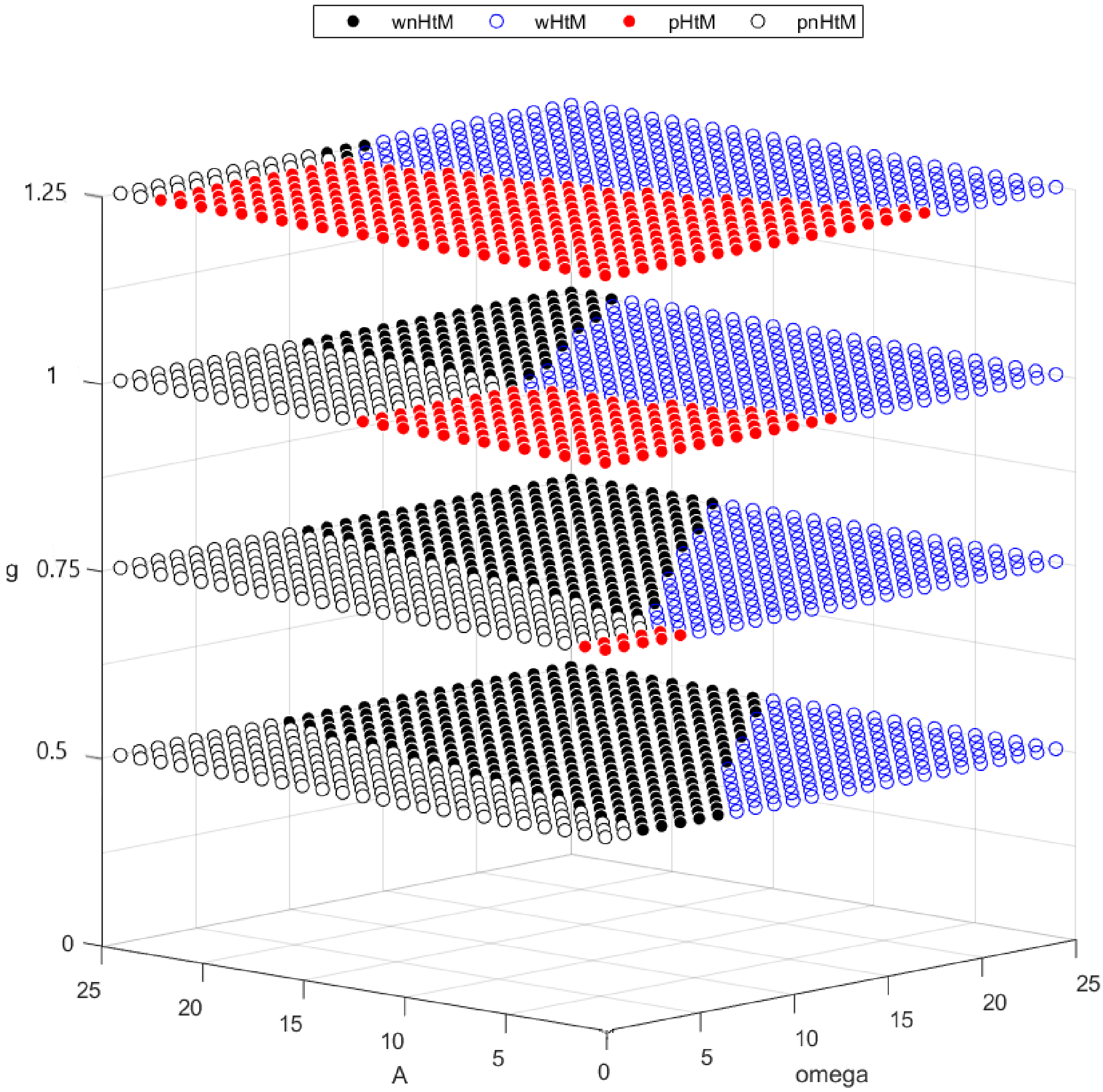


Figure 2.9: *Qualitative distribution of the wHtM (blue rings), pHtM (**red dots**), wnHtM (**black dots**) and pnHtM (black rings) statuses for different values of g (relative labor productivity between periods), ω (initial endowment) and A (innate ability).*

The numerical solutions illustrate how the incidence for all 4 household types depends on exogenous parameters of interest. It is difficult to match precisely relative values of stocks of liquid or illiquid assets and human capital created with so stylized a model to the empirical values, presented Section 2.2. Nevertheless, even in the simple framework the results can match some important features of the data and give unambiguous predictions.

The presented model can be interpreted as an analysis of a households' asset allocation problem over the life cycle, where the life cycle is split into three periods.

The most obvious conclusion is that the occurrence of the *hand-to-mouth* status depends positively on the relation of income in the second period to the one in the previous period (y_2/y_1), which means that the parameter g drives the *HtM* behavior by increasing the income in the final period of the model. Moreover, the higher the initial endowment ω, the higher the stocks of human capital and illiquid assets as households have more resources to invest. However, higher ω lowers, ceteris paribus, the stock of liquid assets m_2 as households can invest more in a that gives a higher interest rate $R > 1$. Finally, higher innate ability increases the parameter β and therefore also the likelihood of being a *nHtM* household as consumers are more patient and are more likely to save in the intermediate period in liquid assets m_2. The innate ability itself does not affect the *HtM* behavior much as it affects simultaneously productivity in both periods.

The economic interpretation of the descriptive results presented in the preceding paragraphs gives the following answer on the incidence of the 4 household types:

- *Poor hand-to-mouth* households comprise those experiencing income growth (i.e. those households with high g) but also those who experience stagnating or even decreasing labor income. Both groups are dominated by initially poor and less able (less patient) consumers. It is noteworthy to point out the low human capital accumulation of these households that equals their entire initial endowment though. Households with the characteristics of these types are depicted in panels 'poor HtM A' and 'poor HtM B' of Figure 2.10.[10]

- *Wealthy hand-to-mouth* households are initially rich and able (patient) with high labor income growth. These households have very large human capital stocks. The *hand-to-mouth* behavior is caused by high innate ability, patience and large g: they all make the human capital productive and desirable as it gives return also in the final period. The liquid assets m_2 are superseded by the illiquid ones and by the liquidity created in the final period by the labor income as the investment in the the human capital gives higher returns. An example of the *wHtM* consumer is depicted in panel 'wealthy HtM' of Figure 2.10.

- *Wealthy and poor non-hand-to-mouth* consumers are a nonhomogeneous group with respect to their initial characteristics yet they all exhibit stagnating or declining

[10]It should be kept in mind that for the *poor HtM* consumers the income graph is identical to the consumption graph, while the illiquid asset line is identical to the liquid asset one as they are both zero. The consumption graph overlaps with the income graph for the *wealthy HtM* as well.

labor income.[11] They can be either initially rich and able with constant labor productivity or initially poorer and less able with declining labor income. The first group (see panel 'wealthy nHtM' in Figure 2.10) accumulates a lot of human capital and illiquid assets while the other one (panel 'poor nHtM') is more constrained because of a lower initial endowment. Both groups are patient enough to accumulate positive savings in the intermediate period and not to exhibit the *HtM* behavior. Households depicted as 'wealthy nHtM' have very large holdings of illiquid assets in comparison with other household types and households described as 'poor nHtM' match the empirical characteristics of their type relatively close.

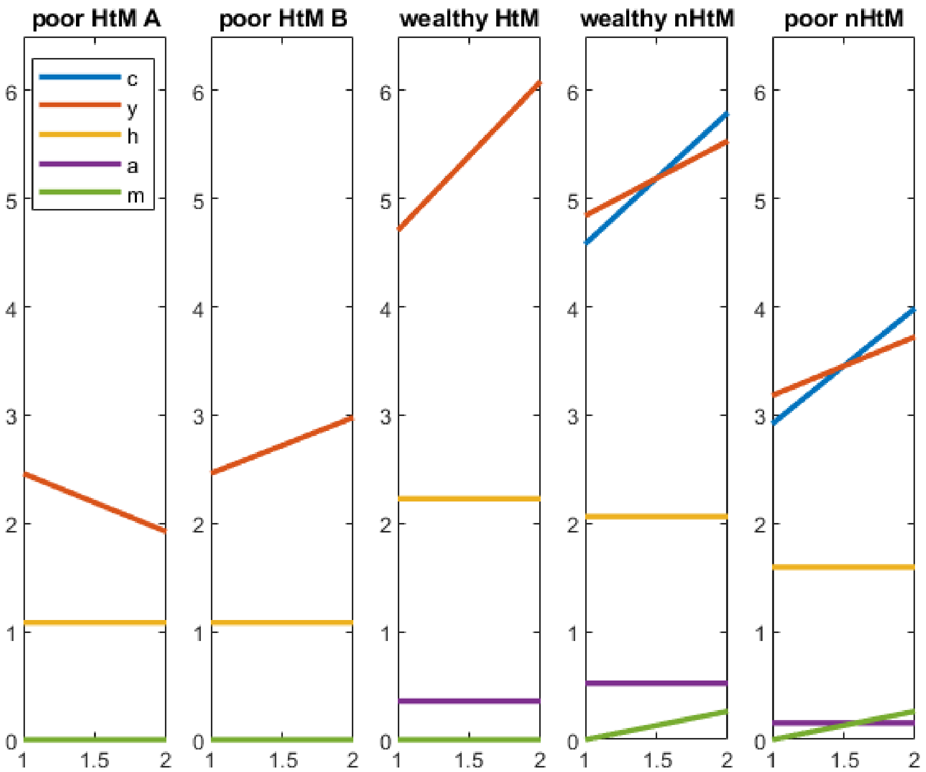


Figure 2.10: *The values of asset holdings (a, h, m_1, m_2), consumption (c) and income (y) in the intermediate period 1 and in the final period 2 for 5 distinct households. Note that the liquid asset line m depicts m_1 and m_2 depending on the period and income y is the labor income plus illiquid savings $R{\cdot}a$, without liquid savings m_2. Characteristics of the households are given in the Table 2.4.*

[11] Note that in Figure 2.10 the income is defined as the labor income plus the interest on illiquid assets.

	status	(A, ω)	g
1	*poor HtM* A	(2,2)	0.75
2	*poor HtM* B	(2,2)	1.25
3	*wealthy HtM*	(20,20)	1.25
4	*wealthy nHtM*	(23,20)	1
5	*poor nHtM*	(10,10)	0.75

Table 2.4: *Characteristics of the households in Figure 2.10*

A limitation of the model is the fact that it constrains the household's optimization problem to only 3 periods, which makes it impossible to generate a 'Ben-Porath' hump-shaped, non-monotonic path of human capital investment/stock. Moreover, the model operates in a non-stochastic environment, in which there are less incentives to save liquid assets that could smooth negative shocks. Thus, it does not allow for a full analysis of *hand-to-mouth* behavior. The *HtM* status is of transitory nature and, empirically, large flows between all 4 groups of consumers are observed (see Table 2.3 and Figure 2.2). The existing empirical research, including Section 2.2 of this paper, does not shed much light on how persistent the *pHtM*, *wHtM* and *nHtM* status is in the long run (i.e. if households change their type more than once or twice in the life cycle). Nevertheless, aforementioned figures show that the analysis of the *HtM* behavior should allow for transitions.

Without a formal model allowing for more periods, such statements are only speculative, yet it could be presumed that in the very beginning of the life-cycle households expect increasing income paths and they accumulate human capital as *poor HtM*, *wealthy HtM* and *poor nHtM* types. Over time they transform gradually into *wealthy non-hand-to-mouth* households or *wealthy hand-to-mouth* by accumulating illiquid assets and expecting their income to stagnate. Such predictions are in line with the theoretical model and are confirmed by the empirical analysis with one caveat: the *wealthy non-hand-to-mouth* consumers also experience increasing income. This feature could be probably captured by a model with stochastic shocks that would introduce precautionary saving motives.

2.6 Conclusions

The model presented in this paper is an extension of Kaplan et al. (2014) with the novelty of endogenizing income by introducing labor supply and human capital accumulation.

This extended framework allows to analyse characteristics of *nHtM*, *wHtM* and *pHtM* households together with factors affecting the deliberate choices of these household types. Although the model is too technically complex to obtain a tractable analytical solution, the effects of the exogenous parameters on the solutions can be analyzed numerically.

The model predicts that *poor hand-to-mouth* consumers are initially poorer and less able while *wealthy hand-to-mouth* consumers are those with higher initial endowment and higher abilities. Both groups experience income growth. *Non-hand-to-mouth* consumers experience stagnating or decreasing labor income path and are more heterogeneous when it comes to their initial characteristics, which gives rise to different behavior regarding accumulation of illiquid assets and human capital.

The major conclusion of the model is the statement that the *wealthy/poor hand-to-mouth* status is chosen endogenously. Households wish to optimize their utility, taking their characteristics as granted, and choose preferred values of asset and human capital stocks. Particularly, it is worth pointing out the optimal educational decisions: they may imply deviations from full consumption smoothing by inducing decisions leading to being liquidity constrained.

I show the possibility of existence of *nHtM* (both *poor* and *wealthy*), *pHtM* and *wHtM* consumers. Their emergence depends on the slope of the endogenous income path, which means that the characteristics of consumers like time preference β, innate ability A or initial endowment ω are crucial in making households' decisions about their $(n)HtM$ status or the decisions about accumulating illiquid assets a and human capital h.

Some of the results of the model are similar to KVW. Firstly, both *HtM* and *nHtM* households emerge, where one of the determinants of their appearance is the slope of the income path. If the household has higher total income in the final period than in the intermediate one, $y_1 < y_2$, then it is more likely to be *HtM* as it is more likely to decide to save no liquid assets. If $y_1 > y_2$ instead, then the household is more likely to be *nHtM* because of holding a positive stock of liquid assets.

Secondly, saving both in liquid and illiquid assets simultaneously does not arise in the optimum because of the fact that the return factor of the liquid asset is lower than of the illiquid one, $R > 1$, which is again in line with KVW. Furthermore, households may be willing to attach less value to consumption smoothing if they receive sufficiently high return on the illiquid assets, which gives rise to *HtM* households.

A novelty in comparison with KVW is the introduction of endogenous labor income, which provides a motive to invest in human capital as a part of the portfolio. Because of the assumptions on the production function, we can rule out zero human capital and also zero labor supply, which is in accordance with empirics.

This makes it possible to have larger sets of *wealthy* and *poor hand-to-mouth* consumer types as, contrary to the KVW model, they can also be households with zero liquid wealth in the first period, i.e. $m_1 = 0$. This is because our assumptions preclude zero income so that consumers can get positive consumption in the intermediate period without a necessity of having positive liquid assets $m_1 > 0$. The assumptions on the production function imply that the set of *nHtM* consumer types is also larger as it consists of those dividing their initial endowment between the human capital and illiquid assets and those investing only in the human capital, deciding to have no illiquid assets. In Section 2.2 these two groups of households were called the *wealthy nHtM* and the *poor nHtM*, respectively.

The model should be perceived with caution because of its stylized nature. A 3-period (effectively, a 2-period) optimization makes it impossible to obtain the empirically observed non-monotonic relationships of assets or human capital stocks over the life cycle. Furthermore, the model abstracts from the stochastic shocks and revisions of expectations. Therefore, the next step in the analysis of the *wealthy* and *poor hand-to-mouth* households should be the extension of the life-cycle horizon of consumers to many, for example, annual periods and an introduction of risk. Although such a model, in view of multiple periods, assets and decisions, must be expected to be numerically complicated, it could explain the dynamic nature of the *hand-to-mouth* behavior.

2.7 Appendix A: technical aspects of the theoretical model

2.7.1 Solution of the model with backward induction

The model is solved using the standard Kuhn-Tucker multiplier method with inequality constraints.

Final Period 2:

$$\max_{l_2} \quad u(\underbrace{g \cdot A \cdot f(h, l_2) + m_2 + R \cdot a}_{c_2},\ l_2,\ a)$$
$$s.t.$$
$$l_2 \geq 0$$

$$L = u\Big(g \cdot A \cdot f(h, l_2) + m_2 + R \cdot a,\ l_2\Big) + \lambda \cdot l_2$$
$$\frac{\partial L}{\partial l_2} = u_{c_2} \cdot f_{l_2} \cdot g \cdot A + u_{l_2} + \lambda = 0$$
$$\lambda \geq 0,\ l_2 \geq 0,\ \lambda \cdot l_2 = 0$$

The assumptions about the production function $f(h, l_t)$ differ depending on the value of h: if $h = 0$ then $f(0, l_t) = 0$ and therefore $f_l(0, l_t) = 0$, while for $h > 0$ the marginal productivity of labor is strictly positive.

Therefore, we have to distinguish two cases and analyse the labor supply l_2 for $h > 0$ and for $h = 0$. If $h > 0$ then the optimality condition implies that $l_2 > 0$ and $\lambda = 0$. In the optimum the condition is $u_{c_2} \cdot f_{l_2} \cdot g \cdot A = -u_{l_2}$. If $h = 0$ then the optimal choice is $l_2 = 0$ and $\lambda = -u_{l_2}$ because $f_{l_2} = 0$ implies $u_{c_2} \cdot f_{l_2} \cdot g \cdot A = 0$ and then $\lambda = -u_{l_2}$. $\lambda \neq 0$ implies $l_2 = 0$ to fulfill the condition $\lambda \cdot l_2 = 0$.

Intermediate Period 1:

$$\max_{m_2, l_1} \quad u(\underbrace{A \cdot f(h, l_1) + m_1 - m_2}_{c_1},\ l_1,\ a) + \beta \cdot u(\underbrace{g \cdot A \cdot f(h, l_2^{opt}) + m_2 + R \cdot a}_{c_2},\ l_2^{opt},\ a)$$

$$s.t.$$

$$l_1 \geq 0, \quad m_2 \geq 0$$

$$L = u\Big(A \cdot f(h, l_1) + m_1 - m_2,\ l_1,\ a\Big) + \beta \cdot u\Big(g \cdot A \cdot f(h, l_2^{opt}) + m_2 + R \cdot a,\ l_2^{opt},\ a\Big) + \eta \cdot l_1 + \rho \cdot m_2$$
$$\frac{\partial L}{\partial m_2} = -u_{c_1} + \beta \cdot u_{c_2} + \rho = 0$$
$$\frac{\partial L}{\partial l_1} = u_{c_1} \cdot f_{l_1} \cdot A + u_{l_1} + \eta = 0$$
$$\rho \geq 0,\ m_2 \geq 0,\ \rho \cdot m_2 = 0$$
$$\eta \geq 0,\ l_1 \geq 0,\ \eta \cdot l_1 = 0$$

As before, we have to analyse two cases because of the assumptions on the production function. If $h > 0$ then $l_1 > 0$, $\eta = 0$ and the optimality condition is $u_{c_1} \cdot f_{l_1} \cdot A = -u_{l_1}$. If $h = 0$ then $l_1 = 0$ and $\eta = 0$.

The zero labor supply decision $l_2 = l_1 = 0$, which is the optimal choice for $h = 0$, is ruled out finally because of the Inada conditions on the production and utility functions (for details see Section 2.7.2 of the Technical Appendix). Hence, the following analysis focuses only on the choice of positive labor supply.

The stock of liquid assets in the second period m_2 can be either positive or it can be zero, which distinguishes between *HtM* and *nHtM* households. It is the relation between income in the intermediate and in the final period together with the time preference β that determine the *HtM* behavior. As β is allowed to be both larger and smaller than 1 it is not possible to make a general statement about the shape of the income path generating the discussed behavior. Because of the fact that the income path is endogenous, the 'deep' parameters like time preference, innate ability or initial endowment are those that make a difference between *HtM* and *nHtM* status.

$$
\begin{array}{cc}
\textit{Hand-to-Mouth } (HtM) & \textit{non-Hand-to-Mouth } (nHtM) \\
m_2 = 0 \quad \Rightarrow \quad \rho > 0 & m_2 > 0 \quad \Rightarrow \quad \rho = 0 \\
u_{c_1} > \beta \cdot u_{c_2} & u_{c_1} = \beta \cdot u_{c_2} \\
c_1 \gtreqless c_2 & c_1 \gtreqless c_2 \\
\underbrace{A \cdot f(h, l_1) + m_1}_{c_1} \gtreqless \underbrace{g \cdot A \cdot f(h, l_2) + Ra}_{c_2} & \underbrace{A \cdot f(h, l_1) + m_1 - m_2}_{c_1} \gtreqless \underbrace{g \cdot A \cdot f(h, l_2) + m_2 + R \cdot a}_{c_2} \\
m_2 = 0 & 1/2 \cdot \Big[\underbrace{A \cdot f(h, l_1) + m_1}_{\textit{income in } t=1} - \underbrace{(g \cdot A \cdot f(h, l_2) + Ra)}_{\textit{income in } t=2} \Big] \gtreqless m_2 > 0
\end{array}
$$

Initial Period 0:

$$
\max_{a,h,m_1} \quad u(\underbrace{A \cdot f(h, l_1^{opt}) + m_1 - m_2^{opt}}_{c_1},\ l_1^{opt},\ a) + \beta \cdot u(\underbrace{g \cdot A \cdot f(h, l_2^{opt}) + m_2^{opt} + R \cdot a}_{c_2},\ l_2^{opt},\ a)
$$

$$
s.t.
$$

$$
\omega = a + h + m_1
$$

$$
a \geq 0, \quad h \geq 0, \quad m_1 \geq 0
$$

$$
\begin{aligned}
& L = u\Big(A \cdot f(h, l_1^{opt}) + m_1 - m_2^{opt},\ l_1^{opt},\ a\Big) + \beta \cdot u\Big(g \cdot A \cdot f(h, l_2^{opt}) + m_2^{opt} + R \cdot a,\ l_2^{opt},\ a\Big) + \dots \\
& \quad + \delta \cdot a + \theta \cdot h + \kappa \cdot m_1 + \tau \cdot (\omega - a - h - m_1) \\
& \frac{\partial L}{\partial m_1} = u_{c_1} + \kappa = \tau \\
& \frac{\partial L}{\partial a} = \beta \cdot R \cdot u_{c_2} + u_a + \beta \cdot u_a + \delta = \tau \\
& \frac{\partial L}{\partial h} = u_{c_1} \cdot A \cdot f_h(h, l_1) + \beta \cdot u_{c_2} \cdot g \cdot A \cdot f_h(h, l_2) + \theta = \tau \\
& \delta \geq 0,\ a \geq 0,\ \delta \cdot a = 0 \\
& \theta \geq 0,\ h \geq 0,\ \theta \cdot h = 0 \\
& \kappa \geq 0,\ m_1 \geq 0,\ \kappa \cdot m_1 = 0 \\
& \tau > 0,\ \omega = a + h + m_1
\end{aligned}
$$

The analysis of decision making in the initial period 0 has to distinguish two cases: the one with $m_2 > 0$ and the one in which $m_2 = 0$. Further, in both of these cases 8 potential decisions about the choice of a, h, m_1 have to be investigated as each of these variables can be either positive or zero.

a) *Hand-to-Mouth* households:

$m_2 = 0$ with $\rho > 0 \Rightarrow$ HtM household with $u_{c_1} > \beta \cdot u_{c_2}$

The assumption about $f(h, l_i)$ with $l_i > 0$, $lim_{h \to 0} f_h(h, l_i) = \infty$ rules out the following situations:

1. $a = 0,\ h = 0,\ m_1 = 0$	2. $a > 0,\ h = 0,\ m_1 = 0$
3. $a = 0,\ h = 0,\ m_1 > 0$	4. $a > 0,\ h = 0,\ m_1 > 0$

Potential solutions that do not violate $lim_{h \to 0} f_h(h, l_i) = \infty$ are the ones with $h > 0$:

$$\text{wealthy HtM} \begin{cases} 1. & a > 0,\ h > 0,\ m_1 > 0 \\ 2. & a > 0,\ h > 0,\ m_1 = 0 \end{cases} \qquad \text{poor HtM} \begin{cases} 3. & a = 0,\ h > 0,\ m_1 > 0 \\ 4. & a = 0,\ h > 0,\ m_1 = 0 \end{cases}$$

In line with the definition of *HtM* households by 'KVW' (the ones with $m_2 = 0$) are either wealthy or poor depending on their illiquid asset position a. Contrary to 'KVW' the model does not preclude the existence of households with $m_1 = 0$ because of assuming a more general situation in which income in the first period is endogenous and positive without the necessity of having $m_1 > 0$. Therefore, *poor* and *wealthy HtM* can have positive or zero liquid asset stocks in the first period.

b) *non-Hand-to-Mouth* households:

$m_2 > 0$ with $\rho = 0 \Rightarrow$ *nHtM* household with $u_{c_1} = \beta \cdot u_{c_2}$

Assumptions $lim_{h \to 0} f_h(h, l_i) = \infty$ and $R > 1$ imply that the following decisions are not optimal:

1. $a = 0,\ h = 0,\ m_1 = 0$	2. $a = 0,\ h = 0,\ m_1 > 0$
3. $a > 0,\ h = 0,\ m_1 = 0$	4. $a > 0,\ h = 0,\ m_1 > 0$
5. $a > 0,\ h > 0,\ m_1 > 0$	6. $a = 0,\ h > 0,\ m_1 > 0$

Cases 5. and 6. are not optimal because saving in the liquid asset is return dominated if households save for the future in the illiquid asset.

The model leaves 2 potential solutions in the case of *nHtM* households:

1. $a > 0,\ h > 0,\ m_1 = 0$
2. $a = 0,\ h > 0,\ m_1 = 0$

In both situations $m_1 = 0$ as *nHtM* consumers are those with such an income path that does not give enough resources in the final period to match the preferred consumption profile (such a path implies that the preferred consumption profile implies a necessity of saving) so there is no point in taking additional liquid resources from the period 0 to the period 1 as a household does not need additional resources in the intermediate period but needs them in the final one and returns on the illiquid asset dominate the ones on liquid assets. In the first case we can observe the choice of portfolio that guarantees equating the marginal utility gains between investing one more unit of endowment in illiquid assets and human capital: $A \cdot u_{c_1} \cdot [f_h(h, l_1) + g \cdot f_h(h, l_2)] = R \cdot u_{c_1} + u_a \cdot (1 + \beta)$. The second case is a situation in which the marginal increase in utility would be higher for investing an additional unit of endowment in the human capital than in the illiquid asset even when investing the whole initial endowment ω in education.

2.7.2 Ruling out the decision of zero human capital and zero labor supply

As already mentioned, the optimal labor supply decisions in the final period 2 and in the intermediate period 1 differ depending on the value of human capital h that is chosen in the initial period 0. The decisions made in the initial period with positive labor supply in the intermediate and final period have been extensively analyzed in the main body of the text (special attention should be attached to the fact that positive labor supply rules out zero human capital). In this subsection it is described how the decision of zero labor supply together with zero human capital is ruled out. In any case, if such an optimum existed, it would not be interesting because it is empirically implausible that agents cannot generate labor income by exerting at least some labor effort.

The following paragraphs present the solution to the consumer's problem by backward induction method focusing on the optimal choice of labor supposing $h = 0$.

Final period 2

If $h = 0$ then the optimality condition $\frac{\partial L}{\partial l_2} = u_{c_2} \cdot f_{l_2} \cdot g \cdot A + u_{l_2} + \lambda = 0$ with $\lambda \geq 0,\ l_2 \geq 0,\ \lambda l_2 = 0$ becomes $u_{l_2} + \lambda = 0$ because $f_{l_2}(0, l_2) = 0$. The only possibility to satisfy the optimality and complementary slackness conditions is to choose $l_2 = \lambda = 0$ as $u_{l_2}(c_2, 0) = 0$.

Intermediate period 1

The problem in the intermediate period is identical as the problem in the final period regarding the labor supply. If $h = 0$ it is optimal to choose $l_1 = \eta = 0$. The problem regarding the choice of m_2 is identical as in the main body of the text i.e. as in the case of positive labor supply.

Initial period 0

The optimality condition $\frac{\partial L}{\partial h} = u_{c_1} \cdot A \cdot f_h(h, l_1) + \beta \cdot u_{c_2} \cdot g \cdot A \cdot f_h(h, l_2) + \theta = \tau$ implies $\theta = \tau$ because $f_h = 0$ if $l_2 = l_1 = 0$. Since τ is a strictly positive multiplier of the equality $\omega = a + m_1 + h$, $\theta = \tau > 0$. Therefore, h must be zero to fulfill the complementary slackness condition $\theta h = 0$.

Further, a standard analysis is done: we have to analyse four combinations of a and m_1 for *HtM* households ($m_2 = 0$) and four cases of *nHtM* households ($m_2 > 0$) as $h = 0$ rules out cases with $h > 0$. An inspection of the cases shows that the only possible one is the one with $a > 0, m_1 > 0, m_2 = 0$. It is so because we assume no exogenous income and we rule out situations that $c_1 = 0$ or $c_2 = 0$ as $\lim_{c_i \to 0} u_{c_i}(c_i, 0) = \infty$, or we rule out $a = 0, m_1 = 0$ as $\omega = a + m_1 > 0$. Moreover, we can rule out *nHtM* behavior with positive m_2 because it is not optimal to save in the liquid asset m that gives a lower rate of return than the illiquid asset a.

Thus, the solution that can be optimal has to fulfill: $a > 0, m_1 > 0, m_2 = 0, h = 0$. Nevertheless, it can be ruled out using a simple variational argument that it is strictly dominated by a choice of positive labor supply and positive human capital. Let us take into account that in this situation $c_1 = m_1$ and $c_2 = R{\cdot}a$ so we have $u(m_1, 0) + \beta{\cdot}u(R{\cdot}a, 0)$. Imagine that a small fraction $\epsilon > 0$ of either a or m_1 is used to invest in human capital and a small amount of labor l_1 and l_2 is supplied in line with the optimality conditions. This means that the agent loses a small amount of utility created by ϵ through a or m_1. However, the agent gets more utility from human capital, because of the Inada conditions, as its marginal productivity approaches infinity and therefore outweighs the rate of return on a or m_1. The marginal disutility from labor supply is negligible because of the condition $\lim_{l_i \to 0} u_{l_i} = 0$. Hence, choices with zero labor supply and zero human capital can be ruled out.

2.7.3 Sufficient conditions and the concavity of optimization problems

The analysis above lists first-order conditions. This subsection proves that they are sufficient conditions for characterizing maxima as the optimization problems in all three periods are (strictly) concave in the space of decision variables. Because of the fact that the solution $h = l_1 = l_2 = 0$ is ruled out (see the Technical Appendix 2.7.2) the sufficiency analysis drops this possibility. Practically, it means that the Hessian matrices are negative definite instead of being negative semidefinite and the maximization problems are strictly concave.

Final period 2

The optimization problem in the final period is given by the following Lagrangian function and first-order condition:

$$L = u\Big(g \cdot A \cdot f(h, l_2) + m_2 + R \cdot a,\ l_2,\ a\Big) + \lambda \cdot l_2$$
$$\frac{\partial L}{\partial l_2} = u_{c_2} \cdot f_{l_2} \cdot g \cdot A + u_{l_2} + \lambda = 0$$
$$\lambda \geq 0,\ l_2 \geq 0,\ \lambda \cdot l_2 = 0$$

The second-order derivative of the Lagrangian is given as follows:

$$\frac{\partial^2 L}{\partial^2 l_2} = (u_{c_2 c_2} \cdot g \cdot A \cdot f_{l_2} + u_{c_2 l_2}) \cdot g \cdot A \cdot f_{l_2} + u_{c_2} \cdot g \cdot A \cdot f_{l_2 l_2} + u_{l_2 l_2}$$

Hessian is therefore negative for all $l_2 > 0$, because of the assumptions on utility and production functions: $\forall_{l>0} \quad u_{cc} < 0, u_{cl} = 0, f_l > 0, f_{ll} < 0, u_{ll} < 0$. The Lagrangian is then strictly concave $\forall_{l>0}$ so that first-order conditions are characterizing a maximum.

Intermediate period 1

The optimization problem in this period is stated by the following Lagrangian and first-order conditions (cross-derivatives are zero because of the separability of the utility function):

$$L = u\Big(A{\cdot}f(h, l_1) + m_1 - m_2,\ l_1,\ a\Big) + \beta \cdot u\Big(g{\cdot}A{\cdot}f(h, l_2^{opt}) + m_2 + R \cdot a,\ l_2^{opt},\ a\Big) + \eta \cdot l_1 + \rho \cdot m_2$$

$$\frac{\partial L}{\partial m_2} = -u_{c_1} + \beta u_{c_2} + \rho = 0$$

$$\frac{\partial L}{\partial l_1} = u_{c_1} \cdot A \cdot f_{l_1} + u_{l_1} + \eta = 0$$

$$\rho \geq 0,\ m_2 \geq 0,\ \rho \cdot m_2 = 0$$

$$\eta \geq 0,\ l_1 \geq 0,\ \eta \cdot l_1 = 0$$

The second order derivatives are:

$$\frac{\partial^2 L}{\partial^2 m_2} = u_{c_1 c_1} + \beta \cdot u_{c_2 c_2}$$

$$\frac{\partial^2 L}{\partial^2 l_1} = (u_{c_1 c_1} \cdot A \cdot f_{l_1} + u_{c_1 l_1}) \cdot A \cdot f_{l_1} + u_{c_1} \cdot A \cdot f_{l_1 l_1} + u_{l_1 c_1} \cdot A \cdot f_{l_1} + u_{l_1 l_1}$$

$$\frac{\partial^2 L}{\partial m_2 \partial l_1} = -u_{c_1 c_1} \cdot A \cdot f_{l_1} - u_{c_1 l_1}$$

The Hessian matrix composed of these second-order derivatives is negative definite for the positive labor supply as both $\frac{\partial^2 L}{\partial^2 m_2}$ and $\frac{\partial^2 L}{\partial^2 l_1}$ are negative and $\frac{\partial^2 L}{\partial^2 m_2} \cdot \frac{\partial^2 L}{\partial^2 l_1} - (\frac{\partial^2 L}{\partial m_2 \partial l_1})^2 = u_{c_1 c_1} \cdot u_{c_1} \cdot A \cdot f_{l_1 l_1} + u_{c_1 c_1} u_{l_1 l_1} + \beta u_{c_2 c_2} \cdot u_{c_1 c_1} \cdot A^2 \cdot f_{l_1}^2 + \beta u_{c_2 c_2} \cdot u_{c_1} \cdot A \cdot f_{l_1 l_1} + \beta u_{c_2 c_2} \cdot u_{l_1 l_1} > 0$. Therefore, the Hessian is strictly concave $\forall l_2 > 0, m_2 \geq 0$ and first-order conditions are sufficient for a maximum.

Initial period 0

The optimization problem in this period is stated by the following Lagrangian and first-order conditions (cross-derivatives are zero because of the separability of the utility function):

$$L = u\Big(A{\cdot}f(h, l_1) + m_1 - m_2,\ l_1,\ a\Big) + \beta \cdot u\Big(g{\cdot}A{\cdot}f(h, l_2^{opt}) + m_2 + R \cdot a,\ l_2^{opt},\ a\Big) + \dots$$
$$+ \delta \cdot a + \theta \cdot h + \kappa \cdot m_1 + \tau \cdot (\omega - a - h - m_1)$$

$$\frac{\partial L}{\partial m_1} = u_{c_1} + \kappa = \tau$$

$$\frac{\partial L}{\partial a} = \beta \cdot R \cdot u_{c_2} + u_a + \beta \cdot u_a + \delta = \tau$$

$$\frac{\partial L}{\partial h} = u_{c_1} \cdot A \cdot f_h(h, l_1) + \beta \cdot u_{c_2} \cdot g \cdot A \cdot f_h(h, l_2) + \theta = \tau$$

$$\delta \geq 0,\ a \geq 0,\ \delta \cdot a = 0$$

$$\theta \geq 0,\ h \geq 0,\ \theta \cdot h = 0$$

$$\kappa \geq 0,\ m_1 \geq 0,\ \kappa \cdot m_1 = 0$$

$$\tau > 0,\ \omega = a + h + m_1$$

The second-order derivatives are:

$$\frac{\partial^2 L}{\partial^2 m_1} = u_{c_1 c_1}$$

$$\frac{\partial^2 L}{\partial^2 a} = \beta \cdot R^2 \cdot u_{c_2 c_2} + u_{aa} + \beta u_{aa}$$

$$\frac{\partial^2 L}{\partial^2 h} = u_{c_1 c_1} \cdot A^2 \cdot f_h^2(h, l_1) + u_{c_1} \cdot A \cdot f_{hh}(h, l_1) + \beta u_{c_2 c_2} \cdot g^2 \cdot A^2 \cdot f_h^2(h, l_2) + \beta g \ \cdot A \cdot u_{c_2} \cdot f_{hh}(h, l_2)$$

$$\frac{\partial^2 L}{\partial m_1 \partial a} = u_{c_1 a} = 0$$

$$\frac{\partial^2 L}{\partial m_1 \partial h} = u_{c_1 c_1} \cdot A \cdot f_h(h, l_1)$$

$$\frac{\partial^2 L}{\partial a \partial h} = \beta \cdot R \cdot g \cdot A \cdot u_{c_2 c_2} \cdot f_h(h, l_2)$$

The sufficient conditions for a maximum are satisfied as the Hessian composed of these partial derivatives is negative definite. It is so as the first leading principal minor is negative: $u_{c_1 c_1} < 0$, the second leading principal minor is positive: $u_{c_1 c_1} \cdot \big(\beta \cdot R^2 \cdot u_{c_2 c_2} + u_{aa} + \beta \cdot u_{aa}\big) - (u_{c_1 a})^2 > 0$ and the determinant of the whole Hessian matrix (the third leading principal minor) is negative for $h > 0, l_1 > 0, l_2 > 0$, which means negative definiteness.

2.7.4 System of equations given by the model

The model consists of the following 8 unknowns $\{a, h, m_1, m_2, l_1, l_2, c_1, c_2\}$, 7 Kuhn-Tucker multipliers because of the imposed non-negativity conditions and one equality condition $\{\theta, \delta, \kappa, \rho, \eta, \lambda, \tau\}$ and 12 parameters $\{\sigma, \mu, \alpha, \gamma, \beta, \omega, \psi, \varphi, \chi, g, A, R\}$. In order to solve the model for 15 unknowns we need 15 equations which are listed below:

1. $l_2: \quad u_{c_2} \cdot f_{l_2} \cdot g \cdot A + u_{l_2} + \lambda = 0$
2. $l_1: \quad u_{c_1} \cdot f_{l_1} \cdot A + u_{l_1} + \eta = 0$
3. $m_1: \quad u_{c_1} + \kappa - \tau = 0$
4. $a: \quad \beta \cdot R \cdot u'_{c_2} + u_a + \beta \cdot u_a + \delta - \tau = 0$
5. $h: \quad u_{c_1} \cdot A \cdot f_h(h, l_1) + \beta \cdot u_{c_2} \cdot g \cdot A \cdot f_h(h, l_2) + \theta - \tau = 0$
6. $m_2: \quad -u_{c_1} + \beta u_{c_2} + \rho = 0$
7. $c_1: \quad c_1 = A \cdot f(h, l_1) + m_1 - m_2$
8. $c_2: \quad c_2 = A \cdot f(h, l_2) \cdot g + m_2 + R \cdot a$
9. $\omega = a + h + m_1$
10. $\delta \cdot a = 0$
11. $\theta \cdot h = 0$
12. $\kappa \cdot m_1 = 0$
13. $\rho \cdot m_2 = 0$
14. $\eta \cdot l_1 = 0$
15. $\lambda \cdot l_2 = 0$

If we adopt the production and utility functions, $u(c_i, l_i, a) = \frac{c_i^{1-\sigma}-1}{1-\sigma} + \frac{\psi}{\chi}(a+\varphi)^{\chi} - \frac{\mu}{\alpha} l_i^{\alpha}$ and $f(h, l_i) = h^{\gamma} l_i^{1-\gamma}$ conditions 1-8 change to the following ones:

1. $l_1: \quad c_1^{-\sigma} \cdot A \cdot (1-\gamma) \cdot h^{\gamma} \cdot l_1^{-\gamma} - \mu \cdot l_1^{\alpha-1} + \eta = 0$
2. $l_2: \quad c_2^{-\sigma} \cdot A \cdot (1-\gamma) \cdot h^{\gamma} \cdot l_2^{-\gamma} \cdot g - \mu \cdot l_2^{\alpha-1} + \lambda = 0$
3. $m_1: \quad c_1^{-\sigma} + \kappa - \tau = 0$
4. $a: \quad \beta \cdot R \cdot c_2^{-\sigma} + \psi \cdot (a+\varphi)^{\chi-1} + \beta \cdot \psi \cdot (a+\varphi)^{\chi-1} + \delta - \tau = 0$
5. $h: \quad c_1^{-\sigma} \cdot A \cdot \gamma h^{\gamma-1} \cdot l_1^{1-\gamma} + \beta \cdot c_2^{-\sigma} \cdot g \cdot A \cdot \gamma h^{\gamma-1} \cdot l_2^{1-\gamma} + \theta - \tau = 0$
6. $m_2: \quad -c_1^{-\sigma} + \beta \cdot c_2^{-\sigma} + \rho = 0$
7. $c_1: \quad c_1 = A \cdot h^{\gamma} \cdot l^{1-\gamma} + m_1 - m_2$
8. $c_2: \quad c_2 = g \cdot A \cdot h^{\gamma} \cdot l^{1-\gamma} + m_2 + R \cdot a$

All in all, we have 15 equations that should be solved to find expressions for 15 unknowns (8 variables + 7 multipliers, three out of which are zero as we know from our analysis that l_1,l_2 and h are positive so their multipliers are zero). Because the Cobb-Douglas production function is nonlinear, the equations listed above are nonlinear even for simple utility functions like $u(c_i, l_i, a) = ln\ c_i + ln\ a - \mu l_i^2$. The model does not produce any tractable analytical solutions, even in situations in which the labor supply is assumed to be exogenously given, so numerical methods are indispensable to solve this consumer's problem.

2.8 Appendix B: SCF variables used in the analysis

This section provides names of SCF variables used for the construction of income, liquid wealth and illiquid wealth in the statistical analysis in section 2.2.

Income is defined as the SCF variable '*income*' minus the variables '*kginc*' and '*intdivinc*'.

Liquid wealth is equal to liquid assets minus liquid debt. Liquid assets are the sum of the following SCF variables: '*mma*', '*checking*', '*saving*', '*call*', '*bond*', '*stocks*' and '*nmmf*', with the sum of first four being inflated by the '*cash ratio*', which is computed as explained in the main text. Liquid debt contains credit card and installment debt (variable '*ccbal*').

Illiquid wealth is the sum of illiquid assets net of mortgage debt. In terms of SCF variables, we sum '*houses*', '*oresre*', '*nnresre*', '*retqliq*', '*cashli*', '*cds*' and '*savbnd*', net of '*mrthel*', which is the value of mortgage loans. Illiquid wealth does not contain net wealth in vehicles (variable '*vehic*' minus '*VEH_INST*') nor student loans ('*EDN_INST*').

3 A Comparison of German, Swiss and Polish Fiscal Rules Using Monte Carlo Simulations

This chapter has been prepared together with Michał Ramsza from the Warsaw School of Economics.

Abstract:
This chapter assesses the economic implications of existing fiscal rules in Poland, Switzerland and Germany. In the analysis we establish economic relationships between output, government revenues and expenditures estimating a VAR model on US data for 1960–2015. Imposing fiscal policies implied by a given rule on those relationships, we analyze the consequences for the simulated paths of debts, deficits and expenditures in terms of stability and cyclicality. We find that the Swiss and German rules are strict and stabilize debt at low levels, while the Polish rule stabilizes the debt level at about 50 % of GDP in the long run. All rules imply an anticyclical fiscal policy but their anticyclicality is limited: the difference between the maximum and minimum deficit to GDP ratio implied by changes in the output gap varies, at most, by 1.9 pp and 2.8 pp over the whole business cycle for the Polish and Swiss/German rules, respectively.

JEL Codes: C32, E62, H62, H63
Keywords: fiscal policy, fiscal rules

3.1 Introduction

The Great Recession, together with debt crises in some European countries, has put fiscal policy in the spotlight again and has accentuated the significance of sound and sustainable public finance. It is clear that both neoclassical tax smoothing and active Keynesian policy need enough fiscal space to be effective. However, the fiscal stance of most developed countries does not offer much room for action, which is particularly worrying in view of demographic changes or, a potentially long-lasting, slowdown of economic growth due to secular stagnation.[1]

Wyplosz (2013) argues that a departure from optimal fiscal policy, due to a deficit bias, results in too high, suboptimal debt levels and is a political failure. Potential causes of such a failure are concisely summarized in Alesina and Passalacqua (2016). They include fiscal illusion that results in an inability among voters to understand the notion of the intertemporal budget constraint for the government, political budget cycles, delayed stabilizations with 'wars of attrition' preventing smooth fiscal contractions or common pool problems that result in a failure of certain groups of voters to internalize fully fiscal policy costs. The suboptimality of fiscal policy may be also connected with the procyclicality of government expenditures, see e.g. Alesina, Campante, and Tabellini (2008).

The aforementioned authors summarize potential remedies for the political deficit bias. Firstly, Alesina and Perotti (1996) argue that an improvement of institutions may restrict the process of budget creation (e.g. an improvement of the voting process on budget amendments in the parliament). Secondly, Wyplosz (2008) indicates that fiscal councils, i.e. impartial committees may decide on the budget balance of the government or, at least, assess and comment the fiscal policy led by the government. Thirdly, fiscal rules, as numerical and formal mechanisms, may restrict budget balances and they are the focus of this paper.

Our research serves two purposes. Firstly, we provide a detailed mathematical description of the fiscal rules existing in Switzerland, Poland and Germany. Secondly, we analyze the behavior of these three rules and compare them with a simple balanced-budget rule. We apply these rules to artificially created series of GDP and government revenues of a benchmark economy. The economy is represented by a VAR model, which is estimated on empirical US data. The actual economic relationships between GDP, government revenues and expenditures, described by the VAR model, interact with the fiscal policy implied by

[1] For a discussion of the secular stagnation see e.g. Gordon (2015) or Summers (2015).

each of the analyzed rules. This in turn results in interdependent time series of output, revenues and expenditure that are analyzed.

Analyzing four rules in a unified Monte Carlo simulational framework enables credible comparison of their implications and effectiveness regarding stabilizing output, reducing procyclicality of fiscal policy and reducing debt levels. There is a wide consensus in the existing literature that these features, together with transparency and simplicity of a rule's mechanism, are the most important merits of an effective fiscal rule. Therefore, the comparison focuses on precisely these features.

We find that the rules are capable of stabilizing debt at low levels and they imply anti-cyclical fiscal policy. Nonetheless, the degree of antycyclicality is rather moderate: the maximum difference in deficit to GDP ratios, caused by output gap volatility, over a business cycle does not exceed 1.9 pp for the Polish rule and maximally 2.8 pp for the Swiss and German rules.

The Monte Carlo simulation methodology, which we use for our analysis, has been already used in the analysis of fiscal rules. Examples of such research are Geier (2012), who assesses the Swiss fiscal rule on purely artificial data, and Korniluk (2016), who analyses the Polish expenditure rule based on time series created by an econometric model using data from EU countries. The research in Landon and Smith (2017) is closest to our contribution. They use a very similar approach based on time series generated by a VAR model and compare properties of different fiscal rules. While they focus on a synthetic measure of welfare and do not assess in detail other features of the rules, we focus on procyclicality and stabilization of debt, deficits and expenditures. Moreover, they analyze simplified rules that are not part of actual legislation. The major conclusions of their paper are that rules, in general, help in increasing welfare and in decreasing expenditure volatility. Finally, structural deficit rules (like the Swiss or German ones) deliver best results in terms of welfare maximization.

Problems arising with implementation and effectiveness of rules are well-known on theoretical grounds: the 'rules vs. discretion dilemma' may result in time-inconsistency (these issues are analyzed specifically for fiscal rules in Alfaro and Kanczuk (2016) or Halac and Yared (2014)) with further problems being commitment and self-enforcement issues. Short time series and endogeneity make it difficult to assess the behavior of fiscal rules empirically. Nevertheless, the popularity of this strand of research is increasing together with the rising popularity of rules within last 25 years. Schaechter, Kinda, Budina, and Weber (2012) present an encyclopedic overview of countries adopting fiscal rules together with reviewing their types and features. Econometric research (Debrun and Kumar (2007), Holm-Hadulla, Hauptmeier, and Rother (2012) or Nerlich and Reuter (2016)) claims that fiscal rules are associated with lower deficits, more fiscal space and lower procyclicality of fiscal policy. Nevertheless, they point out that self-selection may convolute the causal effect of fiscal rules as countries with a better fiscal situation or more willingness to follow a conservative fiscal policy may be more eager to implement rules, which then serve rather as signaling devices than as binding policy constraints. Recent research, which includes Grembi, Nannicini, and Troiano (2016) and Guerguil, Mandon, and Tapsoba (2017), using a quasi-experimental setting or propensity-score matching, shows that rules themselves can affect, in the sense of causality, fiscal policy by reducing deficits and procyclicality.

The rest of the paper is organized as follows. Section 3.2 describes in detail the mechanics of all four analyzed rules. Section 3.3 presents the framework in which the rules are assessed. Section 3.4 describes the VAR model generating artificial GDP and revenue time series used in the simulations. Section 3.5 presents the obtained results and Section 3.6 concludes.

The Appendix contains all technical details. Section 3.7 explains all the steps of algorithms used in the simulations. Section 3.8 presents the modified HP filter used in the calculations of the Swiss and German rule. Section 3.9 summarizes all diagnostic checks of the VAR model used in the paper.

3.2 Description of the fiscal rules

This section explains the mechanics behind each of the rules analyzed in the paper, together with a review of the respective literature. The focus of the analysis is economic so that we do not address legal details like the degree of enshrinement of each rule in a legal system of each country, potential loopholes, the degree of budget rule coverage, enforcement sanctions, escape clauses etc.

Fiscal rules in each of the analyzed countries cover a different amount of public expenditure (understood as general government spending) as, firstly, they differ in their central budget coverage by definition and, secondly, central government spending that is subject to these rules differs significantly because of the federal or centralized nature of Germany, Switzerland and Poland. For example, the Swiss rule does not include expenditure on unemployment insurance and the Polish rule does not encompass 'expenditure generated by institutions incapable of creating large deficits' (e.g. the Polish Academy of Science). According to the Eurostat database, central government expenditures in 2015 encompassed

10.5% of GDP in Switzerland, 22.9% in Poland and 12.6% in Germany, while general government expenditure in these countries was, in terms of the GDP fraction, 33.9%, 41.5% and 44.0%, respectively.

GDP, total government expenditures, i.e. including interest payments on existing debt stock, total government revenues and public debt in a budget period t are called Y_t, G_t, R_t and D_t, respectively. We define public deficit as $R_t - G_t$, which means that a positive deficit is, in fact, a surplus. Debt is defined in the same fashion: positive values of D_t mean an accumulation of assets and negative values of D_t mean liabilities. The variable c_t stands for corrections of some variables to be undertaken in period t and is connected with a state of the correction account CA_t, which is used, in some form, in all rules. Corrections show up in all of the analyzed rules, but the simplified balanced budget rule, and pertain to expenditure or deficit limits. Correction accounts CA_t are defined differently for every rule and are explained in detail below. $E_t[x_{t+1}]$ is an expectation of a variable x_{t+1} in period t. The convention applied in the paper, which is in line with reality and necessary from the technical perspective of the simulation, is that a budget for year $t+1$ is planned in year t and its plan is based on projected values $E_t[x_{t+1}]$.

3.2.1 Swiss rule

The Swiss fiscal rule, which is called 'the debt brake' or 'die Schuldenbremse' in German, is described in Geier (2011). It was created in 2000 and is operational since 2003, after its three-year vacatio legis.

The main tenet of the rule is to have the budget structurally balanced over the business cycle. The rule is summarized by the following equation:

$$\bar{G}_{t+1} = E_t[k_{t+1}] \cdot E_t[R_{t+1}], \quad \text{with } E_t[k_{t+1}] = \frac{E_t[Y^*_{t+1}]}{E_t[Y_{t+1}]},$$

where $\bar{G}$ is the expenditure limit for the next period's budget and it is equal to expected revenues R multiplied by the expected business cycle adjustment factor k. The adjustment factor k is a ratio between the long-run trend output Y^* and actual output Y. The logic behind this adjustment is that when the economy is below its trend, i.e. it is in a slowdown phase: $E_t[Y^*_{t+1}] > E_t[Y_{t+1}]$, the adjustment factor k is larger than one allowing expenditures to be larger than revenues, which results in a (cyclical) deficit. The opposite happens when the economy is in a boom phase: k is lower than one as the economy is above its trend and the rule requires a (cyclical) surplus. Both trend output and actual output are real variables.

The expenditure limit $\bar{G}$ is based on expectations, which do not necessarily coincide with their realizations. The discrepancies may come from forecasts errors, as it may be so that $E_{t-1}[R_t] \neq R_t$ or $E_{t-1}[k_t] \neq k_t$, or because initially authorized expenditure may differ from actual spending. In the simulation we abstract from differences between authorized and actual expenditure so the only source of discrepancies are forecast errors. The difference between the expenditure limit and the revised realization of expenditure ceiling $\bar{G}_t - \bar{G}_t^R$ is credited in the compensation account CA in line with the equation:

$$CA_t = CA_{t-1} + (\bar{G}_t - \bar{G}_t^R),$$

where $\bar{G}_t^R$ is a revised expenditure ceiling, i.e. the expenditure limit calculated with realizations of variables Y and R instead of their expectations ($\bar{G}_t^R = k_t \cdot R_t$ with $k_t = Y_t^*/Y_t$).

If the cumulated deficit in the correction account is higher than 6% of the expenditure (i.e. $CA_t/G_t > 0.06$) then the excessive amount must be eliminated by decreasing expenditure limits within next 3 years. The corrective amount is defined as $c = max(0, \frac{CA_t}{G_t} - 0.06)$. The statement that a deficit in the correction account has to be eliminated within 3 years is not precise enough for the simulation algorithm so it is assumed that it is always eliminated in the next year after its occurrence by multiplying $\bar{G}_{t+1}$ with the term $c_{t+1} = (1 - c)$.

An innate feature of the Swiss fiscal rule is its method of trend calculation. The trend Y^* is calculated using a modified HP filter. The modification is presented originally in Bruchez (2003) and it is explained in detail in Appendix 3.8. The modified HP filter applies different weights for observations at the very end of the rolling window of observations used to calculate the trend. The reason for such a modification is the fact that the standard HP filter does not smooth enough observations at the end of sample. Pigoń and Ramsza (2016) confirm that the application of this change may also increase the countercyclical properties of the rule. The rolling window consists of 24 GDP observations, which is, to the best of our knowledge, just a discretionary decision of Swiss authorities. The last observation in the rolling sample is a GDP prediction $E_t[Y_{t+1}]$.

The rule is praised (see eg. Beljean and Geier (2013)) for its transparency and simplicity. It is pointed out that it leads to surpluses and decreases in government expenditure, even in nominal terms. It is unknown though if this outcome is caused by the construction of the rule or by favorable conditions in which the Swiss economy is currently operating.

3.2.2 German rule

The German fiscal rule, often claimed to be inspired by the Swiss rule, bears the name of its Helvetic counterpart and is also called 'the (German) debt brake' or 'die (Deutsche) Schuldenbremse' in German. The rule is best documented by the official paper of Federal Ministry of Finance, Germany (2015), which serves as a basic reference for the rule's mechanics. The German debt brake entered the German constitution in 2011 with the federal budget in 2016 being the first under the official scope of the rule. Budgets between 2011 and 2016 were subject to transitory constraints.

The rule states that the structural federal budget should be 'nearly' balanced as the maximum allowed structural deficit is set to be 0.35% of GDP. The rule is best specified by the following equation (all variables are nominal ones):

$$\begin{aligned} E_t[R_{t+1}] - \bar{G}_{t+1} = & -0.0035 \cdot E_t[Y_{t+1}] + E_t[F_{t+1}] \\ & + E_t[\epsilon_{t+1}] \cdot (E_t[Y_{t+1}] - E_t[Y^*_{t+1}]) - c_{t+1} \ , \end{aligned}$$

where $E_t[R_{t+1}] - \bar{G}_{t+1}$ is the maximum permissible projected deficit ($\bar{G}_{t+1}$ is the maximum allowed expenditure), $0.0035 \cdot E_t[Y_{t+1}]$ is the maximum allowed structural deficit equal to 0.35% of GDP, $E_t[F_{t+1}]$ is the balance of financial transactions (i.e. those transactions related to financial assets, e.g. privatization proceeds), $E_t[\epsilon_{t+1}] \cdot (E_t[Y_{t+1}] - E_t[Y^*_{t+1}])$ stands for the cyclical component of the budget balance and c_{t+1} is the correction coming from the stance of the correction account CA_t. The cyclical term is the multiplication of the output gap $E_t[Y_{t+1}] - E_t[Y^*_{t+1}]$, where $E_t[Y^*_{t+1}]$ is the expected potential output, with the semi-elasticity ϵ of federal budget balance with respect to the output gap. The semi-elasticity ϵ measures the impact of a change in economic activity on federal revenues and expenditures, which affect together the budget balance. The correction term c_{t+1} is connected with the compensation account CA_t, whose goal is to make sure the rule works not only with the projected but also with the actual (realized) budget. The state of the correction account is determined as follows:

$$CA_t = CA_{t-1} + (R_t - G_t) - (R_t - \bar{G}^R_t) \ ,$$

where $R_t - G_t$ is the actual deficit[2] and $R_t - \bar{G}^R_t$ is the revised borrowing limit. G_t is the actual expenditure, which is equal in our simulation to $\bar{G}_t$ as there are no unplanned expenditures and all planned ones are undertaken, and $\bar{G}^R_t$ is the maximum allowed expenditure revised with respect to the cyclical component CC^R. The revision means that

[2]When $R - G$ is positive it is in fact a surplus.

instead of just output gap projections also the actual realizations of variables are used in the following way:

$$CC_t^R = E_{t-1}[\epsilon_t] \cdot \Big(Y_t - E_{t-1}[Y_t^*] \Big)$$

The rest of the equation for revised expenditure limit is the same as for the expenditure limit before revision, which means that the revision takes into account only an adjustment of cyclical factors. If the accumulated deficits in the correction account are larger than 1% of GDP (i.e. $CA_t < -0.01 \cdot Y_t$) then the excessive deficit must lower the maximum allowed expenditure as a correction term. The correction cannot be larger than 0.35% of GDP though. Finally, a correction is applied only if the economy is in an upturn. It all means that:

$$c_{t+1} = \max\{CA_t + 0.01 \cdot Y_t\ ,\ -0.0035 \cdot Y_t\}$$

if

$$CA_t < -0.01 \cdot Y_t \quad \text{and} \quad E_t[Y_{t+1}] > E_t[Y_{t+1}^*]$$

and 0 otherwise.

The framework of the rule does not provide any special treatment of one-off extraordinary revenues (e.g. auctions of TV frequencies), contrary to the Polish mechanism. The existence of any funds being outside of the scope of the rule's limits is not allowed. The rule pertains also to state ('Bundesländer') budgets with a difference that their budgets must be structurally fully balanced. The law regarding states starts to be binding in 2020.

The rule involves the necessity of calculating the cyclical component of the government budget. It is stipulated in the law that the method that must be applied in the calculation of this component is the European Commission's production function approach used together with semi-elasticities of the budget balance with respect to the output gap. A detailed exposition of the production function method is presented in the paper by European Commission (2014b), while the way of obtaining semi-elasticities is given in European Commission (2014a). In order to get the potential output Y^*, which is needed to calculate the output gap, the Cobb-Douglas production function must be applied with the use of potential values of capital, labor, TFP and capital/labor weights. The projections, made by German fiscal authorities together with calculations of the budget balance semi-elasticities, are subject to many arbitrary decisions and, potentially, give enough degrees of freedom to manipulate effectively the fiscal policy. Being unable to credibly project the potential output using the EU Commission approach, we have decided to apply the modified HP filter, used in the same way as in the Swiss fiscal rule, to calculate the trend output, as a measure for the potential output. This approach makes it possible to focus rather on the comparison of various aspects of fiscal rules than on various trend/potential output calculation methods. The sum of semi-elasticities applied in the simulations is constant and equal to 0.205, which is the value obtained by the German Ministry of Finance in 2015.

The German fiscal rule has been criticized in Truger and Will (2012). The main reason for their critique is the use of the European Commission production function trend calculation method, which is made even more opaque by the allowance in German law to apply any modifications that are 'justified by the newest state of knowledge'. Truger and Will (2012) point out that the German rule, also through channels other than the trend calculation, is prone to various interpretations and manipulations. Other problems envisaged by these authors are the arbitrariness of the 0.35% structural deficit limit, tendency for procyclicality and high conservativeness of the rule, which are all problems of more subjective nature. It is worth noting that the German fiscal policy after 2011 is even more conservative than it would be when sticking precisely to the rule. Paetz, Rietzler, and Truger (2016) claim that it is so because of very favorable conditions for the German economy and the fact that the rule itself is procyclical: in good times the rule is not that strict but had the environment been worse the rule would have been binding as the economic conditions would worsen more than the expenditure limit would increase.

3.2.3 Polish rule

'The stabilizing expenditure rule' in Poland ('stabilizująca reguła wydatkowa' in Polish) is described in Korniluk (2016). It was added to the Polish legal system in 2013 and the first national budget calculated with its spending limit was in 2015.

The principle of the rule is to let government expenditure grow not faster than the rate of the medium-term real GDP growth. The rule is summarized by the following equation:

$$\bar{G}_{t+1} = G_t \cdot E_t[\Pi^*_{t+1}] \cdot (y^*_{t+1} + c_{t+1}) + E_t[\Delta\ dR_{t+1}],$$

where $\bar{G}$ is the maximum allowed expenditure limit for next period, G is the government expenditure in a given period, $E[\Pi^*]$ is the central bank inflation target, y^* is a medium-term real GDP growth, c is the correction term explained below and $E[\Delta\ dR]$ is a change in 'large discretionary revenue'. The medium-term real GDP growth y^*, which is a geometric mean over 8 years, is given by the following formula:[3]

$$y^*_{t+1} = \Big(\frac{E_t(Y_{t+1})}{Y_{t-7}}\Big)^{\frac{1}{8}}.$$

The Polish rule uses the correction account CA to store deviations of deficits $R - G$, relative to GDP, from the medium-term objective that is set to -1% of GDP:[4]

$$CA_t = CA_{t-1} + \frac{R_t - G_t}{Y_t} + 0.01.$$

Finally, corrections c are given as follows:

- if $\frac{R_t - G_t}{Y_t} < -0.03$ or $\frac{D_t}{Y_t} < -0.55$ then $c_{t+1} = -0.02$;
- if $\frac{R_t - G_t}{Y_t} \geqslant -0.03$ and $-0.55 \leqslant \frac{D_t}{Y_t} < -0.50$ and $E_t[\frac{Y_{t+1}}{Y_t}] \geqslant y^*_{t+1} - 0.02$ then $c_{t+1} = -0.015$;
- if $\frac{R_t - G_t}{Y_t} \geqslant -0.03$ and $\frac{D_t}{Y_t} \geqslant -0.50$ and $CA_t < -0.06$ and $E_t[\frac{Y_{t+1}}{Y_t}] \geqslant y^*_{t+1} - 0.02$ then $c_{t+1} = -0.015$;
- if $\frac{R_t - G_t}{Y_t} \geqslant -0.03$ and $\frac{D_t}{Y_t} \geqslant -0.50$ and $CA_t > 0.06$ and $E_t[\frac{Y_{t+1}}{Y_t}] \leqslant y^*_{t+1} + 0.02$ then $c_{t+1} = 0.015$;
- otherwise, $c_{t+1} = 0$.

The rule does not contain unobservable terms like structural/cyclical balances or trends, whose calculation method can be disputable or opaque. Nonetheless, it is far from being transparent because of the complicated correction mechanism. This mechanism serves the purpose to decrease the growth of expenditure (corrections -2pp and -1.5pp) when the government is heavily indebted, it just ran a high deficit, GDP growth is high or the control account accumulated substantial deficits. In an opposite situation, when the economic growth is low, deficits and debt are low enough and the control account accumulated surpluses, public expenditure is allowed to grow faster than the medium-term economic growth rate as a positive correction applies ($+1.5$pp).

[3]Empirical research indicates that business cycles in Poland, similarly to other developed countries, have a maximum length of 8 years, for details see Korniluk (2016). The formula is equivalent to $y^*_{t+1} = \big(\frac{Y_{t-6}}{Y_{t-7}} \cdot \frac{Y_{t-5}}{Y_{t-6}} \cdot \ldots \cdot \frac{E_t[Y_{t+1}]}{Y_t}\big)^{\frac{1}{8}}$.

[4]Recall that a positive deficit $R - G$ is in fact a surplus. The same applies to the debt: a positive value of D means an accumulation of assets.

It is worth noting that in the initial formula of the Polish expenditure rule there appeared a projection of CPI dynamics, which was replaced by the central bank (National Bank of Poland - NBP) inflation target in 2015. The NBP inflation target was higher than the actual inflation, which allowed larger expenditure. Clearly, such a change does not instill confidence in the rule, particularly in view of the fact that it is not a part of the constitution but it can be changed as an ordinary law. Additional changes in the rule framework were an inclusion of discretionary revenues in the expenditure limit and a change of debt thresholds in the correction definitions from 55% of GDP and 50% of GDP to 48% and 43%, respectively. The last change was implied by a pension system reform and redemption of some pension bonds. In our simulation we use the 'old' debt levels i.e. 50% and 55%.

The results of a simulation in Korniluk (2016) indicate that the Polish fiscal rule should lead to debt levels around 20% of GDP. Nonetheless, the rule does not imply an anticyclical fiscal policy because the output gap and cyclically adjusted budget balances are not correlated, which suggests acyclicality.

3.2.4 Balanced budget rule

We define the balanced budget rule in this paper so that government expenditure equals realized government revenue:

$$\bar{G}_{t+1} = R_{t+1}.$$

To the best of our knowledge such a rule is not operational in any developed country at the national level. It is popular though at the US state or Swiss cantonal level. Nevertheless, it must be pointed out that these rules vary considerably as they deal differently with deficits realized ex post. This is so because the balanced-budget rule should be rather written as $\bar{G}_{t+1} = E_t[R_{t+1}]$ with a precise instruction on what to do in case of deviations from the expected values. In order to avoid making fully discretionary decisions on the rule mechanics we have decided to apply its most stringent form: the government cannot run any deficits at all and must equate its spending to its realized revenue all the time.

Although the rule is extremely transparent and simple conceptually because there is no need to calculate unobservable structural or cyclical components, its limited popularity can be traced back to its procyclicality. This rule is treated then in this research as a benchmark. We expect it to be the most procyclical rule among the analyzed rules. Empirical analysis of US state rules (Alesina and Bayoumi (1996)) and Swiss cantonal rules (Luechinger and Schaltegger (2013)) show that variations of balanced budget rules lower deficits and can lead to more accurate revenue projections. Moreover, Alesina and Bayoumi (1996) show that balanced budgets do not necessarily increase output volatility, yet this claim may be invalid in view of the fact that state governments in the US have little impact on output stability in comparison with the federal government.

3.3 Simulation methodology

The main component of the simulation framework is a reduced-form VAR(2) model that is explained in detail in Section 3.4. The econometric model is used on US data from years 1960 to 2015 and contains the following endogenous variables: output Y, public revenue R and public expenditure G, as well as an exogenous variable '$crisis$', which indicates an occurrence of recessions.

The strategy applied in our simulation is to use the VAR parameter estimates $\hat{\beta}$, based on the empirical data, to compute three endogenous variables Y_{t+1}, G_{t+1} and R_{t+1}, knowing their initial values in period t, as well as the value of the exogenous variable $crisis_{t+1}$. Then, the value of expenditures in period $t+1$ (i.e. G_{t+1}), which is predicted with the VAR model, is substituted with the value $\bar{G}_{t+1}$, obtained with a given fiscal rule. Fiscal rule expenditure limits are functions of current and past variables

$$\bar{G}_{t+1} = f(Y_t, R_t, G_t, Y_{t-1}, R_{t-1}, G_{t-1}, ...),$$

which makes their computation feasible. Finally, we move to a next step of simulation, in which the new current state variables are output, revenues and expenditures obtained in the previous step using the VAR relationship supplemented with the expenditure policy implied by a fiscal rule. The precise algorithm of simulations for every rule is described step by step in Appendix 3.7.

The simulation approach relies on a series of assumptions, which are summarized and explained below:

- Expenditure limits are binding, i.e. politicians are willing to spend as much as possible and they exploit all the space given by the rules. The purpose of this assumption is that we want to assess the properties of the fiscal policy implied by the rules, i.e. when they are binding.
- Fiscal rules encompass all the public expenditure without any exceptions for any special funds or expenditure types.

- All planned expenditures are incurred and there are no unplanned expenditures. There are no irregular one-off revenues and there are no financial transactions (e.g. privatization), which reduces respective terms in the Polish and German rules: $\forall t\ E_t[F_{t+1}] = 0$ and $\forall t\ E_t[\Delta\ dR_{t+1}] = 0$.

- Revenues are 'exogenous' from politicians' perspective so that all adjustments required to respect rule's limits are made through changing expenditures.

- There is no inflation, which means that nominal variables are equal to real variables. The inflation term in the Polish rule is always equal to one: $\forall t\ E_t[\Pi^*_{t+1}] = 1$.

- The simulation abstracts from default risk and interest payments. The latter assumption, although seemingly strict, is, in fact, fully warranted by the construction of the rules, which specify the maximum allowed *total*, not *primary*, deficits. As government expenditure is equal to primary expenditures plus interest payments, $G = GP + i \cdot D$, an increase in the interest rate i, for a given maximum spending limit G, must lead to lower primary expenditures GP and does not affect total deficit $R - G$ nor debt level D. It is beyond the scope of this chapter to assess which fraction of the interest payments in total government spending is beyond social acceptance. Furthermore, we assume that the debt level is neutral for output growth. Debt accumulation is defined as $D_t = D_{t-1} + (R_t - G_t)$.

- Computation of expenditure limits for all rules, but the balanced-budget one, involves projecting revenues and output one period (year) ahead. In all places where such a forecast is required the following procedure is used. Firstly, logs of a time series are calculated. The resulting time series is fed into the function that automatically identifies and estimates the best ARIMA model. The function uses a stepwise method based on Akaike information criterion.[5] The length of a time series that is used to make a forecast is not constant, further observations are added to the time series as the simulation sample gets longer. Based on the identified and estimated model a point forecast is created.[6] We apply ARIMA forecast functions because these models have parsimonious specifications and it can be shown (see Favero and Marcellino (2005)) that ARIMA fiscal forecasts perform very well.

- It is important to note that all fiscal rules use projections $E_t[x_{t+1}]$ but also $E_t[x_t]$. This is so because the process of planning a budget for period $t + 1$ starts well in the middle of period t so variable x_t is not yet fully realized. In the simulations, variables x_t are assumed to be already known in period t or, equivalently, to be

[5] `http://reference.wolfram.com/language/ref/TimeSeriesModelFit.html`
[6] `http://reference.wolfram.com/language/ref/TimeSeriesForecast.html`

perfectly forecasted. This last assumption seems to be rather innocuous considering the fact that projections of annual GDP or revenues are rather accurate when the government has partial data from the first or first two quarters of a given year.

- At the time of the introduction of the rules, countries start without any deficits and surpluses on their correction accounts CA. This assumption is realistic because all three rules started with a neutral 'record' when they were introduced.

- The German rule uses the modified HP filter (identical to the one used in the Swiss rule) to calculate the trend/potential output. This modification is due to the arbitrariness and complexity of the production function method specified in the actual German rule. The modified HP filter is described in detail in Appendix 3.8.

- Values of output Y_{t+1}, expenditures G_{t+1} and revenues R_{t+1} are computed using a reduced-form VAR and previous values of Y, G and R, without imposing any structural identification. A substitution of VAR-generated expenditures G_{t+1} with the ones specified by fiscal rule limits (i.e. $\bar{G}_{t+1}$) does not affect contemporaneous output Y_{t+1} nor revenues R_{t+1}.

- In every simulated path for every rule an independent path of exogenous crisis shocks is generated. The binary variable 'crisis' is created using a Markov chain, whose characteristics were calibrated to mimic the actual path of crises in years 1960-2015 in the US.

- The only source of randomness in the simulations are the VAR error terms and a stochastic occurrence of exogenous economic crises. They are the reason why realizations of output Y_t and revenues R_t may differ from their expectations $E_{t-1}[Y_t]$ and $E_{t-1}[R_t]$, which affect the expenditure limits $\bar{G}_t$.

The simulation uses the VAR model to mimic the economic relationships between output, government revenues and government expenditures in a realistic way. Moreover, the simulation includes an exogenous variable, which describes the impact of economic crises on endogenous variables. The use of this variable lets the VAR model better fit the data and better map the nature of a business cycle impact on budget balances. Binary variables for 'crises' create deeper recessions than just VAR error terms alone as they would average all recessions and would not put enough strain on budget balances. Therefore, the approach we have used makes measuring how fiscal rules react in difficult times more credible.

The framework adopted in this research is subject to the Lucas critique as it imposes new policies on economic relationships estimated on past behavior. It implies that we cannot answer the question 'what would have happened with the US debt and deficits had the American government adopted one of the fiscal rules at hand ?'. In order to answer this question we would have to assume that we hold the behavior of households constant against different government policies, which is an assumption that we do not want to maintain. Alternatively, assuming that the Ricardian equivalence holds would imply that interdependencies between output, government expenditure, government revenues and private consumption stay relatively constant as government expenditures and private consumption are perfect (or, at least, close) substitutes. Leaving aside the empirical validity of the Ricardian equivalence, under such an assumption it would not make sense to assess anticyclicality and volatility of fiscal policy as private consumption would adjust to substitute government expenditures.

Instead, we build a framework that makes it possible to assess the mechanics of some fiscal rules by applying them to realistic, although artificial, economic relationships. Our model credibly mimics the behavior of business cycles, especially with respect to the strain that is put by a recession on the government budget. It allows us to compare some properties of the fiscal policies implied by the rules at hand and draw conclusions about rules' fundamental characteristics.

3.4 VAR model used for simulations and the data

The basic step of the simulation is based on a reduced-form VAR(2) model. The model takes a standard form:

$$V_t = \beta_0 + \beta_1 \cdot V_{t-1} + \beta_2 \cdot V_{t-2} + \beta_3 \cdot X_t + e_t \ ,$$

where V_t is a vector of endogenous variables, X_t is a vector of exogenous variables, e_t is the vector of error terms and β_0 is a vector of constants. Vector V_t consists of logs of real total output, revenues and expenditures, $V_t = [\ln(Y_t), \ln(R_t), \ln(G_t)]'$. In addition to the endogenous variables, one exogenous variable is used. The exogenous variable is a binary indicator that encodes information about economic crises occurring in a given year, $X_t = [crisis_t]'$.

The endogenous variables are US, annual, real (measured in 2009 US dollars) output, federal revenues and federal expenditures between 1960 and 2015. The source of data is Federal Reserve Bank of St.Louis (FRED). The exogenous variable is a binary indicator for years in which there was an economic recession recorded by NBER, whose origins were deemed to be exogenous to the American fiscal policy. The variable 'crisis' encompasses the economic recessions related to the 1st and 2nd Oil Crisis, 1st Gulf War, September 11th attack and the Great Recession. It implies values of one for the following years in the data sample: $\{1974, 1975, 1982, 1991, 2001, 2008, 2009\}$.

Based on the mentioned variables a number of lags was selected using information criteria. A lack of any theoretical considerations implying a given number of lags for models using annual data a parsimonious two-lag specification was chosen. The model was then estimated equation by equation with the OLS method (equivalent to conditional maximum likelihood estimation). The estimated model has been put through standard diagnostics including tests for normality, serial correlation etc. The only problematic part is related to the possible heteroskedasticity. Since it does not seem to be severe and the model is estimated using OLS technique, which is robust to heteroscedasticity in view of the model application, the estimated values of parameters are used in further simulations. Details of the VAR model results and its diagnostics are presented in Appendix 3.9.

Figure 3.1 depicts impulse response function graphs. The IRFs are calculated after imposing structural form restrictions on the VAR model. These are in the form of a Cholesky identification with the following assumptions on contemporaneous interdependencies:

- Expenditures are not affected contemporaneously by GDP nor revenues.
- GDP is affected by expenditures but not by revenues.
- Revenues are affected by both GDP and expenditures.

The structural identification is similar in spirit to Blanchard and Perotti (2002) by using the institutional framework of fiscal policy to form the restrictions. It is assumed that expenditures are decided before a given fiscal year and, therefore, they are not affected contemporaneously by other variables. They can affect though the level of economic activity and, then, government revenues. Moreover, it is assumed that the level of economic activity affects tax revenues immediately but there is no reverse effect. This is a discretionary, but necessary, assumption that is needed to achieve identification. The structural identification is identical as in Landon and Smith (2017).

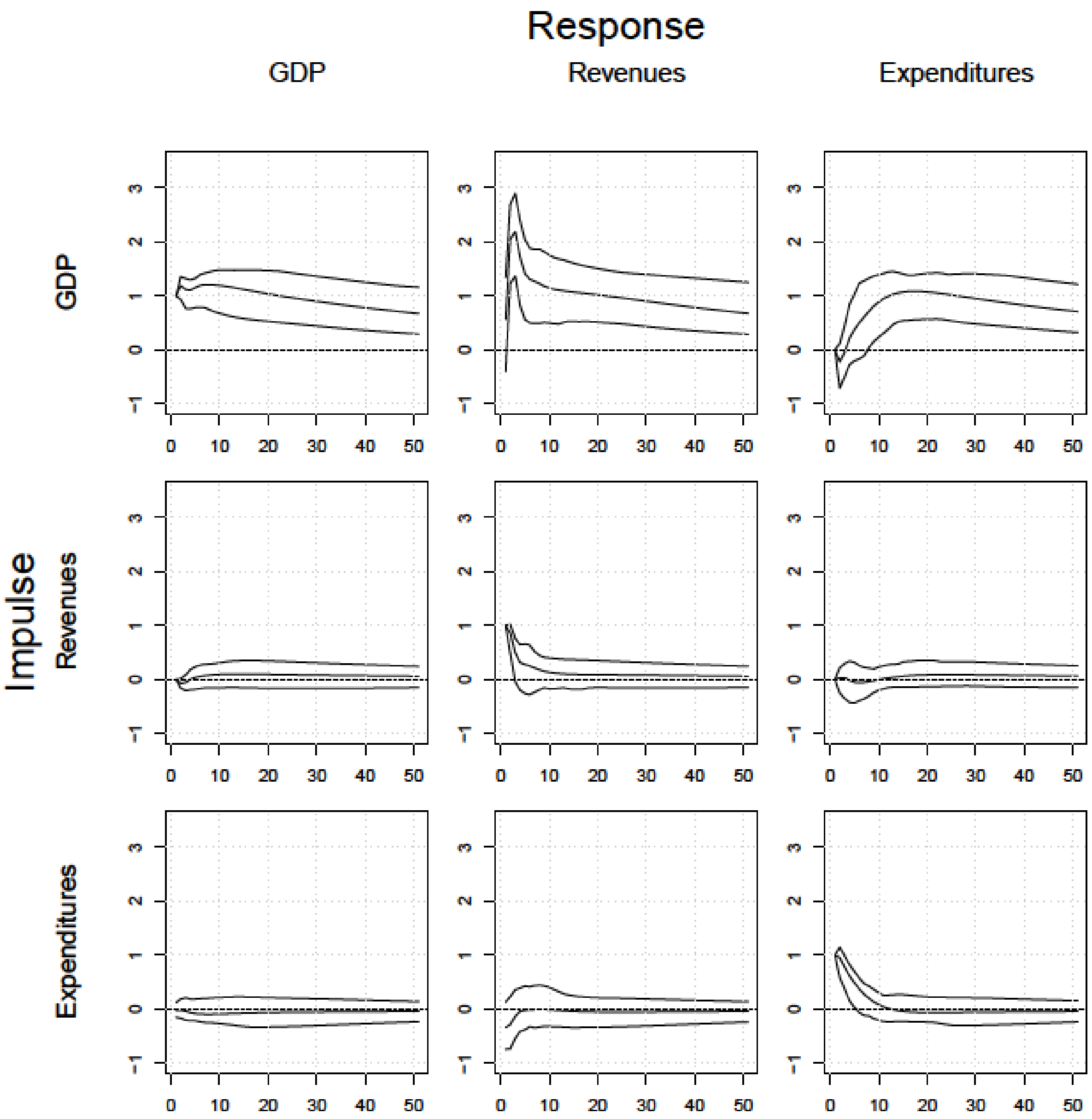


Figure 3.1: *Impulse response functions for a unitary shock in GDP, revenues and expenditures. Confidence intervals are the 5th and 95th percentiles.*

The obtained IRF results indicate insignificant responses of GDP to shocks in expenditures and revenues, which is in line with Landon and Smith (2017), Burriel et al. (2009) or other authors. This may be resulting from the fact that impulse responses of GDP are deemed to be short-lasting with their positive effects up to 3-5 quarters, which may be hardly visible in annual data.

Exogenous paths of crises are generated to mimic empirical data. The maximum-likelihood estimation with binary variables in years $\{1974, 1975, 1982, 1991, 2001, 2008, 2009\}$ over the whole sample 1960-2015 leads to a Markov chain with the following transition probabilities between crisis (c) and non-crisis (nc) states:

$$\begin{bmatrix} P(nc|nc) & P(c|nc) \\ P(nc|c) & P(c|c) \end{bmatrix} = \begin{bmatrix} 0.896 & 0.104 \\ 0.714 & 0.286 \end{bmatrix}.$$

For every simulation a crisis path is generated independently using the described probability matrix. On average, a crisis occurs every 8.6 years.

3.5 Results

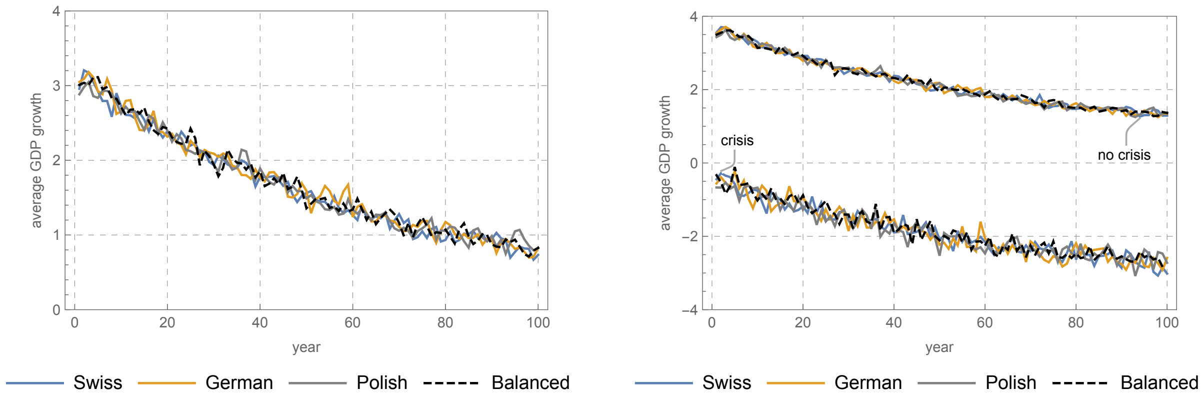
The simulation serves the purpose to analyze the trajectory of debt and deficit levels implied by the rules and to measure their potential for expenditure stabilization. Moreover, the procyclicality of the induced fiscal policy is assessed.

Each simulation starts with the empirical US values from years 1960 and 1961 for all endogenous variables and, then, the VAR model without any fiscal rule is run for 30 periods. This is necessary because the modified HP filter applied in the Swiss and German rules uses a 24-observation rolling window to measure the trend. After 30 periods the fiscal rules are applied and the simulation is run for another 100 periods. The result of these 100 periods is reported in this section. We run 500 simulations for each rule.

The initial debt level for all rules is set to 0% and 50% of GDP because the Polish rule has a built-in mechanism trying to prevent debt accumulation over 50% and 55% of GDP. An inclusion of 2 starting debt levels serves the purpose to asses if the rule is effective in attaining its goal. Other rules are expected to behave identically in both scenarios as their formulas do not depend on debt levels. In order to conserve space only some of the results with 50% of the initial debt to GDP ratio are shown.

A general result of the simulations, coming directly from the construction of the VAR model, is that the rules do not differ with respect to the output stabilization as the fiscal policy affects the output in a very limited way. It is shown by the size of the output gap, defined as $(Y - Y^*)/Y$, whose values, averaged over all 500 simulations for each of the rules, are around 0 and differ between the rules by the order of magnitude of tenths of a basis point of GDP.[7] The same applies to average minimum and maximum output gaps, whose values are around −4.8% and 4.8% of GDP, respectively.

Finally, the simulations reflect a gradual drop of average GDP growth rates from slightly above 3% in the first period to slightly less than 1% in the last period of the simulation, see Figure 3.2a. Decreasing growth rates affect debt accumulation processes and put additional pressure on the rules. Figure 3.2b shows average growth rates in periods in which a crisis occurs and in which there is no crisis. Both graphs indicate no differences between the four assessed rules.



(a) *Average overall growth rates*

(b) *Average growth rates for crisis and non-crisis periods*

Figure 3.2: *Average growth rates for* 0% *of initial debt level. Values are averaged across all 500 simulations of a given rule at given points in time. Note that Figures (a) and (b) have different scales on the vertical axes.*

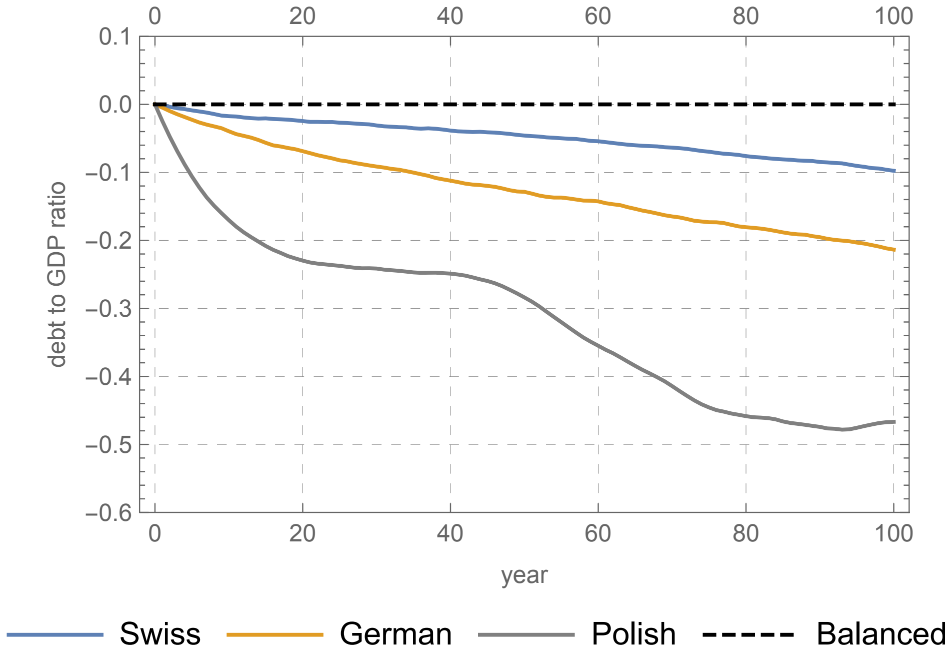
3.5.1 Debt accumulation and stabilization of deficits and expenditures

Figures 3.3a and 3.3b show debt accumulation paths, in debt to GDP ratios, for all four rules over 100 years. Depicted lines are average paths calculated over 500 simulations for every period (median paths would be very similar, which applies for all graphs). The volatility of debt accumulation paths is shown in Figure 3.4 by presenting shaded areas that contain debt to GDP ratios between the 5^{th} and 95^{th} percentile for every period.

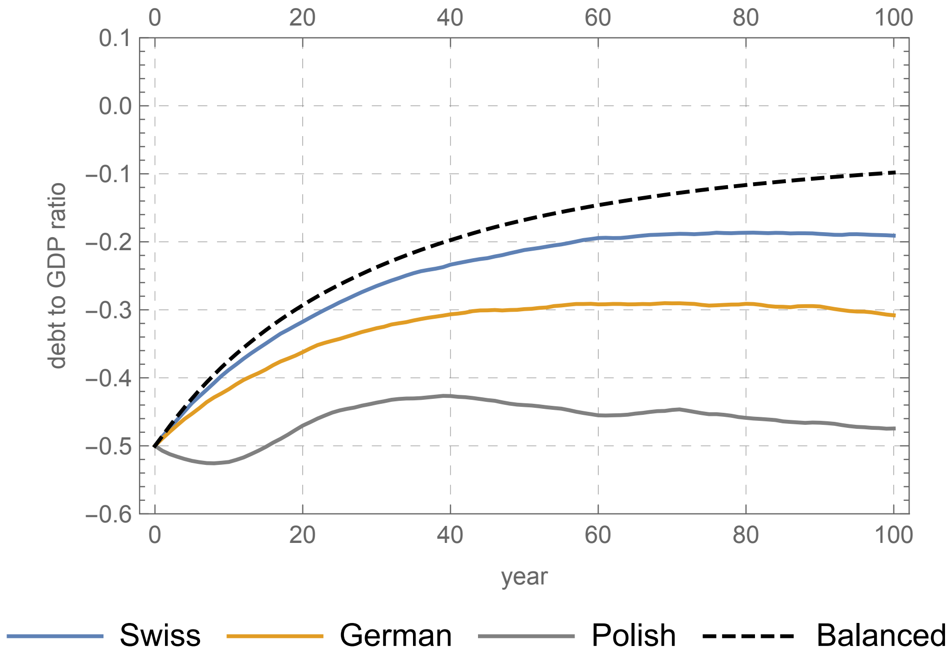
Figure 3.5 presents density functions estimated on histograms of deficits to GDP ratios calculated over all 100 periods for every one of 500 simulations. Figure 3.6 shows in turn

[7] Y^* is the potential output calculated independently for every simulation as a HP trend of the whole 100-year series of GDP.

deficit paths for the whole simulated period, averaged over all 500 scenarios. This graph enables a comparison of average deficits ‘across’ different simulated scenarios. Finally, Table 3.1 gives an overview on basic descriptive statistics on deficits to GDP ratios and expenditures to GDP ratios. All statistics in the table are calculated as average values on 500 simulations of given statistics that are, in turn, calculated on the whole 100-period path of a simulation at hand.

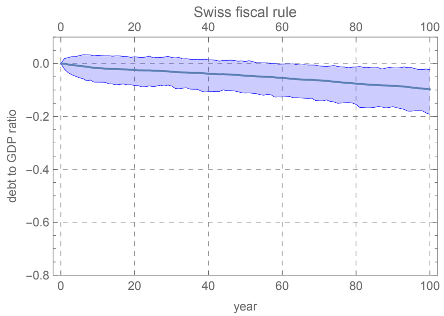


(a) *Average debt accumulation, 0% initial debt*

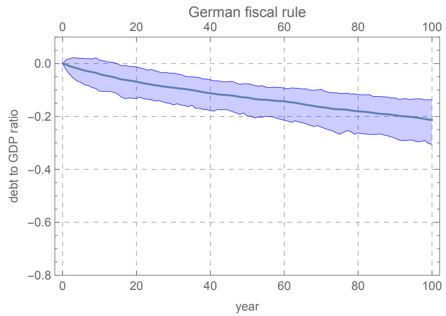


(b) *Average debt accumulation, 50% initial debt*

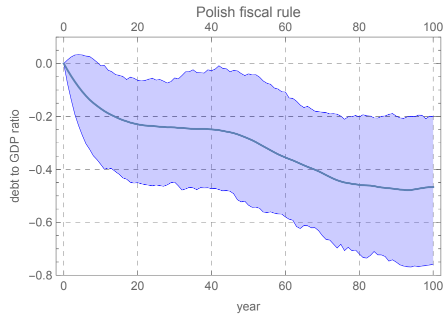
Figure 3.3: *Average paths of debt accumulation for* 0% *and* 50% *of initial debt level. Values are averaged across all 500 simulations of a given rule in a given point in time.*



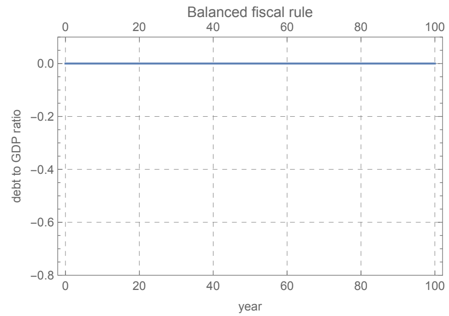
(a) *Swiss rule, 0% initial debt*



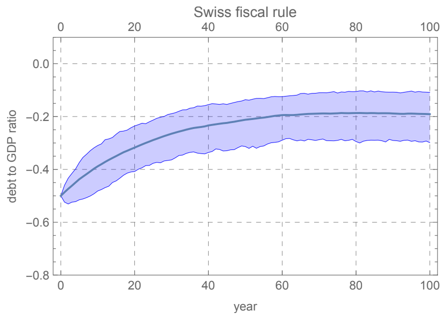
(b) *German rule, 0% initial debt*



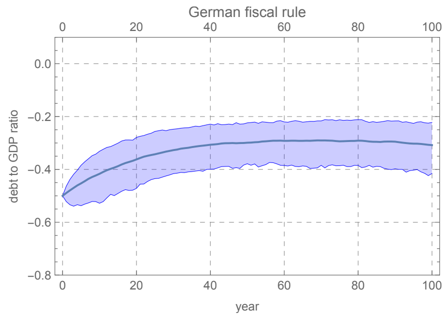
(c) *Polish rule, 0% initial debt*



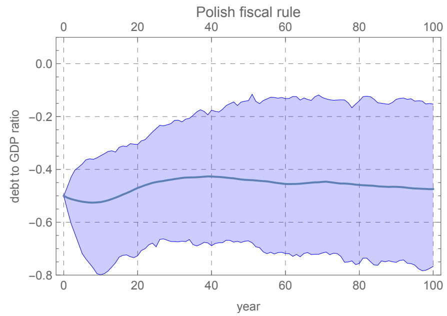
(d) *Balanced rule, 0% initial debt*



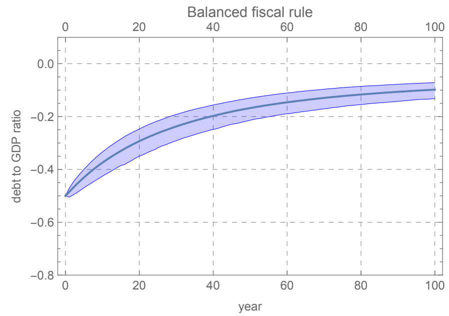
(e) *Swiss rule, 50% initial debt*



(f) *German rule, 50% initial debt*



(g) *Polish rule, 50% initial debt*



(h) *Balanced rule, 50% initial debt*

Figure 3.4: *Debt accumulation paths for different rules and initial debt levels. Thick lines are average paths over all 500 simulations in a given time point. Shaded area is the 90% interval between the* 5^{th} *and* 95^{th} *percentile values.*

Figures 3.3a and 3.3b show that all rules prevent explosive debt paths. Nevertheless, they differ significantly when it comes to details of the debt stabilization process. The balanced budget, Swiss and German rules rely on a similar tenet, which is balancing the real or structural budget and which implies their qualitatively similar behavior. The German rule is less stringent than the Swiss one by construction, i.e. by allowing the structural deficit to be 0.35% of GDP, and it leads to a larger accumulation of debt in the long run. The Swiss scheme mimics in turn quite closely the balanced budget rule in terms of the debt (de)accumulation and only in the very long run, that is after 100 years, differences are clearly visible. A systematically different result is given by the Polish expenditure rule. Debt is stabilized at much higher levels than in case of the other rules and its accumulation path is not necessarily strictly monotonic. Nevertheless, the rule prevents, on average, debt accumulation above 50%, which means that it attains its prescribed goal. Lastly, the Swiss, German and balanced-budget rules offer similar, and rather limited, volatility in their debt to GDP paths, which is reflected by Figure 3.4. The Polish rule proves here to be much less reliable with this respect.

The debt levels converge to equilibria given by the formula $d^* = (r - g)/y$, where $d^* = (D/Y)^*$ is the equilibrium debt to GDP ratio, $r = R/Y$ and $g = G/Y$ are the expenditure and revenue to GDP ratios and y is the GDP growth rate.[8] Taking into account that deficit to GDP ratios $r - g$ implied by the rules oscillate around their long-term average values, the equilibrium debt ratio d^* is a function of the GDP growth rates that differ across time: $d^*(y_t)$. Because of the decreasing GDP growth rates y_t, the long-run debt equilibria $d^*(y_t)$ implied by the Swiss and German fiscal rules also decrease, which is visible at the end of the simulated sample.

Expectations regarding the debt stabilisation features of the German and Swiss rules, listed in Truger and Will (2012) and Beljean and Geier (2013) respectively, are not based on theoretical nor empirical inference but rather on intuition. In our research we corroborate them formally and confirm that these rules are austere and do not allow for large debt accumulation, even in times of decreasing income growth rates. The results we obtain with respect to debt stabilisation properties of the Polish rule are not in line with the results of Korniluk (2016), who projects the debt level to stabilise at around 20% of GDP in the long run. As explained in the previous paragraph, such results rely heavily on the GDP dynamics. In our model the Polish rule allows for larger, but seemingly still stable, debt levels.

[8]It is a formula for an equilibrium of the debt accumulation difference equation $\Delta(D/Y) = R/Y - G/Y - y \cdot (D/Y)$. Interest rate i does not show up in the equation as interest payments are already included in expenditures G.

Results on the debt stabilization are easily explained by the incidence of deficit to GDP ratios, presented in Figure 3.5. The Swiss rule leads to deficits that are heavily and symmetrically concentrated around zero, while the German rule allows for slightly higher deficits. For both these rules deficits and surpluses are nearly always between +/- 3% of GDP. Deficits governed by the Polish rule are centered around their desired level at -1% of GDP yet their dispersion is much more pronounced as they vary between about +/- 6% of GDP.

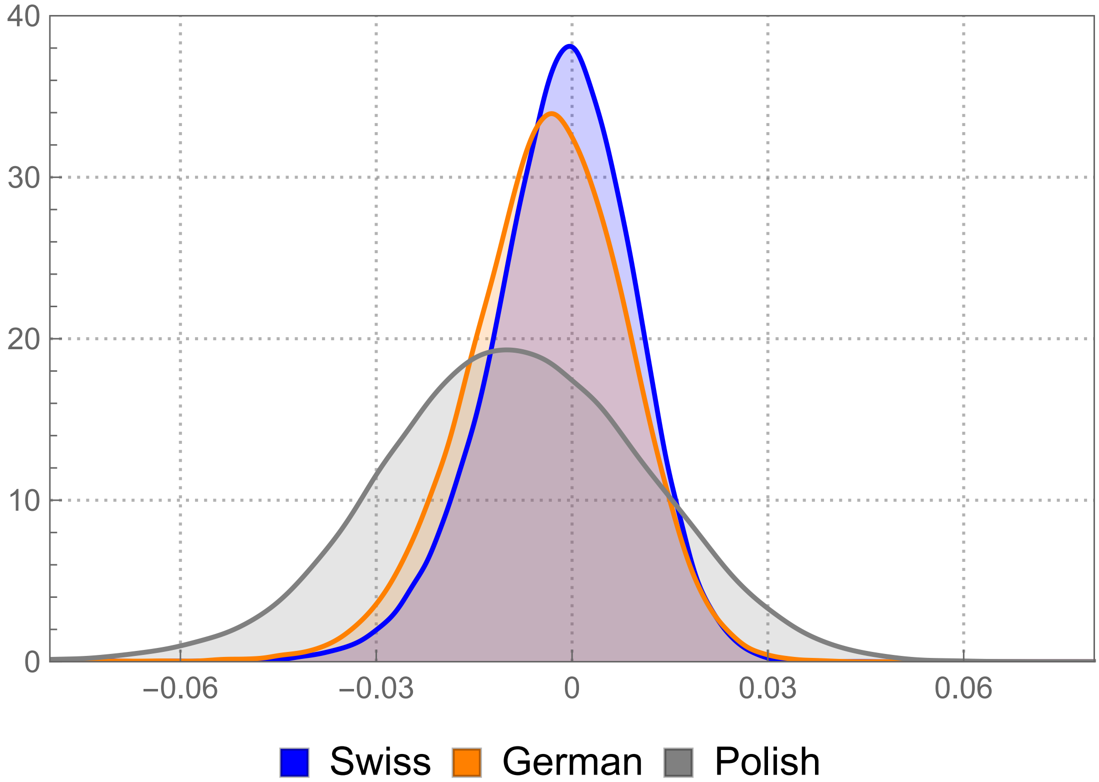


Figure 3.5: *Estimated density functions for all deficit to GDP ratios (averaged across time and simulated scenarios) with 0% of initial debt.*

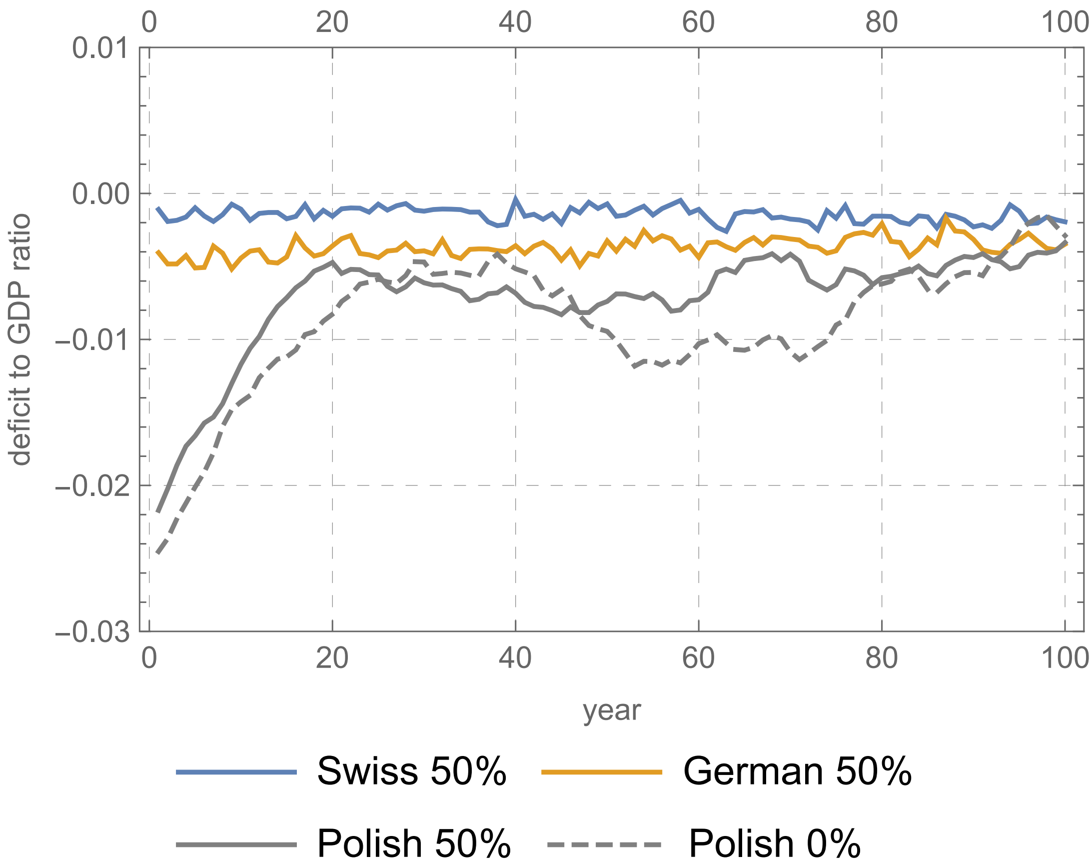


Figure 3.6: *Paths of average deficits for given rules at a given point in time.*

Figure 3.6 shows average deficit paths for the Swiss, German and Polish rules. Swiss and German rules are not affected by the initial debt levels so their results are not reported twice. The trajectory of the Polish rules depends on the starting point of debt, which is visible in the graph. A large debt ratio induces more frugal fiscal policy till debt-stabilizing deficit ratios are attained. When the debt levels are below the limit of 50% of GDP decreasing income growth rates can lead to larger debt accumulation until the debt limit is reached.

Table 3.1 sums up basic statistics on deficit and expenditure ratios, measured for each simulated path and, then, averaged across different simulation scenarios. Standard deviations and ranges indicate that the Polish rule is more volatile when it comes to both deficits and expenditures.

The Polish rule offers the highest degree of stabilizing expenditures period to period, which is proved by the value of expenditure autocorrelation measure. The Swiss and the German rules stabilize deficits at the cost of curbing expenditures, which have to vary significantly

Table 3.1: *Descriptive statistics of deficit to GDP ratios and expenditures to GDP ratios 'within' simulated paths for 0% of initial debt.*

Deficit to GDP ratios

Statistics	Balanced	Polish	Swiss	German
min	0.	−0.058	−0.032	−0.036
max	0.	0.035	0.023	0.024
mean	0.	−0.009	−0.002	−0.004
std. dev.	0.	0.020	0.010	0.012
range	0.	0.093	0.055	0.061
autocorr(1)	1	0.849	0.194	0.269

Expenditure to GDP ratios

Statistics	Balanced	Polish	Swiss	German
min	0.130	0.137	0.130	0.131
max	0.196	0.213	0.203	0.207
mean	0.163	0.174	0.164	0.167
std.dev.	0.014	0.019	0.015	0.016
range	0.065	0.077	0.072	0.077
autocorr(1)	0.791	0.944	0.658	0.687

between years. In case of the Polish rule it is the stabilization of expenditures, which are very persistent year to year, that leads to deficits that are volatile in the long-run, i.e. over the whole 100-period simulation. High persistence of expenditure ratios induced by the Polish rule leads to longer adjustment phases and their large long-run variation, measured by their range or standard deviation.

The reason behind this can be, for example, an introduction of the rule in a long-lasting good phase of the business cycle, which lets expenditure ratios grow to large levels, while their decrease caused by worse economic conditions is slow. Figure 3.6 shows that it takes the Polish rule even 20 years before the large expenditures, created by more profligate non-rule VAR model, are lowered enough to stabilize debt. Swiss and German rules sharply decrease expenditures in such circumstances even by a couple of GDP pp within one period. This implies however that, in practice, the Polish rule can be implemented without any pre-introduction transitory periods, while the other rules may necessitate some fiscal policy adjustments till they can fully shape deficits according to their tenets.

3.5.2 Procyclicality measures

We measures procyclicality as in Alesina et al. (2008) or Guerguil et al. (2017). Specifically, we rely on the following regressions:

$$\Delta \frac{G_t}{Y_t} = \beta_0 + \beta_1 \cdot \Delta \frac{Y_t - Y_t^*}{Y_t} + \epsilon_t, \tag{P1}$$

$$\Delta \frac{R_t - G_t}{Y_t} = \beta_0 + \beta_1 \cdot \Delta \frac{Y_t - Y_t^*}{Y_t} + \epsilon_t, \tag{P2}$$

where G_t/Y_t is a ratio of expenditure to GDP, $(R_t - G_t)/Y_t$ is the deficit to GDP ratio and $(Y_t - Y_t^*)/Y_t$ is the output gap to GDP ratio. Δ indicates a year on year change in the given variables. The trend output Y_t^* is calculated using standard HP filter on the whole 100-period time sample in which fiscal rules are active. The smoothing parameter is equal to $\lambda = 100$, which is used by, for example, Backhus and Kehoe (1992). The parameter of interest is β_1, which describes a reaction of government expenditure ratios to GDP with respect to changes in the output gap.

The above regressions, when used in an empirical setting, are expected to contain an endogenous variable, which is the output gap. The reason for this is the fact that there can be a reverse causality, caused by simultaneous interaction between fiscal policy and output. In order not to include the impact of fiscal multipliers on $\hat{\beta}_1$ instruments are used (e.g. lagged explanatory variables). The reduced-form VAR setting we have adopted precludes endogeneity by assuming no effect of fiscal policy on contemporaneous output. Therefore, no additional modifications are needed to estimate consistently the equations at hand using OLS. Results of the regressions are given in Table 3.2, which presents estimates of parameters β_1 that were calculated as averages on 500 independent paths for each of the four rules. As differences between estimates from regressions starting with 0% and 50% of initial debt to GDP ratios are negligible only the former ones are presented.

Table 3.2: *The procyclicality metrics for average paths (0% initial debt).*

Statistic	Balanced	Polish	Swiss	German
metric $P1$	0.011	−0.203	−0.304	−0.315
p-value	0.011	0.000	0.000	0.001
R^2	0.010	0.600	0.212	0.211

Statistic	Balanced	Polish	Swiss	German
metric $P2$	–	0.219	0.335	0.344
p-value	–	0.007	0.001	0.001
R^2	–	0.153	0.194	0.182

Table 3.2 indicates that all three analyzed rules are anticylical as intended, which is proven by negative and positive coefficients of metrics P1 and P2, respectively. The interpretation of the coefficients means that an increase in the output gap by 1 percentage point decreases the expenditure to GDP ratio by about 0.20 pp, 0.32 pp and 0.30 pp for the Polish, German and Swiss rule, respectively. When we hold the revenue to GDP ratio

constant, it implies that the deficit to GDP ratio changes in one-to-one proportion with the expenditure to GDP ratio, but with a reversed sign, which is corroborated by the P2 metric. No variation of deficits in the balanced-budget rule makes it impossible to obtain sensible results of the P2 regression. Metric P1 for the balanced-budget rule is close to zero, suggesting acyclicality. This result is driven by low correlation between the output gap and government revenues in the empirical US data, which implies only slight procyclicality of the revenues to GDP ratio in the VAR model used in simulations.

Taking into account that the average range of the output gap for all the rules is about 9.6 pp, it means that the deficit to GDP ratio should vary with respect to changes of the output gap in the business cycle up to, maximally, 1.9 pp for the Polish rule and up to about 2.8 pp for the Swiss and German rules, taking other factors constant.

It is crucial to note that the above results are based on the responsiveness of deficit and expenditure ratios to changes in the business cycle measured by the output gap, which is calculated 'ex post', i.e. using all realized periods. The maximum responsiveness of deficits or expenditures to the business cycle is not to be confused with the actually realized deficit ratios, which are depicted in Figure 3.5. Discrepancies between these values contain all other factors affecting deficits than the output gap. For example, they include debt levels (in case of the Polish rule) or the fact that rules miscalculate the 'ex post' output gap using projections and only partial time series samples. Moreover, an inclusion of the term $\Delta\ R_t/Y_t$ to the regressions helps explaining a larger portion of expenditure or deficit variation as revenues contain also shocks that are orthogonal to the output gap.[9]

The results confirm that all the rules are in fact anticyclical. Nonetheless, it must be pointed out that the degree of the anticyclicality is rather small economically. According to the obtained results, the difference in the deficit to GDP ratios implied by the differences in output gaps in the whole 100-year period is up to about 1.9 pp or 2.8 pp for the Polish and German/Swiss rule, respectively. Such a difference between maximum and minimum deficit ratios can be perceived as moderate.

3.5.3 Transparency and political issues

Finally, we would like to mention the lack of transparency and proneness to manipulation, from which suffer all analyzed rules. The goal of fiscal rules is to reduce the deficit bias by minimizing politicians' discretionary decisions regarding expenditure limits. Although the

[9]In case of the balanced budget rule, $\Delta\ R_t/Y_t$ explains obviously 100% of the variance of $\Delta\ G_t/Y_t$.

rules may, at least to some extent, reduce the discretion with respect to the budget balance, they seem to be far from eliminating it completely. The complication and arbitrariness of measuring potential or trend output and the possibility to bias output or inflation projections leave a lot of space to manipulate with, theoretically impartial, expenditure or deficit limits. The problem is particularly acute because rules' limits are calculated by fiscal authorities, whose direct supervisors are politicians.[10] Rules turn then into 'black boxes' producing some results that are, theoretically, in accordance with the legal framework but, in fact, may be far from their initial economic intentions.

Fiscal rules thus cannot be perceived as substitutes to fiscal councils but they should be rather treated as their complementary elements. In case fiscal authorities depending on politicians cannot credibly calculate their limitations this task should be delegated to some, ideally, impartial institutions.

3.6 Conclusions

We have analyzed the performance of the Swiss, German, and Polish fiscal rules by comparing them to one another and to the balanced-budget rule that serves as a benchmark.

We find that, firstly, all rules are capable of decreasing or, at least, stabilizing the debt to GDP ratio in the long run. The Swiss rule is very successful in this respect as it mimics the balanced budget rule. The German rule allows for a slightly higher accumulation of debt, which is a result of allowing structural deficits of 0.35% of GDP. The Polish rule stabilizes the debt ratio at slightly less than 50% of GDP. The rules perform well with respect to the debt stabilization, particularly in view of the fact that they have to face decreasing output growth rates.

Secondly, stabilization of debt and deficits comes at the cost of increasing the volatility of the expenditure to GDP ratio. Here, the Polish rule performs best in the short run as its character is inherently measured to stabilize expenditures period-to-period.

Thirdly, the rules perform less well in terms of implying anticyclical fiscal policy than in stabilizing debt. Although they seem to show anticyclical behavior, its magnitude is economically moderate. The rules react to changes in the output gap yet the deficit responses are limited. Taking into account the output gap volatility, responsiveness of

[10] It is worthy of saying here 'nemo iudex in causa sua' or, simply, 'no-one should be a judge in his own cause'.

deficits is equal to, at most, 2.8 pp of GDP between peaks of the business cycle for the Swiss and German rules and even less for the Polish one.

Finally, the discretionary nature of calculations of trend and projections makes an impartial overseeing of the rules' implementation a necessity. It implies politically independent fiscal councils, which could assess a proper working of rules' mechanisms, being not substitutes but being complementary to rules.

3.7 Appendix A: simulation algorithms

All simulation algorithms work along the same lines. Initially a state of economy is defined. The state contains all relevant pieces of information required to calculate a single step of simulation. For various fiscal rules the required pieces of information are different but in general a state contains year, previous total output history, previous revenue history and so on. The description of each algorithm contains precise delineation of a particular definition of state used.

All algorithms simulate paths in steps. After a step for a period t is calculated, that is all variables for period t are known, three consecutive parts of the next step are performed. The first part is related to a fiscal rule. The main goal of the first part of a step is to calculate an upper bound for budget expenditures for the period $t+1$. Depending on a rule it takes doing some forecasting for selected variables and updating control accounts if there is any. The next part of a step is an application of the VAR model described in Section 3.4. The results of this part are values of the three endogenous variables for period $t+1$. Finally, the last part of a step is a cleaning phase also related to a fiscal rule. In the last part of a step the value for expenditures, resulting from application of the VAR model, is overwritten with the values resulting from application of a fiscal rule in the first part of a step. Other calculations in this part may include adjusting control account or other variables internal to a fiscal rule.

Once a step is finished all values for period $t+1$ are known and the procedure is repeated. Each step is time independent and depends only on a state at time t. The result of a step is random. There is only one source of randomness that comes from application of the VAR model. Since we are interested in average behavior of a fiscal rule many simulations are run and resulting paths are averaged.

All algorithms were calculated using Wolfram language.[11] All logarithms used in the text are natural logarithms.

3.7.1 Swiss fiscal rule

A state for the Swiss fiscal rule contains the following components.

`year` A number denoting a year. This is only for convenience and is not used in any calculations.

[11] Codes used in the calculation of these algorithms are available on request as Mathematica notebooks.

`gdp`	A vector of real output required to calculate outcome of the VAR model, thus of length 2.
`revenues`	A vector of revenues required to calculate outcome of the VAR model, thus of length 2.
`expenditures`	A vector of expenditures required to calculate outcome of the VAR model, thus of length 2.
`gdpFS`	A full history of real total output. New values of total output produced by the VAR model are appended to this vector during simulation.
`revenuesFS`	A full history of revenues. New values of revenues produced by the VAR model are appended to this vector during simulation.
`gdpRollingWindow`	A 23 year rolling window of real total output required to calculate adjusted HP filter. At each step a forecast of real total output is appended to this vector producing a rolling window of width 24 as required.
`expendituresFS`	Assumed upper bound to expenditures resulting from the rule.
`deficit`	Deficit calculated as difference between revenues and expenditures.
`debt`	Debt calculated as accumulated deficit.
`correction`	Corrections calculated in the Swiss fiscal rule.
`account`	Control account used in the Swiss fiscal rule.

A step in the algorithm for the Swiss fiscal rule contains the following three parts. It is assumed that after period t all variables are known. Initial values for `deficit`, `debt` and so on can be set for a particular simulation.

1. Based on logarithm of `revenuesFS` the best model in the ARIMA class is found.
2. Based on the best model a forecast for the next period is calculated. This is taken as the expected value of revenues, denoted by $\mathcal{F}(R_{t+1})$, in the period $t+1$ that is unknown after period t but before period $t+1$.
3. Based on the logarithm of `gdpFS` the best model in the ARIMA class is found.

4. Based on the best model a forecast for the next period is calculated. This is taken as the expected value of real total output, denoted by $\mathcal{F}\left(\mathrm{Y}_{t+1}\right)$, in the period $t+1$ that is unknown after period t but before period $t+1$.

5. Based on `gdpRollingWindow` and forecasted value of total output an adjusted HP filter is calculated as described in Bruchez (2003). The last value in the filtered window is taken as the forecast of a trend.

6. Based on the forecasted total output and forecasted trend a ratio k of trend to total output is calculated.

7. Correction `correction` is calculated as maximum of 0 and a ratio of `account` to `gdp` decreased by 0.06.

8. The value of `account` is decreased by the product of calculated `correction` and `expenditures`.

9. Finally, the value of `expendituresFS` is calculated according to the formula

$$\texttt{expendituresFS} = k \cdot \mathcal{F}\left(R_{t+1}\right) \cdot \left(1 - \texttt{correction}\right)$$

The next part of the step is an application of the VAR model. This results in values of `gdp`, `revenues` and `expenditures` being overwritten with new values resulting from the VAR model. More precisely, these vectors are modified with the oldest values dropped and new values appended at the end.

The last part of the step is again related to a fiscal rule but beyond that it just takes care of rolling appropriate vectors and appending new values.

1. A new value of total output, resulting from the VAR model is appended to `gdpFS`.

2. A new value of `deficit` is calculated as the difference between revenues, resulting from the VAR model, and assumed expenditures, resulting from the fiscal rule.

3. A value of `debt` is updated by the value of deficit.

4. The last value of expenditures, resulting from the VAR model, is overwritten with the value resulting from the fiscal rule.

5. A `gdpRollingWindow` is updated.

6. A new value of revenues, resulting from the VAR model, is appended to `revenuesFS`.

7. The whole fiscal rule is recalculated using this time not forecasts but values resulting from the VAR model. This results in a new "right" value of expenditures limit. Using the difference between this new "right" value and the previous value, calculated based on forecasts in the first part of a step, the value of `account` is decreased by a difference between the revised expenditures limits and the previous value of expenditure limits.

This concludes all calculations done within a single step of a simulation.

3.7.2 German fiscal rule

A state for the German fiscal rule contains the following components.

`year`	A number denoting a year. This is only for convenience and is not used in any calculations.
`gdp`	A vector of real output required to calculate outcome of the VAR model, thus of length 2.
`revenues`	A vector of revenues required to calculate outcome of the VAR model, thus of length 2.
`expenditures`	A vector of expenditures required to calculate outcome of the VAR model, thus of length 2.
`gdpFS`	A full history of real total output. New values of real total output produced by the VAR model are appended to this vector during simulation.
`revenuesFS`	A full history of revenues. New values of revenues produced by the VAR model are appended to this vector during simulation.
`gdpRollingWindow`	A 23 year rolling window of real total output require to calculate adjusted HP filter. At each step a forecast of real total output is appended to this vector producing a rolling window of width 24 as required.
`expendituresFS`	Assumed upper bound to expenditures resulting from the fiscal rule.
`gdpFSPrediction`	One year ahead forecast of real total output require by the fiscal rule.
`gdpFSTrendPrediction`	One year ahead forecast of a trend of real total output calculated as in the Swiss fiscal rule.

`elasticity`	Semi-elasticity of budget balance with respect to the output gap. This value is fixed and taken to be 0.205.
`deficit`	Deficit calculated as difference between revenues and expenditures.
`debt`	Debt calculated as accumulated deficit.
`correction`	Corrections calculated in the German fiscal rule.
`account`	Control account used in the German fiscal rule.

A step in the algorithm for the German fiscal rule contains the following three parts. It is assumed that after period t all variables are known. Initial values for `deficit`, `debt` and so on can be set for a particular simulation.

1. Based on the log of `gdpFS` the best ARIMA model is found.
2. Based on the found best model a one ahead forecast `gdpFSPrediction` is calculated.
3. Based on `gdpRollingWindow` and `gdpFSPrediction` using adjusted HP filter, as described in Bruchez (2003), a one ahead total output trend forecast `gdpFSTrendPrediction` is calculated.
4. Based on log of `revenuesFS` the best ARIMA model is found.
5. Based on the found best ARIMA model one ahead revenue forecast is $\mathcal{F}(R_{t+1})$ calculated.
6. In the next step a value of `correction` is calculated according to the following formula

$$\texttt{correction} = \min\left(0.0035 \cdot \mathrm{Y}_t, \left(\frac{\texttt{account}}{\mathrm{Y}_t} - 0.01\right) \cdot \mathrm{Y}_t\right).$$

7. If `gdpFSPrediction` is lower than `gdpFSTrendPrediction` the value of `correction` is set to 0.
8. If `account` is lower than $0.01\mathrm{Y}_t$ the value of `correction` is set to 0.
9. Finally, the value of `expendituresFS` is calculated according to the following formula

$$\begin{aligned}\texttt{expendituresFS} = 0.0035 \cdot \texttt{gdpFSPrediction} \\ - \texttt{elasticity} \cdot (\texttt{gdpFSPrediction} - \texttt{gdpFSTrendPredition}) \\ + \mathcal{F}(R_{t+1}) - \texttt{correction}.\end{aligned}$$

The next part of the step is application of the VAR model. This results in values of gdp, revenues and expenditures being overwritten with new values resulting from the VAR model. More precisely, these vectors are modified with the oldest values dropped and new values appended at the end.

The last part of the step is again related to a fiscal rule but beyond that it just takes care of rolling appropriate vectors and appending new values.

1. The last value of real total output resulting from the VAR model is appended to gdpFS.
2. The last values of revenues resulting from the VAR model is appended to revenuesFS
3. The vector expenditures is rolled and the last value, resulting from the VAR model, is overwritten with expendituresFS.
4. The vector gdpRollingWindow is rolled and the last value of Y_{t+1} is appended.
5. In the next step the control account account is updated. This is done in three steps. In the first step a variable adj is defined as

   $$\texttt{adj} = \mathrm{Y}_{t+1} - \texttt{gdpFSPrediction} .$$

   Next, the variable amab is defined as

   $$\begin{aligned}\texttt{amab} &= 0.0035 \cdot \texttt{gdpFSPrediction} \\ &\quad - \texttt{elasticity} \cdot (\texttt{gdpFSPrediction} - \texttt{gdpFSTrendPrediction} + \texttt{adj}) \\ &\quad - \texttt{correction} - R_{t+1}.\end{aligned}$$

   Finally, the control account account is updated according to the following formula

   $$\texttt{account} = \texttt{account} - (\texttt{amab} - (\texttt{expendituresFS} - R_{t+1})).$$

6. Deficit deficit is calculated as the difference between revenues, resulting from the VAR model, and expenditures, as calculated using fiscal rule.
7. Finally, variable debt is increased by the calculated deficit deficit.

This concludes all calculations done within a single step of a simulation for the German fiscal rule.

3.7.3 Polish fiscal rule

A state for the Polish fiscal rule contains the following components.

`year` A number denoting a year. This is only for convenience and is not used in any calculations.

`gdp` A vector of real output required to calculate outcome of the VAR model, thus of length 2.

`revenues` A vector of revenues required to calculate outcome of the VAR model, thus of length 2.

`expenditures` A vector of expenditures required to calculate outcome of the VAR model, thus of length 2.

`gdpFS` A full history of real total output. New values of real total output produced by the VAR model are appended to this vector during simulation.

`expendituresFS` Assumed upper bound to expenditures resulting from the fiscal rule.

`deficit` Deficit calculated as difference between revenues and expenditures.

`debt` Debt calculated as accumulated deficit.

`account` Control account used in the Polish fiscal rule.

A step in the algorithm for the Polish fiscal rule contains the following three parts. It is assumed that after period t all variables are known. Initial values for `deficit`, `debt` and so on can be set for a particular simulation.

1. Based on `gdpFS` the best ARIMA model is found.
2. Based on the best ARIMA model a one ahead forecast of real total output $\mathcal{F}\left(\mathrm{Y}_{t+1}\right)$ is calculated.
3. Based on the `gdpFS` and calculated $\mathcal{F}\left(\mathrm{Y}_{t+1}\right)$ an expected average growth $\bar{\mathrm{y}}_{t+1}$ is calculated according to the formula

$$\bar{\mathrm{y}}_{t+1} = \left(\frac{\mathcal{F}\left(\mathrm{Y}_{t+1}\right)}{\texttt{gdpFS}_{\ t-7}} \right)^{1/8}$$

4. Based on gdpFS and calculated $\mathcal{F}(\mathrm{Y}_{t+1})$ an expected year-on-year growth is calculated as
$$\hat{\mathrm{y}}_{t+1} = \frac{\mathcal{F}(\mathrm{Y}_{t+1})}{\texttt{gdpFS}_t}.$$

5. In the next step a correction c is calculated. Correction is hardwired to particular predicates, mostly based on ratios of deficit or debt to total output.

   - If $-\texttt{deficit}/\texttt{gdpFS}_t \geq 0.03$ or $-\texttt{debt}/\texttt{gdpFS}_t \geq 0.55$ then $\mathrm{c} = -0.02$.
   - If $-\texttt{deficit}/\texttt{gdpFS}_t < 0.03$ and $0.5 \leq -\texttt{debt}/\texttt{gdpFS}_t < 0.55$ and $\hat{\mathrm{y}}_{t+1} \geq \bar{\mathrm{y}}_{t+1} - 0.02$ then $\mathrm{c} = -0.015$.
   - If $-\texttt{debt}/\texttt{gdpFS}_t < 0.5$ and $-\texttt{deficit}/\texttt{gdpFS}_t < 0.03$ and $\hat{\mathrm{y}}_{t+1} \leq \bar{\mathrm{y}}_{t+1} + 0.02$ and $\texttt{account} > 0.06\,\texttt{gdpFS}_t$ then $\mathrm{c} = 0.015$.
   - If $-\texttt{debt}/\texttt{gdpFS}_t < 0.5$ and $-\texttt{deficit}/\texttt{gdpFS}_t < 0.03$ and $\hat{\mathrm{y}}_{t+1} \geq \bar{\mathrm{y}}_{t+1} - 0.02$ and $\texttt{account} < -0.06\,\texttt{gdpFS}_t$ then $\mathrm{c} = -0.015$.
   - Otherwise, $\mathrm{c} = 0$.

6. Finally, the value of expendituresFS is calculated according to the following formula
$$\texttt{expendituresFS} = \texttt{expenditures}_t \cdot (\bar{\mathrm{y}}_{t+1} + \mathrm{c}).$$

The next part of the step is application of the VAR model. This results in values of gdp, revenues and expenditures being overwritten with new values resulting from the VAR model. More precisely, these vectors are being rolled with the oldest values dropped and new values appended at the end.

The last part of the step is again related to a fiscal rule but beyond that it just takes care of rolling appropriate vectors and appending new values.

1. A new value of total output is appended to gdpFS.

2. A new value of deficit is calculated as a difference between realized revenues, resulting from the VAR model, and assumed expenditures expendituresFS, resulting from the fiscal rule.

3. A new value of debt is calculated by adding deficit.

4. A value of account is updated as
$$\texttt{account} = \texttt{account} + \texttt{deficit} + 0.01 \cdot \texttt{gdpFS}_{t+1}.$$

5. A vector `expenditures` is rolled and a value of expenditures resulting from the VAR model is overwritten with `expendituresFS`, resulting from a fiscal rule.

This concludes all calculations done within a single step of a simulation for the Polish fiscal rule.

3.7.4 Balanced budget rule

The balanced fiscal rule is by far the simplest among the compared fiscal rules. A state for the balanced fiscal rule contains the following components.

- `year` A number denoting a year. This is only for convenience and is not used in any calculations.
- `gdp` A vector of differenced logarithm of real output required to calculate outcome of the VAR model, thus of length 2.
- `revenues` A vector of differenced logarithm of revenues required to calculate outcome of the VAR model, thus of length 2.
- `expenditures` A vector of differenced logarithms of expenditures required to calculate outcome of the VAR model, thus of length 2.
- `deficit` Deficit calculated as difference between revenues and expenditures.
- `debt` Debt calculated as accumulated deficit.

A step in the algorithm for the balanced fiscal rule contains only two parts. It is assumed that after period t all variables are known. Initial values for `deficit`, `debt` and so on can be set for a particular simulation.

The first part of the step is application of the VAR model. This results in values of `gdp`, `revenues` and `expenditures` being overwritten with new values resulting from the VAR model. More precisely, these vectors are modifed with the oldest values dropped and new values appended at the end.

The second part of the step is again related to a fiscal rule but beyond that it just takes care of rolling appropriate vectors and appending new values. In case of the balanced fiscal rule the only part related to the “fiscal rule” is overwriting expenditures with the value of revenues.

1. The value of expenditures is overwritten with the value of revenues.

2. A new value of `deficit` is calculated as assumed expenditures, resulting from the fiscal rule, and revenues resulting from the VAR model, thus in this case always 0. This is, however, calculated as an additional check for correctness.

3. A new value of `debt` is calculated by adding `deficit`.

This concludes all calculations done within a single step of a simulation for the balanced fiscal rule.

3.8 Appendix B: modified HP filter

Given a time series $\{y_t\}$ for $t = 1, \dots, T$, a smoothed (filtered) version of the time series, with the use of the standard HP filter, is defined as $\{y_t^*\}$ for $t = 1, \dots, T$, where values y_t^* are defined as minimizers of the following function:

$$C_{std} = \frac{1}{\lambda} \sum_{t=1}^{T} (y_t - y_t^*)^2 + \sum_{t=2}^{T-1} \left((y_{t+1}^* - y_t^*) - (y_t^* - y_{t-1}^*) \right)^2 .$$

The first part of this expression is an 'error' that is made when substituting the original values y_t with smoothed values y_t^*. The second part captures the 'smoothness' of the trend time series. The coefficient $1/\lambda$ balances the two parts.

The Swiss fiscal rule uses a modified version of the HP filter, described by Bruchez (2003). Trend values y_t^* are defined as minimizers for the following function:

$$C_{mod} = \sum_{t=1}^{T} \frac{1}{\lambda_t} (y_t - y_t^*)^2 + \sum_{t=2}^{T-1} \left((y_{t+1}^* - y_t^*) - (y_t^* - y_{t-1}^*) \right)^2 ,$$

where

$$\lambda_t = \begin{cases} 3\lambda & \text{for } t = 1 \text{ and } t = T, \\ 3/2\ \lambda & \text{for } t = 2 \text{ and } t = T-1, \\ \lambda & \text{for other } t. \end{cases}$$

The above modification, which is effectively confined to applying different weights to observations, defines larger values of λ at the boundaries of a sample that leads to a trend part being more linear there. The modification of the filter is introduced in order to increase smoothness of the trend at the end of a sample, which is, in the context of the Swiss fiscal rule, a 24-observation rolling window of GDP values.

3.9 Appendix C: details of the VAR model

This section presents the VAR model results, lag selection and diagnostics. Coefficients of the estimated VAR model on output ('gdp'), government expenditures ('exp') and government revenues ('rev') with two lags are presented in Table 3.3.

Four information criteria were calculated and an optimal number of lags according to each information criterion used is given in Table 3.4. A lag of order 2 was selected in order to keep the model parsimonious. As a result we operate with 46 degrees of freedom for each equation. This choice is identical as in the model by Landon and Smith (2017).

Table 3.3: *VAR model results*

	Dependent variable:		
	gdp	rev	exp
gdp.lag1	1.229***	1.565***	−0.236
	(0.109)	(0.330)	(0.288)
rev.lag1	−0.071*	0.862***	0.027
	(0.042)	(0.127)	(0.111)
exp.lag1	−0.016	0.040	0.959***
	(0.058)	(0.176)	(0.154)
gdp.lag2	−0.257**	−1.354***	0.420
	(0.117)	(0.355)	(0.311)
rev.lag2	0.095**	−0.140	−0.038
	(0.041)	(0.125)	(0.109)
exp.lag2	0.008	0.030	−0.140
	(0.054)	(0.164)	(0.143)
constant	0.163	−0.420	−0.210
	(0.098)	(0.296)	(0.259)
crisis	−0.040***	−0.056***	0.024
	(0.006)	(0.017)	(0.015)
Observations	54	54	54
Adjusted R^2	0.999	0.993	0.995
F Statistic (df = 7; 46)	9,541.442***	1,055.989***	1,635.778***
Note:			*p<0.1; **p<0.05; ***p<0.01

A series of diagnostics tests has been run. They include calculation of roots and tests of normality, homoscedasticity and serial correlation of error terms. The results of

Table 3.4: *Best lag selections*

Citerion	Number of lags
Akaike	9
Hannan-Quinn	2
Schwarz	2
Final Prediction Error	3

diagnostic checks are presented in Table 3.5. The only potentially worrisome feature is a heteroscedasticity of the error term but the problem does not seem to be of a large magnitude and, in this particular application of the model, does not distort the results.

Table 3.5: *VAR diagnostic checks*

Roots					
0.986	0.791	0.791	0.469	0.469	0.157

Normality			
JB-Test (multivariate)	χ^2-statistic = 7.711,	df = 6,	p-value = 0.260
Skewness only (multivariate)	χ^2-statistic = 5.313,	df = 3,	p-value = 0.150
Kurtosis only (multivariate)	χ^2-statistic = 2.398,	df = 3,	p-value = 0.494

Homoscedasticity			
ARCH Breusch-Pagan LM (multivariate)	χ^2-statistic = 213.92,	df = 180,	p-value = 0.043

Serial autocorrelation (maximum of 5 lags)			
Portmanteau Test (asymptotic)	χ^2-statistic = 37.252,	df = 126,	p-value = 0.090
Breusch-Godfrey LM test	χ^2-statistic = 49.017,	df = 45,	p-value = 0.315
Edgerton-Shukur F test	F-statistic = 0.869,	df_1 = 45, df_2 = 86,	p-value = 0.694

4 Bid Rejection and Litigation in Public Procurement: Evidence from Poland

This chapter has been prepared together with Gyula Seres from the Humboldt University in Berlin.

Abstract:
Procuring authorities often use screening in order to reject risky offers. This study estimates the effect of bid screening and litigation on entry in public procurement using a unique data set on highway construction procurement auctions in Poland. The market exhibits a number of features of interest including a pronounced presence of foreign bidders and a strict bid selection procedure. We demonstrate with an empirical model that applying a screening procedure affects small firms disproportionately and creates a barrier to entry. We find that post-bidding litigation enhances the effect.

JEL Codes: H57, D44, D47
Keywords: procurement, auctions, market design, litigation

4.1 Introduction

Public procurement represents, on average, nearly 13% of GDP and 29% of government expenditure in OECD member countries OECD (2016b). The benefits of using auctions to purchase goods and services are widely recognized. As most public contracts are awarded by highly regulated procedures, their institutional framework may have a profound effect on competition (see eg. Hortaçsu and McAdams (2018)).

This paper contributes to our understanding of these rules by studying the effect of screening on bidding and market entry. We focus on bidding behavior and the effect of the legal framework on highway construction in Poland. In the context of procurement auctions, screening refers to an administrative selection of eligible offers. We develop a stylized model, where bidders form rational expectations about screening outcomes. Hence, the expected probability of rejection is incorporated into the bidding and entry strategy. Our estimates underscore that the procedural rules indicating acceptance or rejection of a tender can have a significant distorting effect on entry and competition. Using a unique data set we conclude that a level playing field in screening adversely affects smaller bidders and hampers their entry.

Market entry in the construction sector is curbed by a number of known factors and that entrants are particularly sensitive to incentives. The reason behind this is the winner-takes-it-all characteristic of most large-stake auctions. In order to foster entry, the auctioneer must create a system with a significant chance of winning for each entrant. It is a well-supported argument regarding dynamic competition with endogenous entry that favoring weak players fosters competition in the long term. The key idea is that, under symmetry, efficient bidders have a high probability of winning. Hence, incentives for entry of weaker firms are limited by a low chance of winning.

To mitigate the entry effect and improve long-run efficiency, governments can enact preference programs favoring small- and medium-size enterprises (see Krasnokutskaya and Seim (2011)). This paper addresses asymmetries from a new angle and argues that the legal system plays a crucial role in shaping competition and market entry. Public procurement differs from other bidding markets in that bids are accompanied by a mandatory complex documentation of the provided good. Given these features, an automatic evaluation of offers is not possible. Even if a contract is awarded to the bidder offering the lowest price, a flawed documentation can result in a rejection of the tender. Rejection rates are determined by the ability of bidders to deliver high-quality documents. As legal remedies are available, rejection can be mitigated by the ability to undertake litigation against opponents. If firms with larger construction capacities have enhanced ability for administrative and legal tasks, firms with capacity constraints suffer from disproportionately high barriers to entry, which ultimately results in higher prices.

To address this issue we analyze highway construction procurement auctions in Poland. Although emerging markets have attracted less attention in the empirical procurement literature so far, this market exhibits a number of characteristics which aid our study. After the accession to the European Union in 2004, the country experienced a large inflow of highly capitalized foreign firms, which actively participated in the ensuring construction boom. This setting provides a clear identification of firms from outside of Poland, which have accumulated substantial experience in competitive bidding. One can reasonably expect that large firms have a decisive advantage in delivering high-quality documentation and have a higher chance of having their tenders accepted.

Our data is based on highway construction contracts awarded through competitive procedures between 2005 and 2017. The available data, provided by the Polish highway procurement authority GDDKiA, allows us to identify the effect of legal motions and screening on bidding strategies and entry decisions.

Large-stake public procurement auctions draw significant attention in the literature. A fundamental theme of studies is dynamic modeling of competition with endogenous entry. The main idea behind studying entry is the dichotomy of rewarding efficiency and encouraging competition. Allocative efficiency is of key importance in auctions, since production is *a priori* determined by the auctioneer. As we know from seminal studies on auction theory, welfare-maximizing efforts are fostered by transparency and price discovery. However, helping efficient bidders means that the designer diminishes incentives of weak bidders to enter. Consequently, bidders face smaller competitive pressure to submit lower prices. This trade-off makes maintaining the chances of winning by weak bidders essential and even asymmetric auction rules are beneficial under some assumptions.

The most related scholarly contributions to our study concern the symmetry and the corresponding legal framework. As we argue, a level playing field discourages entry and may have a negative overall effect on a bidding market. The general idea in auction theory can be attributed to Myerson (1981). Li and Zheng (2009) estimate that the effect of bidder asymmetries is ambiguous in public procurement markets. Athey, Coey, and Levin (2013) suggest that subsidizing small bidders increases revenue and small bidder profit, with little efficiency cost. Bid screening, which is predominantly featured in this article, is studied by Decarolis (2014) and Branzoli and Decarolis (2015) in the context of performance. They conclude that screening prevents performance worsening but reduces cost savings and induces delays in the awarding procedure. Coviello, Moretti, Spagnolo, and Valbonesi (2017) emphasize the role of efficient legal procedures and show that courts' performance result in extensive delays in the procurement procedure.

Jofre-Bonet and Pesendorfer (2003) show that short-run capacity constraints have a profound effect on entry and bidding strategy and propose to control for the backlog of firms. While this approach is highly relevant in the primary market of bidders, its relevance, as we show, is limited if the major players are foreign as their backlog may not be the bottleneck of their ability to undertake new projects. Another strongly related line of research examines subcontracting and its effect on competition. The corresponding literature identifies two distinct effects of horizontal subcontracting. It weakens competition (Spiegel (1993)) and increases productive efficiency (Marion (2015)). As Gil and Marion (2013) demonstrate with empirical evidence, contractual relationships between firms have a positive effect on market entry. Jeziorski and Krasnokutskaya (2016) analyze the effect of restrictions on subcontracting and conclude that it has a positive overall welfare effect.

We also make a contribution to the study of the effect of auction format on bidding behavior. Since it is not feasible to cover all major questions of this area in this introduction, we focus on the differences between the two main formats favored in construction procurement, scoring and lowest-price auctions. Asker and Cantillon (2008) and Asker and Cantillon (2010) build a theoretical model and conclude that scoring auctions dominate price-only auctions under certain assumptions.

The number of other policy-related factors affecting bidding and entry is large. Other seminal papers address the role of publicity requirements (Coviello and Mariniello (2014)), information disclosure by the procurer (De Silva, Dunne, Kankanamge, and Kosmopoulou (2008)), preparation costs (De Silva, Jeitschko, and Kosmopoulou (2009)), subcontracting terms (Miller (2014)), the possibility of renegotiation of contracting terms (Bajari, Houghton, and Tadelis (2014)) and penalties (Lewis and Bajari (2014)).

This paper is structured as follows. Sections 4.2 describes relevant details of the Polish highway procurement market and Section 4.3 gives a descriptive analysis of the dataset we use. Sections 4.4 and 4.5 present main results. The paper concludes with a discussion.

4.2 Highway procurement in Poland

The state of the Polish infrastructure before joining the EU in 2004 was characterized by a deteriorated state of existing roads, lack of highways and absence of beltways bypassing population centers. Government objectives and newly available EU funds resulted in a large-scale construction boom in the number of newly constructed and renovated roads.[1] Our focus is on brand new or significantly upgraded roads that are administered by the General Directorate of National Roads and Highways (GDDKiA), a government agency responsible for building and maintaining roads that are classified as national roads.[2] Tenders in our dataset come from the period 2005–2017.

National roads in Poland are classified as A-class roads, S-class roads or G/GP roads. Road types are distinguished by their technical characteristics like the minimum distance between junctions, maximum allowed elevation, lane width, maximum bearing capacity etc. National roads are about 4.7% of all roads in Poland, when weighted by their length, but they comprise all A and S class highways in the country.

Construction procurement auctions normally follow a one-stage or a two-stage format based on Polish Public Procurement Law.[3] In a one-stage auction, the procurer specifies all details about a given auction and then asks all interested bidders to submit their offers. Firms may bid individually or form a consortium. In the latter case it is required to name a lead partner from the list of consortium members.

Auctions follow different procedures in case of the two-stage format. All firms that are potentially interested in building a given road are assessed in the first stage if they meet financial requirements specified for second-stage participants. These requirements include having enough own liquid resources and having open credit lines. Moreover, firms must prove that the have sufficient technical and human resources to complete the project. If firms decide to form a consortium, its members and leader must be specified before entering the first-stage. Polish Public Procurement Law mandates that the number of bidders invited to the second round must lie between 5 and 20. If there are less than 5 bidders the auction is repeated, and if there are more than 20 bidders fulfilling financial criteria, the most experienced ones are selected.[4] The second stage of an auction admits the preselected potential bidders. Between the two stages a detailed specification of the

[1] For details on infrastructure developments in Poland, see Goujard (2016). The scale of the projects is illustrated by maps in Figure 4.2 in Appendix 4.7.

[2] The agency name in Polish is 'Generalna Dyrekcja Dróg Krajowych i Autostrad', abbreviated as GDDKiA. We do not analyze renovation works and small upgrades but focus on works specified in various editions of 'National Program of Road Construction' ('Krajowy Program Budowy Dróg' in Polish. The document had editions specified for years 2008-2012, 2011-2015 and 2014-2023), which is an official document issued by the government according to which new national roads are constructed. We do not address projects undertaken by local governments because of data availability.

[3] The Law has been passed in 2004 and modified following the ratification of EU Directives 2004/18 and 2014/24. We exclude 3 directly awarded contracts from our sample.

[4] Unfortunately, the data regarding the first stage results and participation of an auction is not available. To the best of our knowledge, in the whole sample period no auction featured less than 5 first-stage participants.

road is published if it was not disclosed before the first stage. In the second stage all accepted firms make a decision if they submit a tender with a corresponding bid.

The procurer never discloses information about the final list of bidders. Both in one- and in two-stage auctions, firms are allowed to ask questions about details of road specifications and their documentation. These inquiries are publicly available and, therefore, firms can infer from them who can be interested in submitting bids. Moreover, in two-stage auctions the list of preselected companies which passed the first stage is publicly available before submitting bids in the second stage of an auction. This implies that firms may be aware of their potential competitors when they formulate their offers.

In both types of auction formats, bids can be rejected if they contain a mistake in the cost estimation, or they do not meet the investment specifications. Moreover, abnormally low tenders may also get rejected. Finally, bidders can be rejected in the first phase of a two-stage auction if they fail to prove their construction capabilities. There are no preferential programs for domestic nor small and medium-sized enterprises. GDDKiA does not take into account any criteria that are not directly pertaining to the offer.

A contract can get dissolved if a firm does not keep up with the specified progress in the construction. Objective difficulties like disadvantageous weather conditions may be a sufficient excuse to explain a delay in the contract. After dissolving a contract, another auction is published to resume construction works.

The awarding criteria are specified by the procurer. The awarding method is either lowest-price rule, essentially a first-price sealed bid auction, or a scoring rule.[5]. The latter encompasses price, quality and completion time being the awarding criteria. Weights for these dimensions of bids vary between auctions. The price determines 70-95% of the score with 90% being the weight in the vast majority of cases. Time may be between 5 and 30% of the total score, while quality may be between 5 and 15%. Time to completion stands for a maximum number of months within which a company pleads to finish the contract and quality is the minimum number of years within which a bidder guarantees to repair the road at his own expense. The procurer specifies the limits for both best potential quality and time declarations.

The first stage of a two-stage auction may start when only a fraction of all necessary documents and decisions are ready, giving more time for GDDKiA to prepare the lacking rest. It implies that participants of the 1st stage of an auction know only vague details

[5]This awarding criterion appears under other labels in legal documents Dimitri, Piga, and Spagnolo (2006) EU Directive 2004/17/EC and 2004/18/EC referred to it as *best price-quality ratio*, whereas the current Directive 2014/24/EU uses the term *most economically advantageous tender* (MEAT).

of the technical specifications of a road that is to be built. The awarding criteria have to be published before the first stage.

Bidders compete for construction contracts that are of one of four types: ‘build’, ‘project & build’ (P&B), ‘reconstruction’ or ‘continuation’ works. The first two types can be granted for a construction of new roads and in the ‘project & build’ contract a construction company is responsible not only for building a road but also for preparing a detailed road project on its own. In the ‘build’ contract type a company builds a road according to a project specified by another company.[6] ‘Reconstruction’ and ‘continuation’ contracts are granted for upgrades of existing roads and continuation of works when a previous contract was dissolved, respectively.

Auction participants can appeal for reversion of decisions to the National Chamber of Appeal (KIO),[7] a court specializing in decisions of contracting public authorities. Appeals against KIO decisions are possible in regular courts, but such motions are rare due to excessive legal costs.[8] When the auction procedure is lawfully finished, selected contractors can protest against GDDKiA decisions in regular courts only. Appeals to KIO can pertain, among others, firms’ protests against rejections of their own bids, claims that a single or multiple other bidders should be rejected or collective actions of many firms against certain aspects of a given procurement procedure.

4.3 Data

Our data is compiled from two distinct datasets provided by the procurement authority GDDKiA, described in Subsection 4.3.1, and the court of appeals KIO, described in Subsection 4.3.2.

[6] A company preparing a project is chosen in another independent public procedure. GDDKiA decided to introduce ‘P&B’ contracts in order to smooth the construction process by reducing potential inconsistencies between projects and their execution.

[7] KIO or Krajowa Izba Odwoławcza in Polish.

[8] Appeal fees paid to KIO are about 2000-3000 EUR, final costs are usually not higher than 5000-6000 EUR. Costs of appeals in regular courts can be as high as millions of EUR. The aforementioned expenses comprise only filing fees and court costs and do not include other legal bills.

4.3.1 Auctions and bids

Our data contains 300 auctions run between January 2005 and December 2017.[9] All procurements concern a major road project coordinated by GDDKiA. The data includes 186 lowest-price and 114 scoring auctions. Table 4.1 presents summary statistics separately for the two awarding criteria. In what follows, similarly as in the whole article, we express all bids in Polish złoty (PLN) in 2016 prices.[10] The total contract value in the sample period is slightly over 30 billion EUR, which corresponds to about 7.08% of nominal Polish GDP in 2016. The descriptive statistics reveal a number of interesting patterns. On average, bids are lower in scoring auctions as they amount to 445 million PLN compared to 497 million PLN in the other format. As the length of road segments is comparable, amounting to 13.69 km on average, this drastic difference can be attributed to the different proportion of road classes. While lowest-price auctions exhibit all three categories in similar numbers, including 44 A-class contracts, scoring auctions are applied mainly in auctions for S-class roads as 89 out of 114 contracts awarded by scoring is classified as S. The average number of bidders is similar for both auction types, with 7.92 bidders in the lowest price and 8.04 in the alternative format.

Figure 4.1 displays the number of the scoring and lowest-price for each year, showing all scoring auctions in a single set. The figure shows also the distribution of one and two-stage auctions over time as well as the distribution of contract types across the years.

The data highlights two important features. Firstly, there is a gradual policy shift from first-price to scoring formats, the latter becoming dominant in 2013. In the two last years of the observed period the new mechanism was used solely. Secondly, the timing of auctions is bimodal, with most of them taking place in the periods 2008-2009 and 2013-2014. This pattern might be related to parliamentary election years (2011 and 2015) as they witnessed a smaller number of observations. An alternative explanation is that the 2012 UEFA European Championship in Poland and Ukraine induced a larger number of new projects in the preceding years. The most plausible explanation is provided by the EU budgets that are prepared in seven-year horizons, the first being in years 2007–2013 and the other one being in years 2014—2020. Considering the fact that European funds constitute a significant fraction of overall funds spent on road investment, EU budget periods may play an important role. To the best of our knowledge, there are no formal sources with statements on which of these reasons is most plausible.

[9] One may argue that data from that period may be distorted by lower markups and falling backlog due to the Great Recession, as evidenced by Gugler, Weichselbaumer, and Zulehner (2015). However, Poland enjoyed a robust and stable growth during the entire period.

[10] 1 PLN corresponded in 2016 to about 0.21 USD or 0.23 EUR.

Table 4.1: *Summary statistics of data on bids and auctions*

	Awarding mechanism		
	Lowest Price	Scoring	All
Auctions	186	114	300
One stage	71	5	76
Two stages	115	109	224
Bids	1,473	918	2391
Accepted bids	1,362	804	2166
Rejected bids	111	114	225
Mean bid (in mln PLN)	497.2	445.3	477.6
Mean price (in mln PLN)	428.9	421.3	426.0
Mean # of bidders	7.92	8.04	7.97
Mean # of minor cons. members	0.84	0.28	0.62
Mean length (in km)	13.53	13.96	13.69
A class roads	44	10	54
S class roads	85	89	174
G/GP class roads	57	15	72
Dissolved contracts	14	1	15

It is important to note that in the scoring auctions, no matter what weights were applied on quality and completion time, there is no variation in non-price components of bids. In all instances firms offered best permitted quality and time, leaving the competition solely in price.

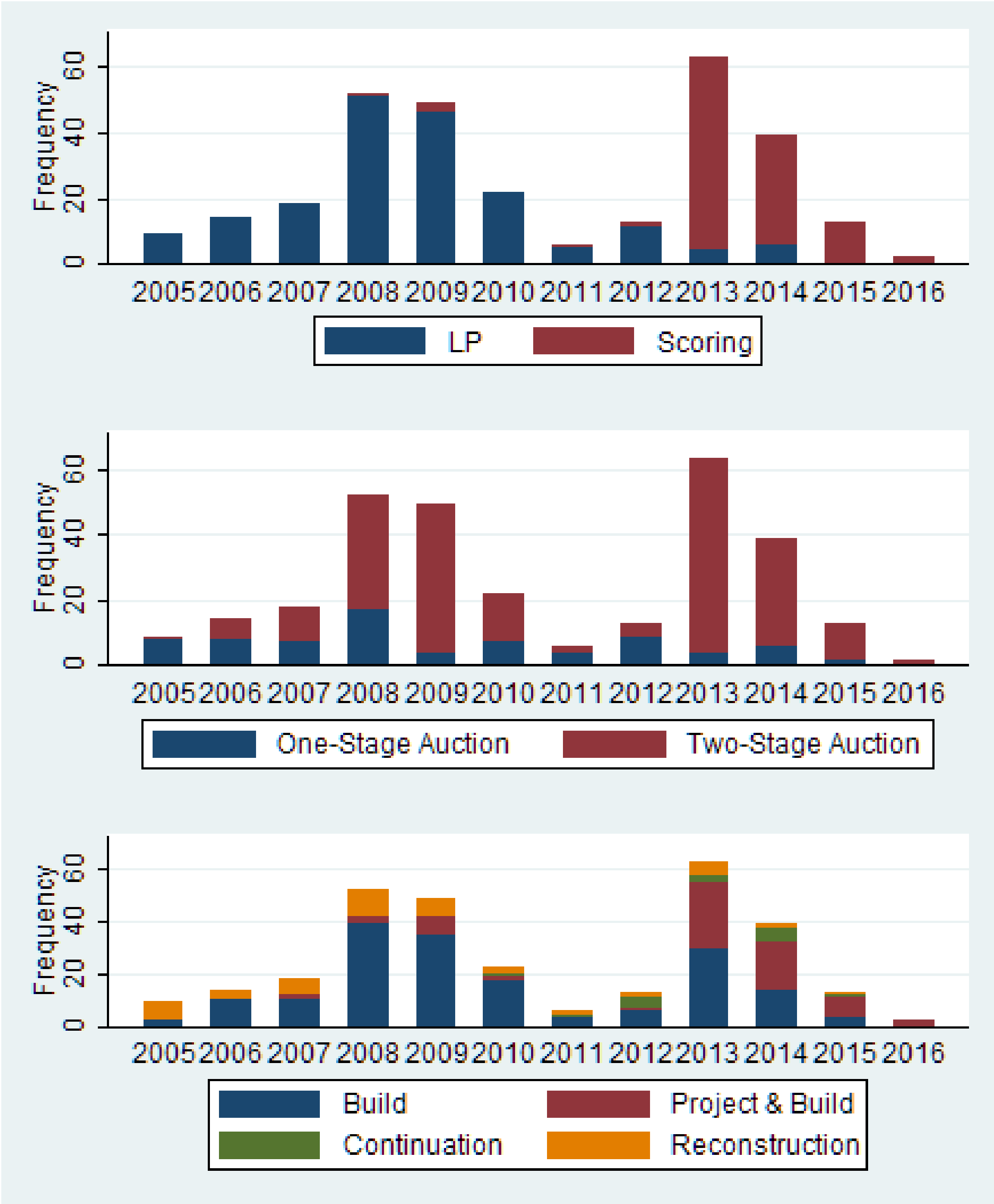
4.3.2 Firms and litigation

The number of foreign main contractors is constantly high. For the entire sample 75.65% of all main contractors have a foreign headquarter or are owned by a foreign construction company.[11] This share does not exhibit a significant time trend.

Foreign firms are more successful and win 80.6 % of all contracts, which is higher than their overall representation. We can note a similar pattern in the value of contracts, whose mean amounts to 365 million for domestic and to 457 million for foreign main contractors. These differences result in a remarkable disparity in the total value of contracts. Foreign companies win a grand total of 110.2 billion PLN out of 131.4 billion, corresponding to a 83.9% aggregate market share.

[11] We classify a firm foreign if it is owned in at least 51 % by foreign capital. Firms may change their status within the years encompassed in the dataset.

Figure 4.1: *Number of auctions across years by auction or contract type.*



A focus of this paper is to assess the effect of performance of large and small firms in screening and litigation on bidding strategies. Therefore, we divide the sample of all firms into two groups. The first group consists of 20 most frequent bidders, whose list is given in a table in Appendix 4.10. Frequency of bids is measured by the number of bids submitted as an individual firm or as a main consortium member. Firms are measured together with their subsidiaries, if such exist. The other group of firms is described as 'small', which is rather a description of their significance on the Polish road construction market, than their overall size. In total, large firms submitted more than 70% of all bids and won nearly 76% of all contracts with the market share of nearly 78%.[12] On the other hand, considering all auctions, 124 small main contracting firms submitted 712 bids. Among these, 331 were submitted by domestic and 381 by foreign firms.

The KIO database reveals that out of the 300 auctions, 81 had an appeal of some sort. Technically, the plaintiffs in each case is a varying number of firms, or their consortia, and the defendant is the procurement authority (GDDKiA). If a decision of the procurer is affirmed, by rejecting all the appeals against it, decisions on the choice of best offers do not change. On the other hand, if firms' appeals are accepted, the GDDKiA decisions are reversed and the procurer is obligated to reassess the competition, taking into account all lawful offers.

For these reasons, we distinguish two types of appeals. Type I ('a defensive appeal') refers to an appeal that aims to change the GDDKiA decision by reversing the rejection of the plaintiffs' offers or to a complaint about a faulty auction procedure in general. Type II ('an offensive appeal') contains a claim, that another bidder's offer contains an error, and therefore, it should be rejected. From all auctions, 41 were followed by type I, and 54 were followed by type II appeals. All auctions considered, 33 of them had more than one appeal and 16 of them exhibited both types. The success rate of appeals was 37.25% as 57 GDDKiA decisions have been reversed out of 153 appeals.

4.4 Analysis of bidding

In what follows, we build up a theoretical model that serves as an exposition of firms' optimal decisions. Next, we analyze bidding empirically using OLS models. These are in turn extended by the use of instrumental variables, in order to mitigate effects of potential endogeneity.

[12] This market concentration is high compared to the Californian highway construction market data with a 73.4% market share of the 20 largest firms, as evidenced by Bajari et al. (2014).

4.4.1 Theoretical model

We follow the footsteps of Levin and Smith (1994) and use a dynamic auction model. A procurer announces a first-price sealed-bid auction for a single contract to n *ex ante* symmetric risk-neutral potential bidders, who may enter the competition. Then, after receiving signals about their types, which are described in detail in the next paragraph, all n bidders make decisions on entry. The number of entering firms is publicly disclosed before making bids and is denoted with k. Bidders who decide to enter the competition simultaneously make bids. The submitter of the lowest offer wins and pays an amount equal to its bid. This framework models the two-stage scheme, which is most often used in GDDKiA auctions.

Each bidder, before making the decision on entry, receives an independently drawn signal c_i about its costs. The signal is drawn from a distribution described by a twice continuously differentiable cumulative density function $F_c(c)$ with a finite support $[\underline{c}, \bar{c}]$.

Unlike Samuelson (1985), we allow for a non-uniform entry costs e_i, which capture a firm's ability to undertake new projects. They include administrative fees and offer preparation costs. Intuitively, the value is heterogeneous among firms due to their size as large firms enjoy smaller entry costs because of economies of scale. The costs are drawn from a distribution described by its twice continuously differentiable cumulative density function $F_e(e)$ with a finite support $[\underline{e}, \bar{e}]$.

We assume that the construction cost c_i and entry cost e_i are independently drawn across all bidders and firm i is fully characterized by its type (c_i, e_i). The cumulative distribution function $F_{c,e}(c, e)$ is assumed to be twice continuously differentiable.

In the auction framework explained above, bidder i maximizes its expected profit in the bidding phase:

$$\underset{b_i}{max} \ \mathbb{E}\left[\pi(b_i, b_{-i}, c_i)\right] = (b_i - c_i) \cdot Pr\{b_i = b_1\}, \tag{4.1}$$

where b_i is the bidder i's bid, b_{-i} is the set of all other bids, c_i is the bidder i's cost signal and b_1 denotes the lowest bid. Firm i enters the auction if the corresponding expected profit is larger than the entry costs, i.e., $max_{b_i} \ \mathbb{E}\left[\pi(b_i, b_{-i}, c_i)\right] \geq e_i$. In case of no participation a bidder receives its outside option that is normalized to 0.

Proposition 1. *For any c_i, the equilibrium bid of a firm $b^*(c_i)$ is a decreasing function of the number of entering bidders k and of the number of potential entrants n.*

Proof. See Appendix 4.11.

□

4.4.2 Bidding policy: OLS results

As Proposition 1 yields a prediction on bidding strategy and it implies that entry puts competitive pressure on bids, we estimate a corresponding empirical bidding model using OLS.

Following Porter and Zona (1993), we calculate backlog as a measure of bidders' used capacities. Then, it serves as a proxy for whether the firm is facing capacity constraints. It is constructed as follows. The variable takes all ongoing projects of a bidder at the point of bidding in a given auction, assuming the capacities are used in a uniform fashion. Backlog is constructed in each day of the sample time period by summing this per day value of all running projects of a given firm. The relevant period runs from signing a contract to the date of opening of a road. As projects are completed, the backlog of a firm goes to zero, unless new contracts are won.

For all finished contracts in our database, both the starting time of the contract and its completion date are available. Given that at the time of observation several roads in our sample are still under construction, we estimate the backlog of unfinished projects using imputed road completion dates.[13] We perform an OLS estimation of the construction duration on the roads that have been completed including all road characteristics and regional dummies to estimate the expected completion time of unfinished roads. We include the logarithm of backlog $ln(backlog + 1)$ in all our regressions.

One of the main points of interest is the analysis of competition on bids. With respect to the competition effects we have to address the heterogeneity in the number of auction stages. One-stage auctions lack an official feedback on the number of eligible bidders. It is doubtful, however, that there is no information leakage in the first-stage auction at all. We can presume then that there is some interim information about participation before bid submission. We test for a heterogeneity in the information structure between auction formats. As the two formats are different in terms of revealed information about participation, we include an interaction term in the model in order to measure potential differences. We estimate:

[13] Jofre-Bonet and Pesendorfer (2003) use planned completion dates for contracts that did not finish by the end of the sample period. However, this data is not available in our sample.

$$Log\ of\ Bid = \beta_0 + \beta_1 \cdot Number\ of\ Bidders + \beta_2 \cdot Number\ of\ Bidders \times OS + \beta \cdot X + \epsilon, \tag{4.2}$$

where X is the vector of all other covariates including contract type, road characteristics and regional dummies and 'OS' stands for 'one stage' binary indicator. Testing statistical significance of β_2 implies a two-sided test of the stage number effect. That is, our model allows us to test if first-stage auctions create a different information setting for bidders. The reason behind this is that potential information leakages about participation, connected with publicly available inquiries submitted to the procurer or other informal channels, may differ. Lastly, in the two-stage format bidders know the identity of their potential competitors.

Table 4.2 lists OLS regression estimates of the model specified in Equation 4.2. The coefficients on the number of bidders in both models are negative, suggesting that higher competition (i.e. more bidders) reduces bids but they are not significant statistically. However, the number of bidders in a one-stage auction exercises a competitive pressure on bids, and the effect is significant at the 0.1% level. Thus, bidders seem to receive some signals about entry before submitting bids.

Our model also suggests that using scoring as the awarding mechanism has no significant effect on bids. The dummy variable *Scoring* concerns the awarding mechanism and it is equal to 1 for scoring auctions. The lack of significance is likely related to the quality requirements we described above. While quality standards are scored, a cap is applied to their values, which is easy to reach and there is no within-auction variance in the quality score as all firms submit the best possible quality and/or completion times. Hence, bidders, in fact, compete in prices.

The sign of the estimated coefficient for backlog is positive as expected, which means that firms with higher load are willing to take on new projects for a higher price. However, we find no conclusive evidence on a significant effect on bids. A potential underlying reason behind a low estimate is that the biggest foreign firms have a small foothold in the local market compared to their total size. Strabag's total revenue in 2016 was 12 billion EUR (54 billion PLN), which exceeds its maximum backlog of 1.01 billion EUR (4.38 billion PLN) significantly.[14] The other major competitors' exposure is similar: Ferrovial's total revenue was 10.8 billion EUR (48 billion PLN) in the same calendar year, comparably higher than its maximal maximum backlog of 6.46 billion PLN.[15] On the other hand,

[14] www.strabag.com

[15] www.ferrovial.com

the backlog of many small firms is mostly zero. Lastly, our measure of backlog is an imperfect proxy for the overall capacity constraints of firms as our sample contains only data on national highway contracts. Most, if not all, companies are involved also in other construction projects, e.g. local roads, railways or apartments, which constitute a substantial portion of their domestic portfolio.

Table 4.2: *OLS and 2SLS estimates on logarithms of bids.*

	OLS		2SLS			
	(1)	(2)	(3)		(4)	
	Log of Bid	Log of Bid	Log of Bid		Log of Bid	
Log of # Bidders	-0.14	-0.13	-0.82**		-0.97**	
	(0.10)	(0.098)	(0.27)		(0.30)	
Log of # Bidders × OS	-0.17***	-0.18***	-0.17***		-0.16**	
	(0.041)	(0.043)	(0.047)		(0.050)	
Log of Backlog	0.0091*	0.0058	0.011*		0.0083*	
	(0.0044)	(0.0040)	(0.0046)		(0.0041)	
Scoring	-0.038	-0.037	-0.026		0.026	
	(0.078)	(0.078)	(0.084)		(0.092)	
Reconstruction	-0.59***	-0.58***	-0.69***		-0.71***	
	(0.12)	(0.10)	(0.14)		(0.13)	
Project & Build	-0.31***	-0.25**	-0.39***		-0.34***	
	(0.075)	(0.079)	(0.082)		(0.093)	
Continuation	-0.60**	-0.65***	-0.89***		-1.06***	
	(0.18)	(0.18)	(0.24)		(0.24)	
Foreign Firm	-0.0086	0.017	-0.050		-0.024	
	(0.026)	(0.024)	(0.030)		(0.026)	
Region Dummies	no	yes	no		yes	
First-stage instrument			(FS1)	(FS2)	(FS1)	(FS2)
IV			0.79***	-0.058***	0.79***	-0.063***
			(0.031)	(0.017)	(0.032)	(0.018)
IV × One Stage			0.0053	1.07***	0.0098	1.07***
			(0.0078)	(0.0044)	(0.0078)	(0.0045)
Specification tests						
Wu-Hausman			45.90***		58.36***	
Durbin-Wu-Hausman			89.05***		112.68***	
Observations	2167	2167	2164		2164	
Adjusted R^2	0.7724	0.8068	0.7415		0.7649	

Standard errors clustered by auction. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Monetary values are expressed in 1 million PLN (2016 prices). All models include road characteristics. Road characteristics: log(length), road class, interaction terms between log(length) and class, number of junctions, dummy variables for bridge and tunnel, indicators for roads in minor and major urban areas, indicator for concrete road, monthly unemployment rate and the log of the average number of warm days in the region.

4.4.3 Estimating bids with an instrumental variable

A concern with our identification strategy is that entry may be linked to unobserved auction characteristics. Such unobserved characteristics may be important factors for road construction, for example, ground stability in the road location or a required non-standard construction material or technique. This would create an endogeneity bias and distort our results. We try to address this concern by using an instrumental variable strategy. We identify a valid instrument that is correlated with the number of bidders in a given auction and uncorrelated with unobserved project characteristics.

A standard technique used in the literature is to instrument the number of bidders with the number of companies registered in the first stage of an auction.[16] Since firms may not know the project characteristics, this is considered a valid instrument, see De Silva et al. (2008). Unfortunately, this number is not universally available in our sample. As an alternative, we use the average number of bidders in the preceding auctions. We use the average number of bidders in auctions in the preceding 365 days (i.e. in one year) and exclude auctions on the same day in order to mitigate the problem of endogeneity due to potentially related contracts.[17]

Exogeneity of the instrument is not possible to verify directly, but seems plausible for the following reasons. It is unlikely that entry would be affected by any characteristics of previous auctions as, conditional on entry, the equilibrium strategy of a firm may only be influenced by its own and other firms' costs, which pertain strictly to a given auction. A potential concern regarding the exclusion restriction is that the instrument is correlated with an unobserved market-related variance. This is, however, unlikely because when firms submit their bids only current costs, driven by the current market situation, are taken into account. Even if the opposite were true, the time lag of the information used for the instrument is rather limited. It is improbable that fundamental changes occur in the market setting within one year.

The IV model is estimated using two-stage least squares with clustered standard errors, using the first-stage regression model with interaction terms:

$$Number\ of\ Bidders = \pi_0 + \pi_1 \cdot IV + \pi_2 \cdot IV \times OS + \pi_3 \cdot X + \epsilon, \tag{FS1}$$

$$Number\ of\ Bidders \times OS = \pi_0 + \pi_1 \cdot IV + \pi_2 \cdot IV \times OS + \pi_3 \cdot X + \epsilon, \tag{FS2}$$

[16]The cited literature refers to them as plan holders.

[17]The procurer divides some roads into separate segments and announces auctions for each segment separately. Such auctions occur on the same day and may share road characteristics.

The proposed instruments turn out to be strong as shown by the significant coefficients in the first-stage equations in Table 4.2. The coefficients of the first-stage equations suggest that our instruments are strongly correlated with the number of bidders and the respective interaction term in the auction. This is consistent with our claim above that the market is stable and entry is not affected erratically by short-term market shocks.

The coefficient estimates of the log number of bidders are negative in all model specifications, meaning that higher competition translates into lower highway prices. However, Table 4.2 reveals that the estimates based on the instrumental variable approach predict much higher competitive pressure than OLS estimates. That is, the use of instrumental variables may be justified. This claim is confirmed by the Wu-Hausman F-test on both models as they yield values 45.90 and 58.36, both significant at the 0.1% level. The Durbin-Wu-Hausman test with statistics of 89.05 and 112.68, is significant also at the 0.1% level. As the null hypotheses, assuming no statistical differences between OLS and 2SLS results, are rejected, this suggests that the OLS estimates are not consistent. The estimates of the fully specified model of Column 4 have the following interpretation: an increase in the number of competing firms by 1% decreases the highway price by about 0.97% in a two-stage auction and by about 1.13% in a one-stage auction. In other words, considering the fact that the average number of bidders is equal to about eight firms, an increase in the number of competitors from eight to nine implies a price decrease of about 12.1% and 14.1% in a two-stage and in a one-stage auction, respectively.

Other aspects of the equations also deserve our attention. The number of stages of the auctions influences cost as two-stage auctions last longer, and thus, imply higher entry costs than one-stage auctions. However, implicit entry costs do not influence bids. Conditional on entry, it is considered to be a sunk cost and cancels out in the first-order conditions.[18] The opposite argument applies to other auction category variables. For instance, '*project & build*' and '*build*' contract types differ in the execution of the plan and involve different costs, conditional on winning. That is, variables that are related to the construction costs enter the bidding equation while the entry cost does not. We find no evidence that foreign and domestic firms bid differently.

[18]Other factors, potentially related to different information structures, led to a stronger negative effect of competition in a one-stage auction type.

4.5 Bid Screening and litigation

To formalize the effect of screening in auctions, we extend the simple model we introduced in Section 4.4, in the spirit of Krasnokutskaya and Seim (2011) and De Silva et al. (2008). We note that success in screening and litigation is strongly related as both indicate a firm's ability to avoid rejection. Thus, we handle these two factors together.

As the procurer exerts effort to screen bids, it is a legitimate question to address its effect on entry. Screening can result in bids' rejections and bidders anticipate this. The rejection rate may vary between firms of different size and country of origin. In this case, small firms' entry may be adversely affected, which reduces bidding and, therefore, firms' pressure to submit low bids. In order to address this puzzle further, we augment our study with litigation data. As appeals against screening decisions are possible, courts may alter the outcome even further. We proceed in two steps. First, we analyze the setting with a theoretical model and formulate our hypotheses. Second, we confront the hypotheses with the data and draw conclusions.

4.5.1 Theoretical model with screening

In this section we enrich the baseline model from Section 4.4.1 with screening. The extended model assumes that the procurer is able to install a screening system that may reject certain bids. For simplicity, we focus on the possible rejection of the most competitive bid only, as it is unlikely in practice that the two lowest bids are both simultaneously rejected in the same auction.

The probability of rejection, conditionally on submitting the lowest bid, is a function of entry cost e_i, expressed as a non-decreasing twice-continuously differentiable function $\alpha(e_i) \geq 0$. The intuition behind this is that firms that are able to carry more projects (i.e. the ones with low e_i) are also the ones with better legal and administrative teams, which are also responsible for preparing and defending offers.

Function $\alpha(e_i)$ captures the disadvantage of small firms and a steep slope of $\alpha(e_i)$ signals a more pronounced difference between firms of different size (or with different entry costs). On the other hand, a constant $\alpha(e_i)$ would mean that legal strength has no influence on screening and the subsequent litigation. The extreme case $\alpha(e_i) = 0$ implies a situation in which no screening takes place and all bids are accepted. This case corresponds to our baseline model. The objective function of the firm becomes, conditionally on entering:

$$\underset{b_i}{max}\ \mathbb{E}\left[\pi(b_i, b_{-i}, c_i, \alpha(e_i))\right] = \Big((1-\alpha(e_i)) \cdot Pr\{b_i = b_1\} + \alpha(e_1) \cdot Pr\{b_i = b_2\}\Big) \cdot (b_i - c_i), \tag{4.3}$$

where e_1 refers to the entry cost of the bidder with the lowest offer.[19] Equation (4.3) is different from (4.1) in two aspects. First, a firm may be rejected if it is ranked first. Second, a firm that is ranked second may win if the first-ranked bidder is rejected. These terms depend on the specification of the screening function $\alpha(\cdot)$, which can be freely set by the procurer, respecting the assumptions above. As in the baseline model, entry occurs if the expected payoff covers the entry cost, $max_{b_i}\ \mathbb{E}\left[\pi(b_i, b_{-i}, c_i, \alpha(e_i))\right] \geq e_i$.

Proposition 2. *For constant $\alpha(e_i) = \alpha$, the marginal effect of α on entry is negative for firms with sufficiently low costs c_i, and positive for firms with sufficiently high costs c_i.*

Proof. See Appendix 4.11.

□

Higher, but uniform for all firms, rejection rate may not favor more competitive firms as the lower probability of winning with the most competitive bid has a stronger effect on expected payoff than the higher probability of winning with submitting the second lowest bid. In terms of Proposition 2, a higher entry among firms with higher cost c_i can be achieved with a screening system that does not discriminate between firm types. Intuitively, if rejections occur more often for firms with higher barriers of entry (i.e. higher values of c_i or e_i), we can observe two effects. Given that differentiated rejection rates imply that favored firms may enjoy higher entry rate but this may be compensated by a larger drop in entry of firms that have higher entry cost e_i. The role of high overall entry is increased level of competition in the spirit of Proposition 1. Lower entry may, in turn, imply less competition and higher bids and, consequently, a higher *ex ante* expected price for the auctioneer. As the effect of the screening function is not uniform for firms with different levels of barriers to entry, it is yet to be tested what the net effects are.

4.5.2 Hypotheses

The hypotheses which are evaluated in our empirical study are general patterns that are implied or assumed by the theoretical model. The main assumption of the model is that the firm size positively affects the chance of a bid to be accepted.

[19] The rest of variables is defined as in the baseline model.

Hypothesis 1. *Small firms have higher rejection rates, $\alpha'(e_i) > 0$.*

Hypothesis 1 can be evaluated by comparing firms by size. As actual firm size is a poor predictor of available resources and it is correlated with entry, we use the total sum of entries of a firm as a proxy for e_i. We conduct a comparison between size categories, and we set different threshold levels to check robustness.

The second hypothesis concerns the strategic effects of screening and is a direct test of Proposition 2.

Hypothesis 2. *The rejection rate $\alpha(e_i)$ has a negative effect on entry.*

If Hypothesis 1 and 2 hold, screening has a negative effect on the entry of small firms. Our framework and data allow for a further analysis. In the rest of the Section we address the overall effect of the screening policy and estimate its effect on total entry and bidding.

4.5.3 Rejections: logit model

In this section we explore empirically the underlying reasons behind bid rejections and we test Hypothesis 1. In the sample, 225 offers have been rejected, which accounts for 9.41% of all bids. There is a sizable difference between the rejection rate of foreign (8.78%) and domestic firms (11.34%). The contrast is even more pronounced if we compare the 20 most frequent bidders (7.44%) with the small firms (14.04%). In what follows we explore this disparity and its consequences on entry.[20]

The available data does not provide us with the bids of rejected tenders, only with the identity of the bidders. However, it is reasonable to assume that Type II appeals (i.e. 'offensive appeals') are directed mainly against low bids. As way of testing this, we include competitiveness as a covariate in our model. The variable is constructed in the following way. A bid of a given company, equal to b_i, is divided by the winning bid in a given auction, which we refer to as b_1, creating a relative bid b_i/b_1. This value is 1 for a winning bid and larger than 1 for everyone else. Then, a mean of this variable is calculated for each firm across all auctions in which a company took part. The value of this firm-specific variable is always at least 1 (it would be equal to 1 if a firm would have won all auctions in

[20]Since no data is available on the first-stage registrations in two-stage auctions, building a model including first-stage rejections is not possible. Consequently, the estimated rejection probability should be interpreted as an estimate of a chance of being rejected after bidding. First-stage participation is considerably cheaper, hence, omitting the impact of rejection probabilities on first-stage decisions is unlikely to bias estimates.

which it participated) and higher values correspond to less competitive bids, on average. This measure of competitiveness is time-invariant.

If our prediction on strategic litigation is correct, small but efficient firms suffer higher rejection rates. Let us consider a binary choice model described by equation (4.4). Competitiveness ('*Comp.*') is the measure defined above. '*Small*' is a binary variable that assumes value 1 if a firm, or a consortium leader, is not among the largest firms in terms of the number of submitted bids.[21] The threshold delimiting *large* and *small* firms is set to have 15, 20, 25 or 30 most frequent bidders as *large* firms.

The logit model explaining rejections is estimated as follows:

$$Pr(\text{Reject}) = \frac{\exp(\beta_0 + \beta_1\text{Comp.} + \beta_2\text{Small} + \beta_3\text{Small} \times \text{ Comp. } + \beta_4 X + \varepsilon)}{1 + \exp(\beta_0 + \beta_1\text{Comp.} + \beta_2\text{Small} + \beta_3\text{Small} \times \text{ Comp. } + \beta_4 X + \varepsilon)} \tag{4.4}$$

Including competitiveness in the model is a sensible choice. Low bids attract more attention as these submissions may win the auction after screening. That is, competitors are more likely to challenge an acceptance decision and the bidder whose bid is rejected is more likely to file an appeal. Coefficient β_3 captures the difference of the effect of competitiveness on rejection probability. Estimate $\beta_1 + \beta_3$ gives us the overall effect of competitiveness for small firms and β_1 is an estimate of the same effect for large firms. In order to directly test Proposition 2, we test if β_3 is significantly lower than zero.

Logit models (1)-(4) in Table 4.3 use different definitions of small and large firms.[22] In all cases, 'large' corresponds to the ranking according to the number of submitted bids. The different thresholds are considered as robustness checks. The most pronounced results are obtained if we draw the threshold at the 20 or 25 largest firms.

Table 4.3 reveals a number of important patterns. The first is that small firms face a much higher rejection rate. As we can see, the effect of firm size is significant at 5% level in all specifications but one. This results is corroborated by the coefficient on capacity in model (5). Larger firms, when measured by their maximum capacities, face lower rejection rates, which is a highly significant result.

It appears that the effect of competitiveness on the occurrence of rejection depends on size. Frequent bidders are mildly favored if they submit low bids in general. The opposite

[21] The label '*large*' refers to a company's presence in the Polish market and not necessarily to its overall size. Most firms listed as '*small*' are smaller in terms of revenue, but there are exceptions, for example Obrascón Huarte Lain or Max Bögl.

[22] The number of observations is lower than the full sample by 10 units as some firms never submitted an accepted bid and, hence, their competitiveness measure cannot be obtained.

is true for small firms. More efficient small firms have a higher likelihood of being rejected than their less efficient counterparts. For example, in model (3), the marginal effect of competitiveness is $\beta_1 = 5.57$ for large and $\beta_1 + \beta_3 = -0.74$ for small firms. The same pattern holds for all models.

Our model also controls for the presence of appeals. The coefficient estimates have the predicted signs. Type I is in favor of the plaintiffs' bids, while Type II claims a rightful rejection of competitors. Hence, the estimates are negative and positive, respectively. Both are significant at the 5% level in all models. This is even more pronounced in case of Type II appeals with coefficient estimates that are significant at the 0.1% level.

Overall, our estimates support Hypothesis 1 as firm size has a sound effect on rejection. The effect of competitiveness is ambiguous. Large firms enjoy lower rejection rate if they submit low bids on average. However, small firms suffer the opposite conjecture.

Table 4.3: *Logit Model on Bid Rejection*

	(1) Rejected	(2) Rejected	(3) Rejected	(4) Rejected	(5) Rejected
Competitiveness	3.41	5.42*	5.57*	4.11*	0.19
	(3.10)	(2.75)	(2.43)	(1.70)	(1.12)
Small15	4.07				
	(3.85)				
Small15× Comp.	-2.85				
	(3.26)				
Small20		8.11*			
		(3.68)			
Small20× Comp.		-6.20*			
		(3.09)			
Small25			8.19*		
			(3.47)		
Small25× Comp.			-6.31*		
			(2.88)		
Small30				6.39*	
				(2.58)	
Small30× Comp.				-4.63*	
				(2.13)	
Log of Capacity					-0.14***
					(0.036)
# of Type I	-0.36*	-0.35*	-0.36*	-0.36*	-0.33*
	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)
# of Type II	0.52***	0.53***	0.53***	0.52***	0.52***
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
Foreign Firm	0.025	0.13	0.17	0.15	0.088
	(0.19)	(0.20)	(0.22)	(0.21)	(0.20)
Constant	-8.87*	-11.2***	-11.4***	-9.73***	-3.92*
	(3.74)	(3.36)	(3.00)	(2.18)	(1.63)
Year Dummies	Yes	Yes	Yes	Yes	Yes
Observations	2381	2381	2381	2381	2381

Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Monetary values are expressed in 1000 PLN (2016 prices).

4.5.4 Bidding and entry with beliefs about rejection

In this section we re-estimate our bidding model with the estimated rejection probabilities. First, we estimate the effect of screening policy and appeals on bidding strategy using fitted values from model (2) of Table 4.3. The estimated specifications are identical to those of Table 4.2, except for the fitted rejection probability as an additional regressor.

Table 4.4 shows that the coefficient estimate of the probability of rejection is positive and significant at the 1% level in each of the models. The estimates assume a value of around 1.16 and 0.79 in the 2SLS models. Using the average rejection probabilities for large and small firms, this implies that a small firm submits, on average, about a $5-7\%$ higher bid.

We interpret positive effect of the anticipated rejection rate as follows. Conditionally on entry, firms with a low chance of winning bid less aggressively, such that they can get a substantial profit if they win. The marginal benefit from aggressive bidding does not substantially increase the chance of winning, but decreases the margin.

The other coefficient estimates of interest are similar to that of Table 4.2 so that we do not comment further on them.

Table 4.4: *OLS and 2SLS estimates on logarithms of bids with rejection probabilities.*

	OLS		2SLS			
	(1)	(2)	(3)		(4)	
	Log of Bid	Log of Bid	Log of Bid		Log of Bid	
Fitted Pr{Reject}	1.22***	0.86**	1.16***		0.79***	
	(0.34)	(0.29)	(0.26)		(0.21)	
Log of # Bidders	-0.14	-0.12	-0.78***		-0.93***	
	(0.098)	(0.097)	(0.080)		(0.10)	
Log of # Bidders ×	-0.16***	-0.18***	-0.16***		-0.16***	
	(0.041)	(0.043)	(0.014)		(0.015)	
Log of Backlog	0.0088*	0.0059	0.011*		0.0083	
	(0.0043)	(0.0040)	(0.0054)		(0.0044)	
Scoring	-0.14	-0.12	-0.12***		-0.050	
	(0.079)	(0.078)	(0.023)		(0.029)	
Reconstruction	-0.61***	-0.59***	-0.71***		-0.71***	
	(0.12)	(0.11)	(0.050)		(0.047)	
Project & Build	-0.26***	-0.21**	-0.34***		-0.30***	
	(0.072)	(0.077)	(0.022)		(0.024)	
Continuation	-0.66***	-0.68***	-0.93***		-1.07***	
	(0.19)	(0.18)	(0.11)		(0.11)	
Foreign Firm	0.0041	0.023	-0.036		-0.017	
	(0.026)	(0.024)	(0.042)		(0.037)	
Region Dummies	No	Yes	No		Yes	
First-stage instrument			(FS1)	(FS2)	(FS1)	(FS2)
IV			0.78***	-0.056***	0.76***	-0.055**
			(0.31)	(0.017)	(0.032)	(0.018)
IV × One Stage			0.004	1.07***	0.0089	1.07***
			(0.0078)	(0.0043)	(0.0079)	(0.0045)
Specification tests						
Wu-Hausman			44.19***		56.88***	
Durbin-Wu-Hausman			85.91***		110.01***	
Observations	2167	2167	2164		2164	
Adjusted R^2	0.7799	0.8104	0.7519		0.7724	

Standard errors clustered by bidder. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Monetary values are expressed in 1 million PLN (2016 prices). All models include road characteristics. Road characteristics: log(length), road class, interaction terms between log(length) and class, number of junctions, dummy variables for bridge and tunnel, indicators for roads in minor and major urban areas, indicator for concrete road, monthly unemployment rate and the log of the average number of warm days in the region.

Our findings suggest that screening affects bids directly. Moreover, another channel deserves our attention. As the estimates suggest, entry has a profound role in shaping bidding strategies. That is, if the prospect of being rejected discourages entry, bids are indirectly affected by bidders facing a smaller number of competitors.

Table 4.5 shows estimates of alternative entry models that include the estimated rejection probability from the respective models of Table 4.3. The unit of observation corresponds to a bidder in a certain auction.

Table 4.5: *Logit Model on Entry*

	(1)	(2)	(3)	(4)	(5)	(6)
	entry	entry	entry	entry	entry	entry
Fitted Pr{Reject}	-4.52**	-3.71*	-4.39*	-3.99*		
	(1.73)	(1.69)	(2.01)	(2.02)		
Mean Annual					-0.67	-0.71
Rejection Rate					(0.87)	(0.89)
Log of Backlog	0.50***	0.42***	0.50***	0.41***	0.52***	0.43***
	(0.043)	(0.045)	(0.047)	(0.049)	(0.040)	(0.043)
Reconstruction	-0.19*	-0.23*	-0.19*	-0.23*	-0.19*	-0.22*
	(0.096)	(0.099)	(0.096)	(0.099)	(0.094)	(0.098)
Project & Build	-0.22*	-0.23**	-0.22*	-0.24*	-0.18*	-0.20*
	(0.090)	(0.090)	(0.091)	(0.092)	(0.082)	(0.083)
Continuation	-0.81***	-0.87***	-0.82***	-0.87***	-0.98***	-1.02***
	(0.19)	(0.19)	(0.18)	(0.19)	(0.17)	(0.18)
Two Stages	-0.12	-0.060	-0.12	-0.059	-0.14	-0.062
	(0.083)	(0.083)	(0.083)	(0.083)	(0.082)	(0.082)
Scoring	0.11	-0.0012	0.098	0.0091	-0.13	-0.20
	(0.16)	(0.16)	(0.16)	(0.16)	(0.14)	(0.14)
Large Firm Threshold	15	15	20	20	15	15
Min. # of Bids	1	3	1	3	1	3
Observations	39600	22800	39600	22800	39600	22800

Standard errors clustered by bidder. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Monetary values are expressed in 1 million PLN (2016 prices). All models include road characteristics and regional and year dummies. Road characteristics: log(length), road class, interaction terms between log(length) and class, number of junctions, dummy variables for bridge and tunnel, indicators for roads in minor and major urban areas, indicator for concrete road, monthly unemployment rate and the log of the average number of warm days in the region.

The models differ in the considered threshold for firm size to check robustness. Furthermore, models (1) and (3) include all firms whereas models (2) and (4) exclude bidders with only 1 or 2 entries. The coefficient estimates clearly show that rejections have a significant negative effect on entry as all coefficients are significant at the 5% level. Another remarkable feature is the effect of backlog that serves as a proxy for a firm's ability to undertake new projects. Contrary to the expected sign, the estimates are positive and highly significant, suggesting that large firms, which are more likely to accumulate backlogs, are also more likely to enter.

That is, the entry and bidding models imply a rather clear pattern. First, rejections increase bids conditionally on entry. Second, they lower entry.

Overall, the evidence strongly supports Hypothesis 2. We find that the expected probability of rejection has a strong negative effect on entry, and this effect persists in bidders of different market presence.

Given that screening affects small firms differently, the total effect of the screening policy on prices is positive. As the model does not account for delays and costs, our model suggests that screening puts a burden on the procurer also by decreasing entry.

Proposition 2 makes a distinction between firms with high and low production cost c_i. As the logit models in Table 4.3 show, competitive large firms are shielded against rejections while small firms suffer higher rates if they are more competitive. This ambiguity does not allow us to learn if the average rejection rate has a negative effect on entry. We revisit Proposition 2 in Models (5) and (6). In comparison to the other models, the estimated individual rejection probability is replaced by the annual observed rejection rate, which is equal to the share of rejected bids in the set of all submitted offers. As the coefficient estimate is not significant, we find that the average rejection rate has no effect on entry. It means that the two opposing effects listed in Proposition 2 cancel out.

The overall effect of screening on entry is tested by imputing an average rejection rate of 9.4% in model (3) of Table 4.5, instead of firm-specific fitted probabilities, which yields an increase of total entry rate from 5.58% to 6.1%. This boost in entry translates to an expected increase in the number of bids equal to 9.36%. These results suggest that a leveled screening policy results in higher total entry.

4.6 Conclusions

Myerson (1981) shows that a level playing field may have adverse effects in auction markets. While this idea is strongly endorsed in the literature, the primary focus is on *ex* *ante* measures supporting small- and medium enterprises. This paper contributes to this rich area of research by extending the scope to bid screening and litigation.

It is a standard procedure that the procurer may reject bids based on financial or technical criteria and an appeal against the decision is possible in the court room. We show, using a unique data set on Polish highway procurement auctions, that the opportunity to make an appeal, combined with the initial screening process of the procurer, adversely affects small firms. While large firms enjoy lower rejection rates if they bid more aggressively, this pattern appears to be the opposite for smaller enterprises. Our observation is consistent with two initial assumptions. First, smaller firms stand a smaller chance in the screening process. Second, by submitting lower bids, opposing bidders exert higher effort to justify their claims and make the procurer or the court of appeals to reject the tender.

Given that foreign tenderers are typically large enterprises, the legal system results in an interesting paradox. While domestic firms could be presumed to have a better understanding of domestic procurement law, foreign firms are rejected with a lower probability. This effect is correlated with the firm size, possibly because large firms hire competent permanent legal teams, if they plan a longer presence in the local market. It is an open question how new entrants would be affected under the comparable circumstances. Given that the biggest players in the market have been present in the country years before the start of the highway-building program, our data does not allow to investigate this issue further.

The primary function of screening is to shield the procurer from risk, including bankruptcy of the contractor. Depending on the development level of the domestic financial markets, screening may be replaced by alternative procedures. Standard solutions include letters of credit issued by banks or surety bonds issued by surety companies. These tools have multiple benefits over elaborate screening. They reduce transaction costs, mitigate extensive delays in contracting and let the market price potential risks. However, their availability may be limited in the Polish insurance market, which barely reaches 2.9% of GDP. This is small in comparison to major European economies, in which surety insurance is a standardized product in public procurement and where insurance markets are much more developed.[23] Depending on the availability of the necessary instruments, insurance of construction contracts could improve the current public procurement system in Poland.

[23] For example, insurance markets account for 7.97% of GDP in Italy and for 9.21% in the United Kingdom, OECD (2016a).

4.7 Appendix A: Highway construction boom in Poland

Figure 4.2 illustrates the progress in highway construction at the beginning and end of the sample period. The maps include A and S class roads.[24]

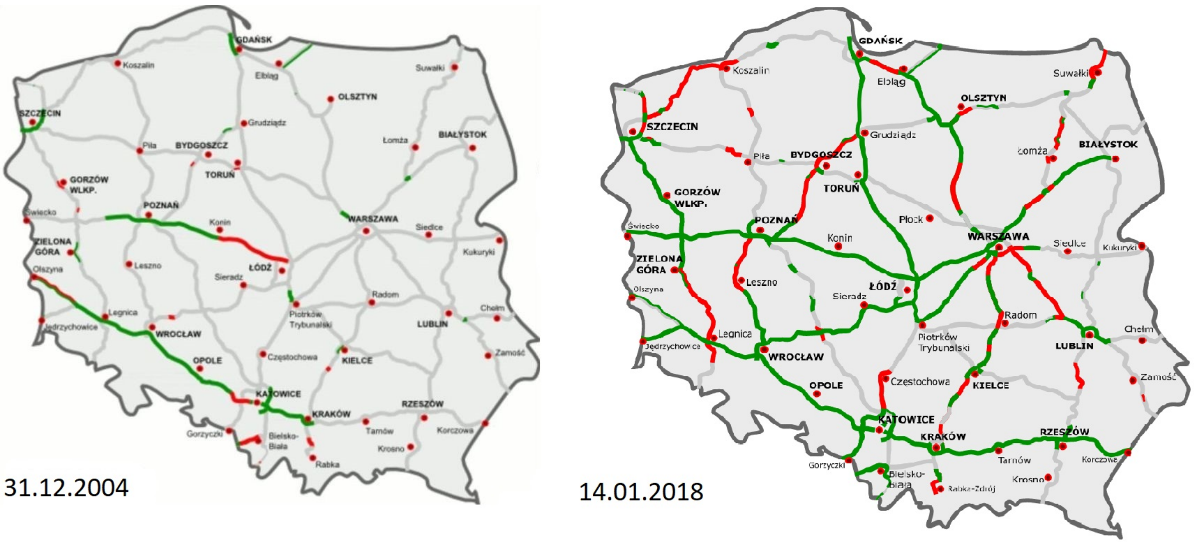


Figure 4.2: *Network of A and S class highways in Poland in 2004 (left) and in 2018 (right). Green roads are ready, red ones are under construction and gray ones are the planned highway network.*

4.8 Appendix B: Data Sources

We compiled data from the following main sources. Most data sources are available only in Polish and have been collected manually from separate documents.

1. **Auction and Road Data:** Summaries and documents on completed auctions are publicly available on the internet website of GDDKiA.[25] The data is presented in a non-standardized format and sometimes it is incomplete. In certain cases the data was provided on request. GDDKiA data contains names of bidders, their bids as well as auction characteristics (i.e. data on the auction mechanisms, number of auction stages, auction dates, dates when the contracts were signed and their types).

[24]The maps exclude the G and GP class, which are most often built as new beltways around cities and towns.

[25]www.gddkia.gov.pl/en/27/about-directorate

Technical data on road characteristics, opening dates etc. is collected from various sources, mainly from GDDKiA services, which provide data on road length, number of junctions, road types, tunnels etc. Certain characteristics were supplemented using local newspapers, Google Maps or Google Views services if GDDKiA data was incomplete or not precise enough. This method was adopted to match junctions and regions (voivodships) with a given road contract and obtain construction material data.

2. **Legal Data:** Documentation of court rulings is publicly available at the website of KIO.[26] The site contains all KIO decisions from 2007 onwards. Standardized documentation includes all essential details of the legal process.

3. **Macroeconomic Data:** Data on unemployment and inflation indices comes from the Central Statistical Office (Główny Urząd Statystyczny, GUS[27])

4. **Firm Data:** The capital ownership of firms was collected from the respective firms' websites.

4.9 Appendix C: List of variables

Data on road characteristics and auctions

Auction Date Date when an auction is announced.

Backlog The total value of uncompleted contracts of a given company at the time of bidding.

Bid Bid expressed in Polish złoty (PLN) in 2016 prices. In 2016, 1 EUR exchanged to 4.36 PLN.

Bridge Dummy variable, takes 1 if there is a large bridge in a given road segment.

Capacity Maximum backlog in the whole period of the sample for a given firm.

[26] www.uzp.gov.pl/kio/english-version/home
[27] www.stat.gov.pl/en/

Competitiveness The average relative bid (bid/winning bid) of a bidder for the entire sample.

Concrete Dummy variable, takes 1 if road is made of concrete, 0 if it is made of asphalt.

Contract Date Date when a contract is signed.

Contract Type Four contract dummies for 'Build', 'Project & Build', 'Continuation' and 'Reconstruction' contract types.

Dissolution Date Date when the contract is dissolved.

Foreign Leader Dummy variable, assumes 1 if the main contractor is foreign or in at least 51% ownership of a foreign firm.

Junctions AS Number of grade-separated junctions with other A/S roads.

Junctions O Number of grade-separated junctions with other roads of non-A/S class.

Large Consortium Member Dummy variable, assumes 1 if there is another large firm listed as a consortium member other than the main contractor.

Length Length of the road in km.

Number of Bids Number of firms or consortia submitting bids in the auction.

Opening Date Date when a given road is opened.

Price Weight The percentage weight on price in the awarding criteria, assumes 1 in a lowest-price auction and lower value in scoring auctions. Its value is at least 0.7 in the data set.

Region Dummy variables corresponding to one of the 16 voivodships (highest-level administrative regions) of Poland.

Road A/S/O Road type classification dummies (A-class, S-class or other road type).

Scoring Dummy variable, takes 1 if the awarding mechanism is a scoring auction, no matter what weights are applied to quality and completion time.

Minor Number of minor consortium members, assumes 0 if there is a single firm. Their identity is available.

Tunnel Dummy variable, takes 1 if there is a tunnel in a given road segment.

Unemployment Rate of registered unemployment in the auction announcement month.

Urban 0/1/2 Dummy variables corresponding to the area in which a road is located: uninhabited area, minor urban area and major urban area, respectively.

Utilization Rate The backlog of a firm divided by the maximum backlog of that firm during the entire period. If a firm does not win any contract, the utilization rate is always zero.

Vegetation Average number of vegetation days in a given region per year, i.e. days with the average temperature above 5 Celsius degrees.

Data on appeals

Defendant Name of company which offer is subjected to litigation.[28]

Plaintiff Name of plaintiff(s).

Type I Appeal Number of appeals related to the auction which either claim that a rejected tender is valid or that the procedure has general irregularities. It does not depend on the number of plaintiffs of individual cases.

Type II Appeal Number of appeals which claim that a tender submitted by a party different from the plaintiff is invalid. It does not depend on the number of plaintiffs of individual cases.

Verdict Dummy variable, takes 1 if the court upheld GDDKiA's decision.

[28] Technically, the defendant is GDDKiA in every case.

4.10 Appendix D: List of large firms

Table 4.6 lists the 20 largest firms ranked by the number of submitted bids as individual firms or as main consortium members. Subsidiaries are counted together with the owner. Hermann Kirchner's shares are owned 100% by Strabag since 2012. Hochtief since 2011 and Pol-Aqua since 2009 are owned by Dragados. Other subsidiaries have unbroken ownership the sample period. In the sample, 124 small main contracting firms submitted 712 bids. Among these, 331 were submitted by domestic and 381 by foreign firms.

Table 4.6: *List of 20 most frequent bidders.*

Name	bids	won	m. share	rej. rate	foreign
Ferrovial & Budimex & M. Kraków	269	51	17.8%	2.2%	yes
Strabag & Dywidag & Heilit+Woerner	244	42	13.1%	5.6%	yes
Mota-Engil	138	19	5.1%	4.3%	yes
Dragados & Pol-Aqua & Hochtief	126	21	7.7%	7%	yes
Vinci & Eurovia & Warbud	110	11	3.3%	9.9%	yes
Acciona & Mostostal Warszawa	93	11	3.7%	3.2%	yes
Bilfinger	70	7	3.1%	10.0%	yes
Astaldi	64	9	5.3%	9.4%	yes
Polimex-Mostostal	57	8	5.4%	11.7%	no
Intercor	54	5	1.3%	11.1%	no
Mosty Łódź	51	1	0.3%	11.7%	no
Mirbud & Kobylarnia	51	6	1.1%	17.7%	no
Skanska	51	10	1.9%	5.9%	yes
MSF	49	1	0.2%	10.2%	yes
Aldesa	47	1	0.2%	15.2%	yes
Porr & Teerag	46	5	1.8%	13%	yes
Bunte	42	2	0.7%	23.8%	yes
Hermann Kirchner	41	8	2.4%	4.9%	yes
Salini & Todini	40	7	2.8%	7.5%	yes
PBDM Mińsk Mazowiecki	38	2	0.1%	2.6%	no
Total share	70.2%	75.6%	77.7%	-	-

4.11 Appendix E: Proofs of Propositions

Proposition 1: *For any c_i, the equilibrium bid of a firm $b^*(c_i)$ is a decreasing function of the number of entering bidders k and of the number of potential entrants n.*

Proof. Let us denote the entry function with $I(c_i, e_i) : \{c_i, e_i\} \rightarrow \{0, 1\}$, in which $I(c_i, e_i) = 1$ indicates entry. First, we show that the set $\{c_i, e_i\} : I(c_i, e_i) = 1$ is compact and satisfies the condition that $I(c_i, e_i) \geq I(c_i', e_i')$ if $c_i' \geq c_i$ and $e_i' \geq e_i$. That is, higher costs hinder entry. This result is derived directly from the entry condition $max_{b_i}\ \mathbb{E}\left[\pi(b_i, b_{-i}, c_i)\right] \geq e_i$. For any strategy b_{-i} of outside bidders, higher entry cost e_i means that there is a lower critical value of c_i such that the condition is satisfied. The logic is similar for c_i.

The optimal bid does not depend on the entry cost, but only on c_i, as e_i is sunk cost at the point of bidding. Bidder i faces the problem $\underset{b_i}{max}\ \mathbb{E}\left[\pi(b_i, b_{-i}, c_i)\right] = (b_i - c_i) \cdot Pr\{b_i = b_1\}$. We focus our attention on a symmetric equilibrium, in which the equilibrium bidding function is continuous. Conditionally on that, the number of potential bidders equals n, the number of entries equals k and the cumulative density function of bids is $F_n(b)$, the probability of winning is equal to the probability of b_i being the lowest bid. Function $F_n(b)$ is conditional on the number of potential bidders n, as it determines the set of entering types, i.e. those $\{c_i, e_i\}$ for which $I(c_i, e_i) = 1$. That is, the expected profit is

$$\underset{b_i}{max}\ (b_i - c_i) \cdot Pr\{b_i = b_1\} = (b_i - c_i) \cdot (1 - F_n(b_i))^{k-1}, \tag{4.5}$$

where $(1 - F_n(b_i))^{k-1}$ expresses the probability that b_i is smaller than other $k - 1$ bids.[29] For any e_i, c_i has to satisfy that $(b_i - c_i) \cdot (1 - F_n(b_i))^{k-1} \geq e_i$. Let us denote the critical type with $\hat{c}_i(e_i, n)$ for which $(b_i - \hat{c}_i(e_i, n)) \cdot F_n^k(b_i) = e_i$. The first-order condition of (4.5) is

$$-(b_i - c_i)(k - 1) \cdot (1 - F_n(b_i))^{k-2} \cdot f_n(b_i) + 1 - F_n^{k-1}(b_i) = 0, \tag{4.6}$$

where $f_n(b_i)$ is the corresponding density function. The solution of this differential equation is

$$b^*(c_i) = \underline{c} + \frac{\int_{c_i}^{\underline{c}} (1 - F_n(s))^{k-1}\, ds}{(1 - F_n(c_i))^{k-1}}, \tag{4.7}$$

which is a decreasing function of k Milgrom (2004).

[29]Here we ignore ties as their probability is zero.

Now, let us address the second part of the theorem. The auction stage profit is given by

$$\left(c_i + \frac{\int_{c_i}^{\underline{c}} (1 - F_n(s))^{k-1}\, ds}{(1 - F_n(c_i))^{k-1}} - \underline{c}\right) \cdot (1 - F_n^{k-1}(c_i)). \tag{4.8}$$

This expression is decreasing in k. In order to prove the theorem, we show that $\hat{c}(e_i, n)$ is decreasing in the potential number of bidders. Suppose that $\hat{c}(e_i, n+1) \geq \hat{c}(e_i, n)$. Then, bidder i with $c_i = \hat{c}(e_i, n)$ enters if the number of potential bidders is $n+1$. Given the bidding function is decreasing in k and the *ex ante* probability of entry is larger, the profit of i is lower than e_i, which is a contradiction. Revisiting (4.7) shows that, because the conditional cumulative density function is larger at any $c_i \leq \hat{c}(e_i, n+1)$ then in the case when the number of potential bidders is n, bids are lower. □

Proposition 2: *For constant $\alpha(e_i) = \alpha$, the marginal effect of α on entry is negative for firms with sufficiently low costs c_i, and positive for firms with sufficiently high costs c_i.*

Proof. Entry occurs if a bidder's interim expected payoff is non-negative, hence

$$\mathbb{E}\left[\Pi(b_i, b_{-i}, c_i, \alpha(e_i))\right] \geq e_i \tag{4.9}$$

where $\Pi(b_i, b_{-i}, c_i, \alpha(e_i))$ is the auction payoff, conditionally on entry, in an auction with a screening function $\alpha(\cdot)$. As in the baseline model, the set of types entering the auction is compact. The proof is identical to that of Proposition 1. We denote the set of entering types with $\{c_i, e_i\} : I(c_i, e_i) = 1$.

Next, we characterize the bidding stage and focus on symmetric Bayesian equilibria with a strictly increasing continuous bidding function. Suppose that the number of potential entrants is n, k bidders enter and the conditional distribution function of bids is denoted with $F_{n,\alpha}(b)$.

The objective function of bidder i is

$$\begin{aligned} &\underset{b_i}{max}\ \ (b_i - c_i) \cdot \left((1 - F_{n,\alpha}(b_i))^{k-1}\ (1 - \alpha(e_i)) + \int_{b_i}^{\bar{e}} \alpha(s)(k-1)F_{n,\alpha}(b_i)(1 - F_{n,\alpha}(b_i))^{k-2}\, ds \right) = \\ &\quad \underset{b_i}{max}\ \ (b_i - c_i) \cdot \left((1 - F_{n,\alpha}(b_i))^{k-1}(1 - \alpha(e_i)) + (k-1)F_{n,\alpha}(b_i)(1 - F_{n,\alpha}(b_i))^{k-2} \int_{b_i}^{\bar{e}} \alpha(s)\, ds \right) \end{aligned} \tag{4.10}$$

This objective function is different from that of the baseline model. Bidder i may win in two cases, if b_i is ranked first and not rejected or if it is ranked second, but the first-ranked bid is rejected. The baseline model is an extreme case in which $\alpha(e_i) = 0$.

Now consider the case when the rejection function is constant, $\alpha(e_i) = \alpha \geq 0$. First, we calculate the expected price for the auctioneer given that k firms enter. In a second-price auction it is a dominating strategy to bid one's own valuation, that is, the expected cost to the auctioneer is

$$\begin{aligned} \mathbb{E}((1-\alpha)\cdot b_1 + \alpha \cdot b_2) = \mathbb{E}((1-\alpha)\cdot c_2 + \alpha \cdot c_3) = \\ \mathbb{E}(c_2 + \alpha \cdot (c_3 - c_2)). \end{aligned} \tag{4.11}$$

In this model we can use the Revenue Equivalence Theorem (see Börgers (2015) for details). As types are independently drawn, utility is a linear function and the type space is convex, second and first-price auctions yield the same expected cost to the auctioneer. That is, (4.11) is the expected winning bid in our first-price auction model. Formula (4.11) highlights an important property of using screening. With given entry, it substitutes the winning bid with a more expensive offer.

Using the Payoff Equivalence Theorem (see Börgers (2015) for details), the expected payoff of the winner conditional on facing $k-1$ other entrants equals

$$\begin{aligned} (\mathbb{E}(c_2|c_i = c_1) - c_i)\cdot(1-\alpha)\cdot(1-F_n(c_i))^{k-1} + \\ (\mathbb{E}(c_3|c_i = c_2) - c_i)\cdot \alpha \cdot (k-1)\cdot F_n(c_i)\cdot(1-F_n(c_i))^{k-2}. \end{aligned} \tag{4.12}$$

The first term corresponds to the case if bidder i wins and the second term is the expected payoff from winning as the second lowest bidder.

For example, the chance that exactly k bidders decide to submit a bid equals $\binom{n}{k}(\int_{\underline{e}}^{\overline{e}} \hat{c}_i(e_i))^k \cdot (1 - (\int_{\underline{e}}^{\overline{e}} \hat{c}_i(e_i)))^{n-k}$. From this, the expected payoff with cost c_i equals

$$\begin{aligned} \sum_{k=1}^{n} \Bigg(\binom{n}{k} \cdot \Big((\mathbb{E}(c_2|c_i = c_1) - c_i)\cdot(1-\alpha)\cdot(1-F_n(c_i))^{k-1} + \\ (\mathbb{E}(c_3|c_i = c_2) - c_i)\cdot \alpha \cdot (k-1)\cdot F_n(c_i)\cdot(1-F_n(c_i))^{k-2}\Big)\Bigg). \end{aligned} \tag{4.13}$$

The most competitive type with cost $c_i = \underline{c}$ clearly suffers from higher α as $\lim_{c_i \to \underline{c}} = F_n(c_i) = 0$. Similarly, bidders with very high type enjoy a higher α. For constant

$\alpha(e_i) = \alpha$, there is a right-side environment of $\underline{c}$ on which the marginal entry effect of α is negative on entry, and there is a right-side environment of $\overline{c}$ on which it is positive.

$\square$

5 References

Aiyagari, R. (1994). Uninsured idiosyncratic risk and aggregate saving. *Quaterly Journal of Economics, vol. 109(3), p. 659-684*.

Albano, G. L., Spagnolo, G., & Zanza, M. (2008). Regulating joint bidding in public procurement. *Journal of Competition Law and Economics*, *vol. 5*(2), p. 335–360.

Alesina, A., & Bayoumi, T. (1996). The costs and benefits of fiscal rules: evidence from U.S. states. *NBER Working Paper 5614*.

Alesina, A., Campante, F., & Tabellini, G. (2008). Why is fiscal policy often procyclical? *Journal of the European Economic Association, vol. 6(5), p. 1006-1036*.

Alesina, A., & Passalacqua, A. (2016). The political economy of government debt. In *Handbook of Macroeconomics 2B, ch. 34.* Elsevier.

Alesina, A., & Perotti, R. (1994). The political economy of budget deficits. *NBER Working Paper 4637*.

Alesina, A., & Perotti, R. (1996). Budget deficits and budget institutions. *NBER Working Paper 5556*.

Alfaro, L., & Kanczuk, F. (2016). Fiscal rules and sovereign default. *Harvard Business School Working Paper 16-134*.

Arve, M., & Chiappinelli, O. (2018). *The role of budget constraints in pre-commercial procurement* (Tech. Rep.). mimeo.

Asker, J., & Cantillon, E. (2008). Properties of scoring auctions. *The RAND Journal of Economics*, *vol. 39*(1), p. 69–85.

Asker, J., & Cantillon, E. (2010). Procurement when price and quality matter. *The RAND Journal of Economics*, *vol. 41*(1), p. 1–34.

Athey, S., Coey, D., & Levin, J. (2013). Set-asides and subsidies in auctions. *American Economic Journal: Microeconomics*, *vol. 5*(1), p. 1–27.

Attanasio, O., & Weber, G. (2010). Consumption and saving: models of intertemporal allocation and their implications for public policy. *Journal of Economic Literature, vol. 48(3), p. 693-751*.

Auerbach, A. (2014). Budget rules and fiscal policy: ten lessons from theory and evidence. *German Economic Review, vol. 15(1), p. 84-99*.

Backhus, D., & Kehoe, P. (1992). International evidence on the historical properties of business cycles. *American Economic Review, vol. 82(4), 864-888*.

Bajari, P., Houghton, S., & Tadelis, S. (2014). Bidding for incomplete contracts: an empirical analysis of adaptation costs. *American Economic Review*, *vol. 104*(4), p.

1288–1319.
Balat, J., Komarova, T., & Krasnokutskaya, E. (2017). *Ex-ante and ex-post subcontracting in highway procurement market* (Tech. Rep.). Working Paper, Johns Hopkins University.
Beetsma, R., & Debrun, X. (2016). Fiscal councils: rationale and effectiveness. *IMF Working Paper 16/86*.
Beljean, T., & Geier, A. (2013). The Swiss debt brake - has it been a success ? *Swiss Journal of Economics and Statistics, vol. 2(2), p. 115-135*.
BenPorath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy, vol. 75(4), p. 352-365*.
Bhuller, M., Mogstad, M., & Salvanes, K. (2017). Life cycle earnings, education permiums and internal rates of return. *Journal of Labor Economics, vol. 35(4), p. 993-1030*.
Blanchard, O., Dell'Ariccia, G., & P, P. M. (2010). Rethinking macroeconomic policy. *IMF Working Paper*.
Blanchard, O., & Perotti, R. (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *Quarterly Journal of Economics, vol. 107(4), p. 1329-1368*.
Blundell, R., Pistaferri, L., & Preston, I. (2008). Consumption inequality and partial insurance. *American Economic Review, vol. 98(5), p. 1887–1921*.
Bohn, H., & Inman, R. (1996). Balanced budget rules and public deficits: evidence from the U.S. states. *NBER Working Paper 5533*.
Börgers, T. (2015). *An introduction to the theory of mechanism design.* Oxford University Press.
Bovenberg, L., & Jacobs, B. (2005). Redistribution and education subsidies are Siamese twins. *Journal of Public Economics, vol. 89(11-12), p. 2005-2035*.
Branzoli, N., & Decarolis, F. (2015). Entry and subcontracting in public procurement auctions. *Management Science*, *vol. 61*(12), p. 2945–2962.
Brinca, P., Holter, H., Krusell, P., & Malafry, L. (2015). Fiscal multipliers in the 21st century. *Working Paper*.
Brown, M., Haughwout, A., Lee, D., Scally, J., & van der Klaauw, W. (2014). Measuring student debt and its performance. *Federal Reserve Bank of New York Staff Reports*.
Browning, M., & Crossley, T. (2001). The life-cycle model of consumption and saving. *Journal of Economic Perspectives, vol.15(3), p. 3-22*.
Bruchez, P. (2003). A modification of the HP filter aiming at reducing the end-point bias. *The Federal Finance Administration of Switzerland Working Paper 3/2003*.
Brzozowski, M., & Siwińska-Gorzelak, J. (2010). The impact of fiscal rules on fiscal policy volatility. *Journal of Applied Economics, vol. 13(2), p. 205-231*.
Burriel, P., deCastro, F., Garrote, D., Gordo, E., Paredes, J., & Perez, J. (2009). Fiscal

policy shocks in the Euro area and in the US. *ECB Working Paper 1133*.

Caldara, D., & Kamps, C. (2008). What are the effects of fiscal policy shocks? a VAR-based comparative analysis. *ECB Working Paper 877*.

Campbell, J., & Mankiw, G. (1989). Consumption, income and interest rates: reinterpreting time series evidence. *NBER Macroeconomics Annual*.

Card, D. (1999). The causal effect of education on earnings. Elsevier.

Carroll, C. (1997). Buffer-stock saving and the life cycle/permanent income hypothesis. *Quaterly Journal of Economics, vol. 112(1), p. 1-55*.

Carroll, C., Slacalek, J., Tokuoka, K., & White, M. (2015). The distribution of wealth and the marginal propensity to consume. *Working Paper*.

Checherita-Westphal, C., & V.Zdarek. (2017). Fiscal reaction function and fiscal fatigue: evidence for the euro area. *ECB Working Paper 2036*.

Choudhary, A., Hanif, N., & Iqbal, J. (2013). On smoothing macroeconomic time series using HP and modified HP filter. *Munich Personal RePEc Archive*.

Conley, T. G., & Decarolis, F. (2016). Detecting bidders groups in collusive auctions. *American Economic Journal: Microeconomics*, *vol. 8*(2), p. 1–38.

Coviello, D., & Mariniello, M. (2014). Publicity requirements in public procurement: evidence from a regression discontinuity design. *Journal of Public Economics*, *vol. 109*, p. 76–100.

Coviello, D., Moretti, L., Spagnolo, G., & Valbonesi, P. (2017). Court efficiency and procurement performance. *The Scandinavian Journal of Economics*, *vol. 120*(3), p. 826–858.

Cronqvist, H., & Siegel, S. (2015). The origins of savings behavior. *Journal of Political Economy, vol. 123(1), p. 123-169*.

Deaton, A. (1991). Saving and liquidity constraints. *Econometrica, vol. 59(5), p. 1221-1248*.

Debrun, X., & Kumar, M. (2007). The discipline-enhancing role of fiscal institutions: theory and empirical evidence. *IMF Working Paper 07/171*.

Debrun, X., Symanski, S., & Epstein, N. (2008). A new fiscal rule: should Israel go Swiss? *IMF Working Paper 08/87*.

Decarolis, F. (2014). Awarding price, contract performance, and bids screening: evidence from procurement auctions. *American Economic Journal: Applied Economics*, *vol. 6*(1), p. 108–132.

de Haan, J. (2014). Democracy, elections and government budget deficits. *German Economic Review, vol. 15(1), p. 131-142*.

De Silva, D. G., Dunne, T., Kankanamge, A., & Kosmopoulou, G. (2008). The impact of public information on bidding in highway procurement auctions. *European Economic Review*, *vol. 52*(1), p. 150–181.

De Silva, D. G., Jeitschko, T. D., & Kosmopoulou, G. (2009). Entry and bidding in common and private value auctions with an unknown number of rivals. *Review of Industrial Organization*, *vol. 35*(1-2), p. 35–73.

De Silva, D. G., Kosmopoulou, G., & Lamarche, C. (2017). Subcontracting and the survival of plants in the road construction industry: a panel quantile regression analysis. *Journal of Economic Behavior & Organization*, *vol. 137*, p. 113–131.

Deutsche Bundesbank. (2011). *Die Schuldenbremse in Deutschland - wesentliche Inhalte und deren Umsetzung (in German).* Monatsbericht Oktober.

Dimitri, N., Piga, G., & Spagnolo, G. (2006). *Handbook of Procurement.* Cambridge University Press.

Doorn, D. (2001). Consequences of Hodrick-Prescott filtering for parameter estimation on a structural model of inventory behavior. *Proceedings of the Annual Meeting of the American Statistical Association.*

European Central Bank. (2013). *The importance and effectiveness of national fiscal frameworks in the EU.* ECB Monthly Bulletin.

European Commission. (2014a). *Adjusting the budget balance for the business cycle: the EU methodology.* Economic Papers 536.

European Commission. (2014b). *The production function methodology for calculating potential growth rates and output gaps.* Economic Papers 535.

Favero, C., & Giavazzi, F. (2007). Debt and the effects of fiscal policy. *Federal Reserve Bank of Boston Working Paper*.

Favero, C., & Giavazzi, F. (2012). Measuring tax multipliers: the narrative method in fiscal VARs. *American Economic Journal: Economic Policy, vol. 4(2), p. 69-94*.

Favero, C., & Marcellino, M. (2005). Modelling and forecasting fiscal variables for the Euro area. *Oxford Bulletin of Economics and Statistics, vol. 67(s1), p. 755-783*.

Federal Ministry of Finance, Germany. (2015). *German's federal debt brake.* Official Paper.

Feldstein, M. (2009). Rethinking the role of fiscal policy. *NBER Working Paper 14684*.

Foster, K., Schuh, S., & H, H. Z. (2010). The 2015 survey of consumer payment choice: Summary results. *Federal Reserve Bank of Boston*.

Frankel, J., & Schreger, J. (2013). Over-optimistic official forecasts in the Eurozone and fiscal rules. *Review of World Economics, vol. 149(2), p. 247-272*.

Frankel, J., Vegh, C., & Vuletin, G. (2011). On graduation from fiscal procyclicality. *NBER Working Paper 17619*.

Gali, J., Valles, J., & Lopez-Salido, J. D. (2007). Understanding the effects of government spending on consumption. *Journal of the Economic Association, vol. 5(1), p. 227-270*.

Geier, A. (2011). The debt brake - the Swiss fiscal rule at the federal level. *The Federal*

Finance Administration of Switzerland Working Paper 15.

Geier, A. (2012). Application of the Swiss fiscal rule to artificial data: a Monte Carlo simulation. *Swiss Journal of Economics and Statistics, vol. 1(2), p. 37-55*.

Gil, R., & Marion, J. (2013). Self-enforcing agreements and relational contracting: evidence from California highway procurement. *The Journal of Law, Economics, and Organization*, *vol. 29*(2), p. 239–277.

Gordon, R. (2015). Secular stagnation: a supply-side view. *American Economic Review, vol. 105(5), p. 54-59*.

Goujard, A. (2016). Improving transport and energy infrastructure investment in Poland. *OECD Economic Department Working Papers*(1302).

Greene, C., Schuh, S., & Stavins, J. (2013). The 2015 survey of consumer payment choice: Summary results. *Federal Reserve Bank of Boston*.

Greene, C., Schuh, S., & Stavins, J. (2015). The 2015 survey of consumer payment choice: Summary results. *Federal Reserve Bank of Boston*.

Grembi, V., Nannicini, T., & Troiano, U. (2016). Do fiscal rules matter? *American Economic Journal: Applied Economics, vol. 8(3), p. 1-30*.

Guerguil, M., Mandon, P., & Tapsoba, R. (2017). Flexible fiscal rules and countercyclical fiscal policy. *Journal of Macroeconomics, vol. 52, p. 189-220*.

Gugler, K., Weichselbaumer, M., & Zulehner, C. (2015). Competition in the economic crisis: analysis of procurement auctions. *European Economic Review*, *vol. 73*, p. 35–57.

Gupta, S. (2001). The effect of bid rigging on prices: a study of the highway construction industry. *Review of Industrial Organization*, *vol. 19*(4), p. 451–465.

Halac, M., & Yared, P. (2014). Fiscal rules and discretion under persistent shocks. *Econometrica, vol. 82(5), p. 1557-1614*.

Haliassos, M., & Michaelides, A. (2013). Portfolio choice and liquidity constraints. *International Economic Review, vol. 44(1), p. 143-177*.

Heathcote, J., Storesletten, K., & Violante, G. (2009). Quantitative macroeconomics with heterogeneous households. *Working Paper*.

Heathcote, J., Storesletten, K., & Violante, G. (2013). Consumption and labor supply with partial insurance: an analytical framework. *Working Paper*.

Heer, B., & Maussner, A. (2008). Dynamic general equilibrium modelling, computational methods and applications. *Springer*.

Heun, W. (2014). Balanced budget requirements and debt brakes: feasibility and enforcement. *German Economic Review, vol. 15(1), p. 100-115*.

Holm-Hadulla, F., Hauptmeier, S., & Rother, P. (2012). The impact of expenditure rules on budgetary discipline over the cycle. *Applied Economics, vol. 44(25), 3287-3296*.

Hortaçsu, A., & McAdams, D. (2018). Empirical work on auctions of multiple objects.

Journal of Economic Literature, *vol. 56*(1), p. 157–184.

Huggett, M. (1993). The risk-free rate in heterogeneous-agent incomplete-insurance economies. *Journal of Economics Dynamics and Control, vol. 17, p. 953-969*.

Huggett, M. (1996). Wealth distribution in the life-cycle economies. *Journal of Monetary Economics, vol. 38(3), p. 469-494*.

Huggett, M., & Kaplan, G. (2015). How large is the stock component of human capital? *Review of Economic Dynamics, vol. 22, p. 21-51*.

Huggett, M., Ventura, G., & Yaron, A. (2006). Human capital and earnings distribution dynamics. *Journal of Monetary Economics, vol. 53(2), p. 265-290*.

Huggett, M., Ventura, G., & Yaron, A. (2011). Sources of lifetime inequality. *American Economic Review, vol. 101 (7), p. 2923-2954*.

Jappelli, T., & Pistaferri, L. (2010). The consumption response to income changes. *Annual Review of Economics*.

Jeziorski, P., & Krasnokutskaya, E. (2016). Dynamic auction environment with subcontracting. *The RAND Journal of Economics*, *vol. 47*(4), p. 751–791.

Jofre-Bonet, M., & Pesendorfer, M. (2003). Estimation of a dynamic auction game. *Econometrica*, *vol. 71*(5), p. 1443–1489.

Kaplan, G., Moll, B., & Violante, G. (2018). Monetary policy according to HANK. *American Economic Review, vol. 108(3), p. 697-743*.

Kaplan, G., & Violante, G. (2010). How much consumption insurance beyond self-insurance? *American Economic Journal: Macroeconomics, vol. 2(4), p. 53-87*.

Kaplan, G., & Violante, G. (2014). A model of the consumption response to fiscal stimulus payments. *Econometrica, vol. 82(4), p. 1199-1239*.

Kaplan, G., Violante, G., & Weidner, J. (2014). The wealthy hand to mouth. *Brookings Papers*.

King, R., Plosser, C., & Rebelo, S. (1988). Production, growth and business cycles. *Journal of Monetary Economics, vol. 21(2-3), p. 195-232*.

Kopits, G. (2001). Fiscal rules: useful policy framework or unnecessary ornament? *IMF Working Paper*.

Korniluk, D. (2013). Optymalizacja wartości parametrów reguły wydatkowej za pomocą algorytmu genetycznego. *Bank i Kredyt (in Polish), vol. 44(1), p. 67-98*.

Korniluk, D. (2016). The stabilizing expenditure rule in Poland - simulations for 2014-2040. *Eastern European Economics, vol. 54(4), p. 351-375*.

Kosmopoulou, G., & Zhou, X. (2014). Price adjustment policies in procurement contracting: an analysis of bidding behavior. *The Journal of Industrial Economics*, *vol. 62*(1), p. 77–112.

Krasnokutskaya, E., & Seim, K. (2011). Bid preference programs and participation in highway procurement auctions. *The American Economic Review*, *vol. 101*(6), p.

2653–2686.

Krueger, D., & Fernandez-Villaverde, J. (2011). Consumption and saving over the life cycle: how important are consumer durables? *Macroeconomic Dynamics, vol. 15(5), p. 725-770*.

Krueger, D., Perri, F., Pistaferri, L., & Violante, G. (2007). Cross sectional facts for macroeconomics. *Working Paper*.

Krusell, P., & Smith, A. (1998). Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy, vol. 106(5), p. 867-896*.

Kumar, M., Baldacci, E., Schaechter, A., Caceres, C., Kim, D., Debrun, X., … Yakadina, I. (2009). Fiscal rules – anchoring expectations for sustainable public finances. *IMF Working Paper*.

Lalive, R., Schmutzler, A., & Zulehner, C. (2017). The benefits of procurement auctions: competitive pressure vs. selection of efficient suppliers.

Landon, S., & Smith, C. (2017). Does the design of a fiscal rule matter for welfare ? *Economic Modelling, vol. 63, p. 226-237*.

Levin, D., & Smith, J. L. (1994). Equilibrium in auctions with entry. *The American Economic Review*, 585–599.

Lewis, G., & Bajari, P. (2011). Procurement contracting with time incentives: theory and evidence. *The Quarterly Journal of Economics*, *vol. 126*(3), p. 1173–1211.

Lewis, G., & Bajari, P. (2014). Moral hazard, incentive contracts, and risk: evidence from procurement. *The Review of Economic Studies*, *vol. 81*(3), p. 1201–1228.

Li, T., & Zheng, X. (2009). Entry and competition effects in first-price auctions: theory and evidence from procurement auctions. *The Review of Economic Studies*, *vol. 76*(4), p. 1397–1429.

Lucas, R. (2003). Macroeconomic priorities. *American Economic Review, vol. 93(1), p. 1-14*.

Luechinger, S., & Schaltegger, C. (2013). Fiscal rules, budget deficits and budget projections. *International Tax and Public Finance, vol. 20(5), p. 785–807*.

Marion, J. (2015). Sourcing from the enemy: horizontal subcontracting in highway procurement. *The Journal of Industrial Economics*, *vol. 63*(1), p. 100–128.

Milgrom, P. (2004). *Putting auction theory to work*. Cambridge University Press.

Miller, D. P. (2014). Subcontracting and competitive bidding on incomplete procurement contracts. *The RAND Journal of Economics*, *vol. 45*(4), p. 705–746.

Moore, J. (2014). *Procedural rules and their impact on procurement outcomes: evidence from France* (Tech. Rep.).

Myerson, R. B. (1981). Optimal auction design. *Mathematics of Operations Research*, *vol. 6*(1), p. 58–73.

Nerlich, C., & Reuter, W. (2016). Fiscal rules, fiscal space and procyclical fiscal policy.

ECB Working Paper 1872.

OECD. (2016a). *OECD data on insurance spending. Online data.* Retrieved 2018-05-01, from https://data.oecd.org/insurance/insurance-spending.htm

OECD. (2016b). *A stocktaking report on MENA public procurement systems* (Tech. Rep.).

Oh, H., & Reis, R. (2012). Targeted transfers and the fiscal response to the great recession. *Journal of Monetary Economics, vol. 15, p. 50-64*.

Paetz, C., Rietzler, K., & Truger, A. (2016). Die Schuldenbremse im Bundeshaushalt seit 2011: die wahre Belastungsprobe steht noch aus. *IMK Working Paper 117 (in German)*.

Pagel, M., & Vardardottir-Olafsson, A. (2016). The liquid hand-to-mouth: evidence from a personal finance management software. *Working Paper*.

Pigoń, A., & Ramsza, M. (2014). Ocena polskiej reguły fiskalnej. Analiza symulacyjna. *Studia i Prace Wydziału Nauk Ekonomicznych i Zarządzania (in Polish), vol. 35(2), p. 491-502*.

Pigoń, A., & Ramsza, M. (2016). Impact of a modified HP filter on countercyclical behavior of the Swiss fiscal rule. *Equilibrium, a Quarterly Journal of Economics and Economic Policy, vol. 11(4), p. 661-674*.

Porter, R., & Zona, D. (1993). Detection of bid rigging in procurement auctions. *Journal of Political Economy*, *vol. 101*(3), p. 518–538.

Ravn, M., & Uhlig, H. (2002). On adjusting the Hodrick-Prescott filter for the frequency of observations. *The Review of Economics and Statistics, vol. 84(2), p. 371–380*.

Samuelson, W. F. (1985). Competitive bidding with entry costs. *Economics Letters*, *vol. 17*(1-2), p. 53–57.

Schaechter, A., Kinda, T., Budina, N., & Weber, A. (2012). Fiscal rules in response to the crisis – toward the 'next-generation' rules: a new dataset. *IMF Working Paper 12/187*.

Spiegel, Y. (1993). Horizontal subcontracting. *The RAND Journal of Economics*, *vol. 24*(4), p. 570–590.

Stantcheva, S. (2016). Optimal taxation and human capital policies over the life cycle. *Journal of Political Economy, vol. 125(6), p. 1931-1990*.

Summers, L. (2015). Demand side secular stagnation. *American Economic Review, vol. 106(5), p. 503-507*.

Sydsaeter, K., Hammond, P., Seierstad, A., & Strom, A. (2008). Further mathematics in economic analysis. *Prentice Hall*.

Truger, A., & Will, H. (2012). The German debt brake - a shining example for European fiscal policy? *Working Paper*.

Wyplosz, C. (2008). Fiscal policy councils: unlovable or just unloved? *Swedish Economic*

Policy Review, vol. 15, p. 173–192.

Wyplosz, C. (2013). Fiscal rules: theoretical issues and historical experiences. In *Fiscal policy after the financial crisis, ch. 12.* University of Chicago Press.

6 Curriculum Vitae

Education

2014 – 2018	Ph.D. in Economics and Finance, University of St. Gallen
	Main advisor: Prof. Dr. Winfried Koeniger
2011 – 2013	M.Sc., Economics, Warsaw School of Economics
2010 – 2011	M.Res., Economics, Tilburg University (no degree)
2007 – 2010	B.Sc., Economics, Warsaw School of Economics

Professional experience

2018 – onwards	Senior economist, Institute for Structural Research (Warsaw)
2014 – 2018	Research assistant, University of St. Gallen
	Swiss Institute for Empirical Economic Research
	Chair of Macroeconomics, Prof. Dr. Winfried Koeniger
2013 – 2014	Teaching assistant, Warsaw School of Economics
	Chair of Mathematical Economics, Prof. Dr. Michał Ramsza
2011	Internship, Macroeconomic Department of Bank BGŻ (Warsaw)
2010	Internship, Institute for Structural Research (Warsaw)
2009	Internship, Polish Ministry of Finance (Warsaw)

Contact details

Email: adampigon@gmail.com